

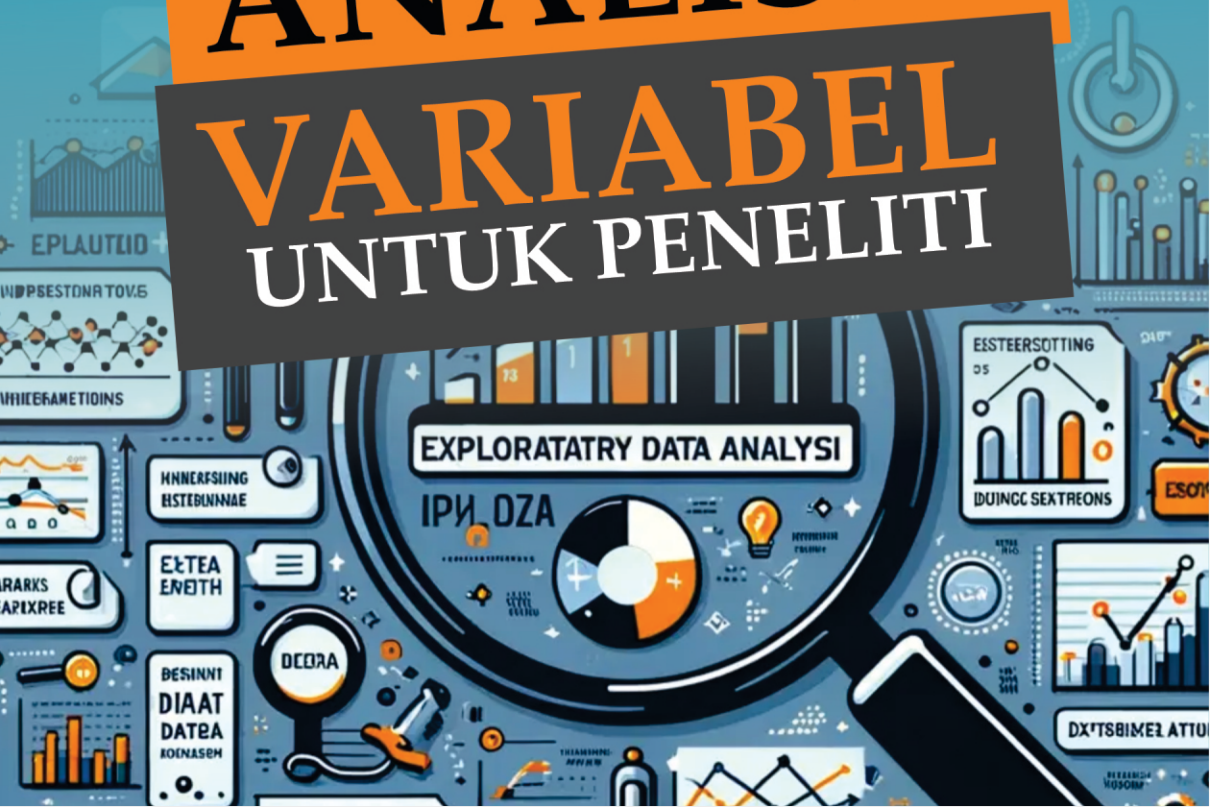
Editor:

Nanny Mayasari, S.Pd., M.Pd.
Qomarotun Nurlaila, S.T., M.T.

Aisyah Mutia Dawis, Rulyanti Susi Wardhani,
Hartono Nurlette, Ritha Widyaprawati,
Lisa Astria Milasari, Mohamad Zaenudin,
Inayatul Inayah, Deny Haryadi, Yoana Nurul Asri,
Nurul Kholisatul 'Ulya

PANDUAN PRAKTIS ANALISIS

VARIABEL UNTUK PENELITI



PANDUAN PRAKTIS ANALISIS VARIABEL UNTUK PENELITI

Penulis

Aisyah Mutia Dawis, Rulyanti Susi Wardhani, Hartono Nurlette,
Ritha Widyaprawati, Lisa Astria Milasari, Mohamad Zaenudin,
Inayatul Inayah, Deny Haryadi, Yoana Nurul Asri, Nurul
Kholisatul 'Ulya

Editor :

Nanny Mayasari, S.Pd., M.Pd.
Qomarotun Nurlaila, S.T., M.T.

Penerbit

TOHAR MEDIA

Panduan Praktis Analisis Variabel untuk Peneliti

Penulis :

Aisyah Mutia Dawis, Rulyanti Susi Wardhani, Hartono Nurlette,
Ritha Widyaprawati, Lisa Astria Milasari, Mohamad Zaenudin,
Inayatul Inayah, Deny Haryadi, Yoana Nurul Asri, Nurul
Kholisatul 'Ulya

Editor:

Nanny Mayasari, S.Pd., M.Pd.
Qomarotun Nurlaila, S.T., M.T.

ISBN: 978-623-8705-31-3

Desain Sampul dan Tata Letak

Ai Siti Khairunisa

Penerbit

CV. Tohar Media

Anggota IKAPI No. 022/SSL/2019

Redaksi :

JL. Rappocini Raya Lr 11 No 13 Makassar

JL. Hamzah dg. Tompo. Perumahan Nayla Regency Blok D No.25 Gowa
Telp. 0852-9999-3635/0852-4352-7215

Email : toharmedia@gmail.com

Website : <https://toharmedia.co.id>

Cetakan Pertama Agustus 2024

Hak Cipta dilindungi undang-undang. Dilarang memperbanyak sebagian atau seluruh isi buku ini dalam bentuk apapun, baik secara elektronik maupun mekanik termasuk memfotocopy, merekam atau dengan menggunakan sistem penyimpanan lainnya, tanpa izin tertulis dari penerbit.

Undang-undang Nomor 19 Tahun 2002 Tentang Hak Cipta

1. Barang siapa dengan sengaja dan tanpa hak mengumumkan atau memperbanyak suatu ciptaan atau memberi izin untuk itu, dipidana dengan pidana penjara paling lama 7 (Tujuh) tahun dan/atau denda paling banyak **Rp. 5.000.000.000,00 (Lima Miliar Rupiah)**
2. Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta atau hak terkait sebagaimana dimaksud pada ayat 1, dipidana paling lama 5 (lima tahun) dan/atau denda paling banyak **Rp. 500.000.000,00 (Lima Ratus Juta Rupiah).**

Selamat datang di buku dengan judul "**Panduan Praktis Analisis Variabel untuk Peneliti.**" Buku ini dirancang untuk memberikan **Referensi** pemahaman menyeluruh tentang teknik analisis variabel yang esensial dalam penelitian ilmiah. Baik Anda seorang peneliti, mahasiswa, atau profesional, buku ini menawarkan panduan praktis untuk menganalisis data dengan akurat dan efektif.

Kami menyajikan konsep dasar analisis variabel serta teknik-teknik penting, termasuk analisis univariant, bivariant, dan multivariant. Selain itu, buku ini mengeksplorasi penerapan big data, machine learning, dan data mining, serta menyediakan panduan dalam pengembangan instrumen penelitian yang efektif dan etis. Kami berharap buku ini menjadi sumber daya yang bermanfaat dalam meningkatkan keterampilan analisis Anda.

Selamat membaca dan semoga buku ini membantu Anda mencapai kesuksesan dalam penelitian.

Salam hangat,

Penulis

DAFTAR ISI



Halaman Depan	_i
Halaman Penerbit	_ii
Kata Pengantar	_iii
Daftar Isi	_iv
Bab 1. Konsep dasar Analisis Variabel	_1
1.1. Pengertian dan Klasifikasi Variabel	_1
1.2. Operasionalisasi Variabel	_6
1.3. Validitas dan Reliabilitas Variabel	_8
Bab 2. Metode Pengumpulan Data	_15
2.1. Definisi Metode Pengumpulan Data	_15
2.2. Definisi dan Pentingnya Pengumpulan Data	_16
2.3. Jenis -Jenis Data	_22
2.4. Data Primer dan Data sekunder	_27
2.5. Proses Pengumpulan Data	_28
2.6. Metode Pengumpulan Data Kualitatif	_30
2.7. Metode Pengumpulan Data Kuantitatif	_50
2.8. Pengukuran dan Skala	_55
2.9. Sumber Data Sekunder	_57
2.10. Etika Pengumpulan Data	_58
Bab 3. Teknik Analisis Data dengan Software Statistik	_61
3.1. Analisis Data	_61
3.2. Mengenal SPSS	_62
3.3. Menyiapkan Data File di SPSS	_66
3.4. Pengenalan Output pada SPSS	_71
Bab 4. Srtudi Univariat	_73
4.1. Univariat	_73
4.2. Peringkasan	_73
4.3. Bentuk Data yang Didistribusikan	_76
4.4. Teori Univariat	_80

Bab 5. Analisis Bivariat _87

- 5.1. Bivariat _87
- 5.2. Pengertian Analisis Bivariat _89
- 5.3. Hubungan asimetris pada Analisis Bivariat _90
- 5.4. Jenis Pengukuran Penelitian Analisis Bivariat _92
- 5.5. Metode Pengujian Bivariat Untuk Hipotesis Penelitian _97

Bab 6. Analisis Multivariat, dengan “v” _101

- 6.1. Multivariat _101
- 6.2. Dasar-dasar Statistik untuk Analisis Multivariat _103
- 6.3. Analisis Faktor (*Factor Analysis*) _109
- 6.4. Analisis Komponen Utama (*Principal Component Analysis – PCA*) _115
- 6.5. Regresi Berganda (*Multiple Regression*) _120
- 6.6. Model Persamaan Struktural (*Structural Equation Modeling - SEM*) _125

Bab 7. Penggunaan *Big Data* dan Analitik Lanjutan _131

- 7.1. Pengantar *Big Data* _131
- 7.2. Sumber dan Jenis Data _133
- 7.3. Teknologi dan Infrastruktur *Big Data* _135
- 7.4. Metodologi Analitik Lanjutan _137
- 7.5. Proses dan Tahapan Analisis *Big Data* _140
- 7.6. Masa Depan *Big Data* dan Analitik Lanjutan _143
- 7.7. Penutup _146

Bab 8. Penerapan dalam Analisis Variabel: Machine Learning dan Data Mining _147

- 8.1. Pemilihan Fitur (*Feature Selection*) _147
- 8.2. Ekstraksi Fitur (*Feature Extraction*) _149
- 8.3. Penerapan Algoritma Machine Learning _152
- 8.4. Data Mining _154
- 8.5. Evaluasi Model _158

Bab 9. Pengembangan Instrumen Penelitian _165

9.1. Pendahuluan _165

9.2. Syarat Instrumen Yang Baik _167

9.3. Langkah-Langkah Pengembangan Instrumen _171

9.4. Cara Memvalidasi Instrumen _173

Bab 10. Etika dalam Penelitian dan Analisis Data _177

10.1. Instrumen Penelitian _177

10.2. Prinsip -Prinsip Etika Penelitian _178

10.3. Pelanggaran Etika Penelitian dan Analisis
Data _186

10.4. Praktik Terbaik dalam Etika Penelitian _189

Daftar Pustaka _192

Panduan Praktis Analisis Variabel untuk Peneliti

Bab 1

Konsep Dasar Analisis Variabel

1.1 Pengertian dan Klasifikasi Variabel

Dalam penelitian, variabel adalah konsep inti yang mewakili karakteristik atau atribut yang dapat diukur, diamati, atau dimanipulasi. Variabel memiliki nilai yang bervariasi antar individu, objek, atau fenomena yang diteliti. Misalnya, usia, jenis kelamin, tingkat pendidikan, pendapatan, atau tingkat kepuasan adalah contoh variabel yang umum digunakan dalam penelitian sosial.

Bayangkan Kita sedang mengamati sekelompok orang. Kita akan melihat bahwa mereka berbeda-beda, bukan? Ada yang tinggi, ada yang pendek. Ada yang berambut hitam, ada yang berambut pirang. Perbedaan-perbedaan inilah yang kita sebut sebagai variabel.

Dalam penelitian, variabel adalah segala sesuatu yang dapat berubah atau memiliki variasi (Kambuaya, Masinambow and Sumual, 2019). Usia, jenis kelamin, tingkat pendidikan, pendapatan, bahkan preferensi musik seseorang, semuanya adalah contoh variabel. Mengapa variabel penting? Karena dengan mempelajari variabel, kita bisa memahami dunia di sekitar kita dengan lebih baik.

Kita bisa membagi variabel menjadi beberapa jenis. Ada variabel yang kita yakini sebagai penyebab, yang disebut variabel independen. Ada juga variabel yang kita anggap sebagai akibat, yang disebut variabel dependen. Misalnya, dalam penelitian tentang pengaruh kebiasaan merokok terhadap kesehatan, kebiasaan merokok adalah variabel independen, sedangkan kesehatan adalah variabel dependen.

Memahami variabel adalah langkah pertama yang krusial dalam melakukan analisis variabel. Dengan memahami jenis-jenis variabel dan bagaimana mereka berinteraksi, kita bisa menggali lebih dalam pola-pola yang tersembunyi dan menemukan jawaban atas pertanyaan-pertanyaan penelitian kita (Elise, 2019).

Variabel dapat diklasifikasikan menjadi beberapa jenis, antara lain:

1. Variabel Independen (Bebas)

Dalam penelitian, variabel adalah segala sesuatu yang dapat berubah atau memiliki variasi. Usia, jenis kelamin, tingkat pendidikan, pendapatan, bahkan preferensi musik seseorang, semuanya adalah contoh variabel. Mengapa variabel penting? Karena dengan mempelajari variabel, kita bisa memahami dunia di sekitar kita dengan lebih baik.

Kita bisa membagi variabel menjadi beberapa jenis. Salah satu yang paling penting adalah variabel independen (Ramírez, Wilson and Soederberg Miller, 2022), yang sering juga disebut variabel bebas. Variabel independen adalah variabel yang kita yakini sebagai penyebab atau yang mempengaruhi variabel lain. Misalnya, dalam penelitian tentang pengaruh kebiasaan merokok terhadap kesehatan, kebiasaan merokok adalah variabel independen (Weigand *et al.*, 2022). Kita menduga bahwa kebiasaan merokok inilah yang menyebabkan perubahan pada variabel lain, yaitu kesehatan.

Variabel independen bisa kita manipulasi atau atur dalam sebuah eksperimen. Misalnya, dalam penelitian tentang pengaruh jenis pupuk terhadap pertumbuhan tanaman, jenis pupuk adalah variabel independen yang bisa kita ubah-ubah (Bustos *et al.*, 2022). Dengan membandingkan pertumbuhan tanaman yang diberi pupuk berbeda, kita bisa melihat bagaimana variabel independen (jenis pupuk) mempengaruhi variabel dependen (pertumbuhan tanaman).

Memahami variabel independen sangat penting dalam merancang penelitian (Apaydin Cirik, Gül and Aksoy, 2022). Dengan mengidentifikasi variabel independen yang tepat, kita bisa menguji hipotesis kita dan menemukan jawaban atas pertanyaan-pertanyaan penelitian kita.

2. Variabel Dependen (Terikat)

Variabel dependen, atau sering disebut juga sebagai variabel terikat, adalah variabel yang menjadi pusat perhatian utama dalam penelitian kita (Romaní-Romaní, Gutiérrez and Azurin-Salazar, 2023). Variabel ini adalah hasil atau efek yang kita amati dan ukur, yang diduga dipengaruhi oleh variabel lain yang kita sebut variabel independen. Sederhananya, variabel dependen adalah "apa yang terjadi" sebagai respons terhadap perubahan atau manipulasi yang kita lakukan pada variabel independen.

Penting untuk diingat bahwa variabel dependen selalu bergantung pada variabel independen. Perubahan pada variabel independen akan menghasilkan perubahan pada variabel dependen (Knight *et al.*, 2022). Misalnya, dalam penelitian tentang efektivitas obat baru terhadap tekanan darah, tekanan darah adalah variabel dependen, sedangkan obat baru adalah variabel independen.

Dalam analisis variabel, kita berusaha untuk memahami bagaimana variabel independen mempengaruhi variabel dependen. Kita ingin mengetahui apakah ada hubungan sebab-akibat antara keduanya, dan jika ada, seberapa kuat hubungan tersebut. Dengan memahami hubungan ini, kita dapat membuat prediksi tentang apa yang akan terjadi pada variabel dependen jika kita mengubah variabel independen.

3. Variabel Moderator

Dalam penelitian, variabel adalah segala sesuatu yang dapat berubah atau memiliki variasi. Usia, jenis kelamin, tingkat pendidikan, pendapatan, bahkan preferensi musik seseorang, semuanya adalah contoh variabel. Mengapa variabel penting? Karena dengan mempelajari variabel, kita bisa memahami dunia di sekitar kita dengan lebih baik.

Kita bisa membagi variabel menjadi beberapa jenis. Ada variabel yang kita yakini sebagai penyebab, yang disebut variabel independen. Ada juga variabel yang kita anggap sebagai akibat, yang disebut variabel dependen. Misalnya, dalam penelitian tentang pengaruh kebiasaan merokok terhadap kesehatan, kebiasaan merokok adalah variabel independen, sedangkan kesehatan adalah variabel dependen.

Selain itu, ada juga variabel moderator, yaitu variabel yang dapat mempengaruhi hubungan antara variabel independen dan dependen. Variabel moderator seperti "rem" atau "gas" dalam hubungan antara dua variabel lainnya. Misalnya, dalam penelitian tentang pengaruh kebiasaan merokok terhadap kesehatan, tingkat stres bisa menjadi variabel moderator (Elise, 2019). Pada orang yang memiliki tingkat stres tinggi, efek negatif dari merokok terhadap kesehatan mungkin akan lebih kuat dibandingkan dengan orang yang memiliki tingkat stres rendah.

Memahami variabel, termasuk variabel moderator, adalah langkah pertama yang krusial dalam melakukan analisis variabel. Dengan memahami jenis-jenis variabel dan bagaimana mereka berinteraksi (Saragih, 2019), kita bisa menggali lebih dalam pola-pola yang tersembunyi dan menemukan jawaban atas pertanyaan-pertanyaan penelitian kita.

4. Variabel Mediator

Kita bisa membagi variabel menjadi beberapa jenis. Ada variabel yang kita yakini sebagai penyebab, yang disebut variabel independen. Ada juga variabel yang kita anggap sebagai akibat, yang disebut variabel dependen. Misalnya, dalam penelitian tentang pengaruh kebiasaan merokok terhadap kesehatan, kebiasaan merokok adalah variabel independen, sedangkan kesehatan adalah variabel dependen.

Namun, terkadang hubungan antara variabel independen dan dependen tidak sesederhana itu. Ada kalanya, ada variabel lain yang berperan sebagai perantara, yang disebut variabel mediator (Bauer *et al.*, 2022). Variabel mediator ini menjelaskan bagaimana atau mengapa variabel independen mempengaruhi variabel dependen (Nuryono *et al.*, 2019). Misalnya, dalam penelitian tentang pengaruh tingkat stres terhadap kinerja kerja, kita mungkin menemukan bahwa tingkat stres mempengaruhi kinerja kerja melalui variabel mediator seperti kualitas tidur atau motivasi kerja.

Memahami variabel, termasuk variabel mediator, adalah langkah pertama yang krusial dalam melakukan analisis variabel. Dengan memahami jenis-jenis variabel dan bagaimana mereka berinteraksi, kita bisa menggali lebih dalam pola-pola yang tersembunyi dan menemukan jawaban atas pertanyaan-pertanyaan penelitian kita.

5. Variabel Kontrol

Dalam penelitian, variabel kontrol adalah faktor-faktor yang sengaja dijaga konstan atau dibatasi pengaruhnya. Tujuannya adalah untuk memastikan bahwa perubahan yang terjadi pada variabel dependen (hasil yang diamati) benar-benar disebabkan oleh variabel independen (faktor yang dimanipulasi), bukan oleh faktor lain yang tidak diinginkan.

Bayangkan Kita ingin meneliti pengaruh jenis pupuk terhadap pertumbuhan tanaman. Dalam hal ini, jenis pupuk adalah variabel independen (Putra *et al.*, 2020), sedangkan pertumbuhan tanaman adalah variabel dependen. Namun, ada banyak faktor lain yang juga bisa mempengaruhi pertumbuhan tanaman, seperti jumlah air, intensitas cahaya matahari, jenis tanah, dan suhu udara. Faktor-faktor inilah yang perlu Kita kontrol agar Kita bisa yakin bahwa perbedaan pertumbuhan tanaman benar-benar disebabkan oleh perbedaan jenis pupuk, bukan oleh faktor lain.

Variabel kontrol sangat penting dalam penelitian karena membantu meningkatkan validitas internal penelitian, yaitu sejauh mana kita bisa yakin bahwa hubungan sebab-akibat yang kita temukan benar-benar valid.

1.2 Operasionalisasi Variabel

Anggaplah penelitian itu seperti sebuah resep masakan. Kita memiliki bahan-bahan mentah (konsep-konsep abstrak), tetapi Kita perlu mengubahnya menjadi sesuatu yang bisa diukur dan diolah (variabel yang terukur). Di sinilah operasionalisasi variabel berperan.

Operasionalisasi variabel adalah proses "menerjemahkan" konsep abstrak menjadi indikator yang konkret dan terukur. Misalnya, Kita ingin meneliti "kinerja karyawan". Konsep ini terlalu umum dan abstrak. Kita perlu mengoperasionalkannya

menjadi indikator yang lebih spesifik, seperti jumlah penjualan per bulan, tingkat kehadiran, atau penilaian kinerja dari atasan.

Proses operasionalisasi melibatkan beberapa langkah. Pertama, Kita perlu mengidentifikasi konsep yang ingin Kita teliti (Alimy *et al.*, 2020). Kemudian, Kita perlu memecah konsep tersebut menjadi dimensi-dimensi yang lebih kecil. Misalnya, "kinerja karyawan" bisa dipecah menjadi dimensi produktivitas, kualitas kerja, dan inisiatif. Setelah itu, Kita perlu memilih indikator yang sesuai untuk setiap dimensi.

Operasionalisasi variabel adalah langkah penting dalam penelitian karena memastikan bahwa variabel yang Kita teliti dapat diukur secara akurat dan konsisten. Dengan demikian, hasil penelitian Kita akan lebih valid dan reliabel.

Langkah-langkah Operasionalisasi Variabel:

1. Identifikasi Konsep

Langkah pertama adalah mengidentifikasi konsep abstrak yang ingin Kita teliti. Konsep ini biasanya merupakan ide umum yang ingin Kita pahami lebih dalam. Contohnya, "kepuasan pelanggan", "motivasi kerja", atau "kualitas hidup".

2. Dekomposisi Konsep

Setelah mengidentifikasi konsep, langkah selanjutnya adalah memecah konsep tersebut menjadi dimensi-dimensi yang lebih spesifik dan terukur. Dimensi adalah aspek-aspek yang membentuk konsep utama. Misalnya, "kepuasan pelanggan" dapat dipecah menjadi dimensi kepuasan terhadap produk, layanan, dan harga.

3. Pemilihan Indikator

Setelah mengidentifikasi dimensi-dimensi konsep, langkah selanjutnya adalah memilih indikator yang sesuai untuk mengukur setiap dimensi. Indikator adalah variabel yang dapat diamati dan diukur yang mencerminkan dimensi

konsep (Rahmawati, Umkabu and Lestari, 2021). Misalnya, indikator untuk dimensi "kepuasan terhadap produk" bisa berupa "kesesuaian produk dengan harapan", "kekitalan produk", atau "daya tahan produk".

4. Penentuan Skala Pengukuran

Langkah terakhir adalah menentukan skala pengukuran yang sesuai untuk setiap indikator. Skala pengukuran menentukan jenis data yang akan dihasilkan oleh indikator tersebut (Mulyani and Mulyani, 2021). Ada empat jenis skala pengukuran: nominal, ordinal, interval, dan rasio. Pemilihan skala pengukuran yang tepat sangat penting karena akan mempengaruhi jenis analisis statistik yang dapat Kita lakukan.

Dengan mengikuti langkah-langkah ini, Kita dapat mengoperasionalkan variabel-variabel dalam penelitian Kita secara sistematis dan terukur. Hal ini akan membantu memastikan bahwa penelitian Kita valid dan reliabel.

1.3 Validitas dan Reliabilitas Variabel

Validitas dan reliabilitas adalah dua konsep penting dalam pengukuran variabel penelitian. Validitas mengacu pada sejauh mana suatu instrumen penelitian (misalnya, kuesioner, tes, atau observasi) benar-benar mengukur apa yang seharusnya diukur. Dengan kata lain, validitas menunjukkan keakuratan dan ketepatan suatu instrumen dalam menangkap konsep yang ingin diukur.

Ada beberapa jenis validitas, termasuk validitas isi (sejauh mana instrumen mencakup semua aspek konsep yang diukur), validitas konstruk (sejauh mana instrumen mengukur konsep teoritis yang mendasarinya), dan validitas kriteria (sejauh mana skor pada instrumen berkorelasi dengan kriteria eksternal yang relevan).

Sementara itu, reliabilitas mengacu pada konsistensi dan stabilitas hasil pengukuran suatu instrumen (Rosmawati, 2022). Instrumen yang reliabel akan memberikan hasil yang konsisten jika digunakan berulang kali pada subjek yang sama atau pada kondisi yang sama. Reliabilitas dapat diuji dengan berbagai metode, seperti uji-ulang (*test-retest*), bentuk paralel (*parallel forms*), dan konsistensi internal (*internal consistency*).

Validitas dan reliabilitas saling terkait. Instrumen yang tidak valid tidak mungkin reliabel, karena instrumen tersebut tidak mengukur apa yang seharusnya diukur. Namun, instrumen yang reliabel belum tentu valid. Instrumen bisa saja memberikan hasil yang konsisten, tetapi hasil tersebut mungkin tidak akurat atau tepat dalam mengukur konsep yang diinginkan. Oleh karena itu, penting untuk memastikan bahwa instrumen penelitian yang digunakan memiliki validitas dan reliabilitas yang tinggi.

Jenis-jenis validitas:

1. Validitas Isi (*Content Validity*)

Validitas isi mengukur sejauh mana item-item atau pertanyaan dalam instrumen penelitian mencakup seluruh aspek dari konsep yang ingin diukur. Misalnya, jika Kita ingin mengukur tingkat depresi seseorang, instrumen yang Kita gunakan harus mencakup semua gejala utama depresi, seperti perasaan sedih, kehilangan minat, gangguan tidur, dan perubahan nafsu makan. Validitas isi biasanya dinilai oleh para ahli di bidang yang relevan.

2. Validitas Konstruk (*Construct Validity*)

Validitas konstruk mengukur sejauh mana instrumen penelitian benar-benar mengukur konsep teoritis yang mendasarinya. Konsep teoritis adalah ide abstrak yang tidak dapat diamati secara langsung, seperti kecerdasan, motivasi, atau kepribadian. Validitas konstruk dapat dinilai melalui

berbagai cara, seperti membandingkan skor pada instrumen dengan skor pada instrumen lain yang mengukur konsep yang sama, atau dengan melihat apakah skor pada instrumen sesuai dengan prediksi teori.

3. Validitas Kriteria (*Criterion Validity*)

Validitas kriteria mengukur sejauh mana skor pada instrumen penelitian berkorelasi dengan kriteria eksternal yang relevan. Kriteria eksternal adalah ukuran lain dari konsep yang sama yang dianggap valid dan reliabel. Misalnya, jika Kita ingin mengukur kemampuan akademik siswa, Kita dapat membandingkan skor pada tes yang Kita kembangkan dengan nilai rapor siswa. Jika skor pada tes Kita berkorelasi tinggi dengan nilai rapor, maka tes Kita memiliki validitas kriteria yang tinggi.

Dengan memahami jenis-jenis validitas ini, Kita dapat memilih dan mengembangkan instrumen penelitian yang tepat untuk mengukur variabel yang Kita minati. Instrumen yang valid akan memberikan hasil yang akurat dan dapat dikitalkan, yang akan membantu Kita menarik kesimpulan yang tepat dari penelitian Kita.

Apa itu Reliabilitas?

Reliabilitas, dalam konteks penelitian, adalah tingkat konsistensi dan stabilitas suatu instrumen pengukuran. Instrumen yang reliabel akan memberikan hasil yang relatif sama ketika digunakan berulang kali pada subjek yang sama atau pada kondisi yang sama.

Bayangkan sebuah timbangan badan. Jika timbangan tersebut reliabel, maka berat badan Kita akan terukur sama setiap kali Kita menimbang diri, selama berat badan Kita tidak berubah. Sebaliknya, jika timbangan tersebut tidak reliabel, maka berat badan Kita akan terukur berbeda-beda setiap kali Kita

menimbang diri, meskipun berat badan Kita sebenarnya tetap sama.

Reliabilitas sangat penting dalam penelitian karena hasil penelitian yang tidak reliabel tidak dapat dipercaya. Jika instrumen pengukuran yang digunakan tidak reliabel, maka hasil penelitian tersebut tidak dapat digeneralisasikan ke populasi yang lebih luas (Indriani *et al.*, 2023).

Berikut penjelasan mengenai metode pengujian reliabilitas:

1. Uji-Ulang (*Test-Retest Reliability*)

Metode ini melibatkan pemberian instrumen yang sama kepada responden yang sama pada dua waktu yang berbeda. Interval waktu antara pengujian pertama dan kedua harus cukup lama untuk menghindari efek memori, tetapi tidak terlalu lama sehingga karakteristik yang diukur dapat berubah. Reliabilitas dihitung dengan mengkorelasikan skor dari kedua pengujian. Jika korelasi tinggi, maka instrumen dianggap reliabel.

Kelebihan: Sederhana dan mudah diterapkan.

Kekurangan: Rentan terhadap efek memori dan perubahan karakteristik responden dari waktu ke waktu.

2. Bentuk Paralel (*Parallel Forms Reliability*)

Metode ini melibatkan penggunaan dua bentuk instrumen yang setara (memiliki item yang berbeda tetapi mengukur konstruk yang sama) pada sampel yang sama. Reliabilitas dihitung dengan mengkorelasikan skor dari kedua bentuk instrumen. Jika korelasi tinggi, maka instrumen dianggap reliabel.

Kelebihan: Menghindari efek memori, lebih akurat dalam mengukur reliabilitas.

Kekurangan: Sulit mengembangkan dua bentuk instrumen yang benar-benar setara.

3. Konsistensi Internal (*Internal Consistency Reliability*)

Metode ini mengukur sejauh mana item-item dalam suatu instrumen saling berkorelasi. Ada beberapa teknik untuk mengukur konsistensi internal, yang paling umum adalah alpha Cronbach. Semakin tinggi nilai alpha Cronbach (mendekati 1), semakin tinggi reliabilitas instrumen.

Kelebihan: Hanya memerlukan satu kali pengujian, mudah diterapkan.

Kekurangan: Tidak memperhitungkan stabilitas hasil pengukuran dari waktu ke waktu.

Pemahaman yang mendalam tentang variabel adalah fondasi yang kokoh bagi setiap penelitian ilmiah. Variabel berperan sebagai alat yang memungkinkan peneliti untuk mengamati, mengukur, dan menganalisis fenomena yang kompleks. Dalam penelitian sosial, variabel seperti usia, jenis kelamin, tingkat pendidikan, dan pendapatan membantu kita memahami perbedaan dan kesamaan di antara individu dan kelompok.

Klasifikasi variabel menjadi variabel independen, dependen, moderator, mediator, dan kontrol memberikan kerangka kerja yang jelas bagi peneliti untuk merancang studi yang komprehensif dan valid. Variabel independen dan dependen membentuk inti dari analisis sebab-akibat, sedangkan variabel moderator dan mediator membantu mengungkap hubungan yang lebih kompleks dan tersembunyi. Variabel kontrol, di sisi lain, memastikan bahwa hasil yang diperoleh adalah akibat dari manipulasi yang dilakukan, bukan faktor eksternal lainnya.

Operasionalisasi variabel adalah langkah krusial yang menerjemahkan konsep abstrak menjadi indikator yang dapat diukur. Proses ini memastikan bahwa penelitian dapat diulang dan hasilnya dapat dikitalkan. Dalam tahap ini, peneliti harus

berhati-hati memilih indikator yang tepat dan skala pengukuran yang sesuai untuk mendapatkan data yang valid dan reliabel.

Validitas dan reliabilitas adalah dua pilar utama dalam evaluasi instrumen penelitian. Validitas memastikan bahwa instrumen benar-benar mengukur apa yang seharusnya diukur, sementara reliabilitas menjamin konsistensi hasil pengukuran. Pengujian validitas dan reliabilitas yang cermat memberikan kepercayaan bahwa data yang dikumpulkan dapat digunakan untuk menarik kesimpulan yang akurat dan bermakna.

Secara keseluruhan, analisis variabel adalah seni dan sains yang memerlukan keahlian, ketelitian, dan integritas. Dengan memahami dan menerapkan prinsip-prinsip ini, peneliti dapat mengungkap wawasan baru, menguji teori, dan memberikan kontribusi berharga bagi pengetahuan ilmiah dan praktik profesional. Kesuksesan penelitian bergantung pada kemampuan untuk mengelola variabel dengan bijaksana, memastikan bahwa setiap langkah penelitian dilaksanakan dengan metode yang paling tepat dan akurat. Dengan demikian, kita dapat mencapai hasil yang tidak hanya valid dan reliabel, tetapi juga relevan dan berdampak bagi masyarakat luas.

Bab 2

Metode Pengumpulan Data

2.1 Definisi Metode Pengumpulan Data

Metode pengumpulan data adalah proses sistematis untuk mengumpulkan informasi atau fakta-fakta yang relevan untuk tujuan penelitian atau analisis. Pengumpulan data ini penting karena data merupakan landasan utama dalam menghasilkan informasi yang akurat dan berguna dalam berbagai bidang seperti ilmu pengetahuan, bisnis, pemerintahan, dan lain sebagainya. Metode pengumpulan data mengacu pada cara atau teknik yang digunakan untuk mendapatkan informasi atau data yang dibutuhkan untuk tujuan penelitian. Proses ini melibatkan berbagai langkah, mulai dari perencanaan dan desain instrumen pengumpulan data, hingga pengumpulan, pengolahan, analisis, dan interpretasi data. Metode ini dapat bervariasi tergantung pada jenis penelitian, tujuan penelitian, sumber daya yang tersedia, dan karakteristik populasi atau objek yang diteliti.

Tujuan utama dari metode pengumpulan data adalah untuk:

1. Mendapatkan Informasi yang Diperlukan: Memperoleh data yang relevan dan diperlukan untuk menjawab pertanyaan penelitian atau mencapai tujuan penelitian yang telah ditetapkan.
2. Memastikan Validitas dan Reliabilitas: Memastikan bahwa data yang dikumpulkan akurat, valid, dan dapat dipercaya untuk digunakan dalam analisis lebih lanjut.

3. Memberikan Dasar untuk Analisis: Menyediakan dasar yang kuat untuk melakukan analisis statistik, kualitatif, atau interpretatif terhadap data yang diperoleh.

2.2 Definisi dan Pentingnya Pengumpulan Data

2.2.1 Pengertian Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mendapatkan, mengukur, dan menganalisis informasi yang relevan dari berbagai sumber untuk menjawab pertanyaan penelitian, menguji hipotesis, dan mengevaluasi hasil (Creswell, W. John dan Creswell, 2018). Proses ini mencakup berbagai metode dan teknik yang digunakan untuk mengumpulkan data kualitatif dan kuantitatif, yang nantinya akan diolah dan dianalisis untuk menghasilkan pengetahuan baru atau memahami fenomena tertentu. Pengumpulan data merupakan langkah krusial dalam penelitian karena data yang dikumpulkan akan menjadi dasar dari seluruh analisis dan kesimpulan penelitian. Adanya data yang akurat dan relevan, peneliti dapat membuat keputusan yang tepat, menyusun strategi, dan memberikan rekomendasi yang berdasarkan bukti.

2.2.2 Pentingnya Data Dalam Penelitian

Data adalah fondasi dari setiap penelitian ilmiah. Tanpa data yang akurat dan relevan, proses penelitian akan menjadi tidak berarti, karena data merupakan basis dari pengumpulan informasi yang digunakan untuk menjawab pertanyaan penelitian, menguji hipotesis, dan membuat kesimpulan. Pengumpulan data yang tepat sangat penting dalam memastikan validitas dan reliabilitas hasil penelitian, yang pada gilirannya mempengaruhi penerimaan dan penerapan temuan penelitian di dunia nyata.

1. Data sebagai Dasar Keputusan

Data memberikan dasar yang kuat untuk membuat keputusan yang berbasis bukti. Dalam konteks penelitian, data yang dikumpulkan melalui metode yang sistematis memungkinkan peneliti untuk memahami fenomena, mengidentifikasi pola, dan mengevaluasi hubungan antara variabel. Sebagai contoh, dalam penelitian medis, data klinis yang dikumpulkan dari pasien dapat membantu dokter menentukan efektivitas pengobatan dan membuat rekomendasi medis yang tepat (Creswell, W. John dan Creswell, 2018). Data adalah fondasi dari setiap penelitian ilmiah. Tanpa data yang akurat dan relevan, proses penelitian akan menjadi tidak berarti, karena data merupakan basis dari pengumpulan informasi yang digunakan untuk menjawab pertanyaan penelitian, menguji hipotesis, dan membuat kesimpulan. Pengumpulan data yang tepat sangat penting dalam memastikan validitas dan reliabilitas hasil penelitian, yang pada gilirannya mempengaruhi penerimaan dan penerapan temuan penelitian di dunia nyata.

2. Validitas dan Reliabilitas Penelitian

Validitas dan reliabilitas merupakan dua konsep kunci dalam penelitian yang sangat bergantung pada data. Validitas merujuk pada sejauh mana alat ukur mengukur apa yang seharusnya diukur, sedangkan reliabilitas berkaitan dengan konsistensi hasil yang diperoleh dari alat ukur yang sama dalam kondisi yang sama. Data yang dikumpulkan dengan cara yang tidak tepat dapat merusak validitas dan reliabilitas penelitian, yang pada akhirnya akan mempengaruhi kredibilitas hasil penelitian (Heriyanto, 2022).

3. Pengembangan Teori dan Model

Data adalah bahan baku utama dalam pengembangan teori dan model ilmiah. Dalam ilmu sosial, data dari survei dan wawancara digunakan untuk mengembangkan dan menguji

teori tentang perilaku manusia dan interaksi sosial. Dalam ilmu alam, data eksperimental digunakan untuk mengembangkan model yang menjelaskan fenomena alam dan memprediksi hasil dari kondisi yang berbeda. Tanpa data, teori dan model ini tidak akan memiliki dasar empiris yang kuat dan tidak dapat diuji atau divalidasi.

4. Identifikasi Pola dan Tren

Salah satu manfaat utama dari pengumpulan data adalah kemampuan untuk mengidentifikasi pola dan tren. Dalam bisnis, data penjualan dan perilaku konsumen digunakan untuk mengidentifikasi tren pasar dan membuat strategi pemasaran yang efektif. Dalam kesehatan masyarakat, data epidemiologi digunakan untuk mengidentifikasi tren penyakit dan merencanakan intervensi kesehatan yang tepat. Pola dan tren ini hanya dapat ditemukan melalui analisis data yang cermat dan menyeluruh (Silverman, 2017).

5. Evaluasi dan Pengukuran Kinerja

Data memungkinkan evaluasi yang objektif dan pengukuran kinerja. Dalam pendidikan, data tentang prestasi siswa digunakan untuk mengevaluasi efektivitas kurikulum dan metode pengajaran. Dalam sektor publik, data tentang kinerja program pemerintah digunakan untuk mengevaluasi dampak program tersebut dan membuat keputusan tentang perbaikan atau pengembangan lebih lanjut. Evaluasi dan pengukuran kinerja yang didasarkan pada data memberikan informasi yang akurat dan terpercaya untuk pengambilan keputusan yang efektif (Patton, 2014; Kriyantono, 2020)

6. Pengurangan Ketidakpastian

Salah satu tujuan utama pengumpulan data adalah untuk mengurangi ketidakpastian dalam penelitian dan pengambilan keputusan. Ketidakpastian dapat timbul dari berbagai sumber, termasuk variabilitas dalam data,

ketidakpastian tentang hubungan antara variabel, dan ketidakpastian tentang hasil dari intervensi atau perubahan kondisi. Dengan mengumpulkan data yang relevan dan akurat, peneliti dapat mengurangi ketidakpastian ini dan membuat prediksi yang lebih akurat serta keputusan yang lebih informatif (Barbrook-Johnson dan Carrick, 2022)

7. Manajemen Risiko

Data juga penting dalam manajemen risiko. Dalam sektor keuangan, data tentang pasar dan kinerja investasi digunakan untuk mengidentifikasi dan mengelola risiko investasi. Dalam manajemen proyek, data tentang kemajuan proyek dan masalah yang dihadapi digunakan untuk mengidentifikasi risiko dan mengambil tindakan mitigasi yang diperlukan. Dengan mengumpulkan dan menganalisis data secara sistematis, organisasi dapat mengidentifikasi potensi risiko lebih awal dan mengembangkan strategi untuk mengelola risiko tersebut (Botter, 1982)

8. Peningkatan Kualitas

Pengumpulan data yang tepat juga penting untuk peningkatan kualitas. Dalam manufaktur, data tentang proses produksi dan kualitas produk digunakan untuk mengidentifikasi area perbaikan dan mengembangkan strategi peningkatan kualitas. Dalam layanan kesehatan, data tentang hasil klinis dan pengalaman pasien digunakan untuk meningkatkan kualitas perawatan dan layanan yang diberikan. Dengan menggunakan data untuk mengidentifikasi masalah dan mengukur hasil, organisasi dapat terus menerus meningkatkan kualitas produk dan layanan mereka.

9. Inovasi dan Pengembangan

Data juga mendorong inovasi dan pengembangan. Dalam industri teknologi, data tentang penggunaan produk dan umpan balik pelanggan digunakan untuk mengembangkan fitur baru dan meningkatkan produk yang ada. Dalam penelitian ilmiah, data tentang hasil eksperimen dan pengamatan digunakan untuk mengembangkan teori dan teknologi baru. Inovasi yang didasarkan pada data memungkinkan organisasi untuk tetap kompetitif dan relevan di pasar yang terus berubah.

2.2.3 Tujuan Pengumpulan Data

Pengumpulan data adalah proses penting dalam penelitian yang memiliki beberapa tujuan utama. Tujuan-tujuan ini berfungsi sebagai panduan untuk memahami dan memanfaatkan data yang dikumpulkan secara efektif. Berikut adalah beberapa tujuan utama pengumpulan data, lengkap dengan sitasi yang relevan:

1. Memvalidasi Hipotesis Penelitian

Tujuan utama pengumpulan data adalah untuk memvalidasi hipotesis penelitian. Dengan mengumpulkan data yang relevan, peneliti dapat menguji apakah hipotesis yang diajukan dapat diterima atau harus ditolak berdasarkan bukti empiris yang diperoleh.

2. Mengembangkan Teori Baru

Pengumpulan data dapat digunakan untuk mengembangkan teori baru atau memodifikasi teori yang sudah ada. Data yang diperoleh dari penelitian dapat mengungkapkan pola atau hubungan baru yang belum pernah teridentifikasi sebelumnya, sehingga berkontribusi pada pengembangan ilmu pengetahuan.

3. Mengevaluasi Kinerja dan Efektivitas

Data digunakan untuk mengevaluasi kinerja dan efektivitas program, kebijakan, atau intervensi tertentu. Misalnya, dalam sektor kesehatan, data digunakan untuk menilai efektivitas program vaksinasi atau intervensi kesehatan lainnya.

4. Mengidentifikasi Masalah dan Kebutuhan

Pengumpulan data membantu mengidentifikasi masalah dan kebutuhan yang ada dalam masyarakat atau organisasi. Data yang dikumpulkan dapat menunjukkan area yang memerlukan perbaikan atau perhatian lebih lanjut.

5. Membuat Keputusan yang Berbasis Bukti

Pengumpulan data yang akurat dan relevan memungkinkan pembuat kebijakan, manajer, dan peneliti untuk membuat keputusan yang berbasis bukti. Keputusan yang didasarkan pada data cenderung lebih dapat diandalkan dan efektif dibandingkan dengan keputusan yang didasarkan pada asumsi atau intuisi semata.

6. Mengukur Variabel dan Mengidentifikasi Hubungan

Pengumpulan data memungkinkan peneliti untuk mengukur variabel secara akurat dan mengidentifikasi hubungan atau korelasi antara variabel tersebut. Hal ini penting untuk memahami dinamika dan mekanisme yang mendasari fenomena yang sedang diteliti.

7. Mengantisipasi dan Mengelola Risiko

Data yang dikumpulkan dapat digunakan untuk mengidentifikasi dan mengelola risiko. Dalam konteks manajemen proyek atau investasi, data tentang tren pasar atau kinerja proyek sebelumnya dapat membantu dalam mengantisipasi dan mengurangi risiko.

8. Meningkatkan Kualitas dan Efisiensi

Pengumpulan data yang sistematis dapat mengidentifikasi area di mana perbaikan diperlukan, sehingga membantu meningkatkan kualitas dan efisiensi proses atau produk. Misalnya, dalam manufaktur, data tentang cacat produk dapat digunakan untuk meningkatkan proses produksi.

9. Mendorong Inovasi

Data yang dikumpulkan dapat menginspirasi inovasi dengan mengidentifikasi peluang baru atau kebutuhan yang belum terpenuhi. Dalam bisnis, data tentang perilaku konsumen dan tren pasar dapat digunakan untuk mengembangkan produk atau layanan baru yang lebih sesuai dengan kebutuhan pasar.

10. Dokumentasi dan Transparansi

Pengumpulan data yang baik memastikan bahwa setiap langkah dalam proses penelitian didokumentasikan dengan baik, sehingga penelitian dapat diulangi dan diverifikasi oleh peneliti lain. Hal ini meningkatkan transparansi dan kredibilitas penelitian.

2.3 Jenis -Jenis Data

2.3.1 Data Kualitatif

Data kualitatif merujuk kepada informasi yang diperoleh dari deskripsi, kata-kata, gambar, atau objek non-angka lainnya. Jenis data ini sering digunakan dalam penelitian ilmu sosial, humaniora, dan bidang-bidang lain di mana peneliti tertarik untuk memahami fenomena secara mendalam dari perspektif yang lebih subjektif dan kompleks.

1. Karakteristik Data Kualitatif

- a. Deskriptif dan Tidak Terstruktur: Data kualitatif tidak dapat diukur secara numerik dan tidak mengikuti skala atau kategori tertentu. Sebaliknya, data ini cenderung berupa kata-kata, citra, suara, atau bentuk-bentuk lain

dari deskripsi yang memberikan pengalaman, pendapat, atau persepsi individu.

- b. Subjektif dan Kontekstual: Informasi dalam data kualitatif sering kali subjektif, karena mewakili pengalaman dan pandangan individu. Data ini juga sangat dipengaruhi oleh konteks sosial, budaya, dan historis di mana informasi tersebut diperoleh.
- c. Mendalam dan Kompleks: Data kualitatif memungkinkan peneliti untuk menjelajahi fenomena dengan cara yang mendalam dan kompleks. Ini memungkinkan untuk memahami nuansa, dinamika, dan konteks yang mungkin tidak tertangkap oleh metode penelitian kuantitatif.

2. Teknik Pengumpulan Data Kualitatif

- a. Wawancara: Melibatkan percakapan terstruktur atau tidak terstruktur antara peneliti dan peserta untuk memahami pandangan mereka.
- b. Observasi: Pengamatan langsung terhadap perilaku atau situasi untuk memahami konteks yang lebih baik.
- c. Studi Kasus: Pendekatan yang memungkinkan peneliti untuk memeriksa kasus tunggal atau beberapa kasus dengan detail yang tinggi.
- d. Analisis Dokumen: Memeriksa dokumen atau artefak untuk mendapatkan wawasan tentang suatu topik.
- e. Grup Fokus: Diskusi terstruktur dengan sekelompok kecil orang untuk memahami pandangan mereka tentang topik tertentu.

3. Proses Analisis Data Kualitatif

- a. Transkripsi: Menerjemahkan data audio atau visual menjadi format yang dapat dianalisis.

- b. Koding: Mengidentifikasi pola atau tema dalam data untuk memahami makna di balik informasi yang terkumpul.
- c. Interpretasi: Mengembangkan pemahaman yang lebih dalam tentang fenomena yang diteliti, sering kali dengan menyusun temuan dalam narasi yang kohesif.
- d. Triangulasi: Membandingkan data dari beberapa sumber atau metode untuk memastikan keandalan dan validitas temuan.

4. Kelebihan dan Keterbatasan Data Kualitatif

- a. Kelebihan: Menyediakan wawasan mendalam tentang kompleksitas manusia, fleksibel dalam memeriksa perubahan dan dinamika sosial, memungkinkan eksplorasi yang mendalam.
- b. Keterbatasan: Data kualitatif cenderung lebih sulit untuk di-generalisasi, membutuhkan waktu yang lebih lama untuk pengumpulan dan analisis, serta rentan terhadap bias peneliti.

5. Contoh Penerapan Data Kualitatif

- a. Penelitian Etnografi: Memeriksa budaya atau subkultur tertentu dengan mendalam.
- b. Analisis Tema: Mengidentifikasi pola pikir atau keyakinan yang muncul dari wawancara atau diskusi.
- c. Penelitian Kasus: Memeriksa studi kasus tunggal atau beberapa kasus untuk memahami konteks yang kaya.

2.3.2 Data Kuantitatif

Data kuantitatif adalah jenis data yang diukur dalam bentuk angka atau kuantitas, dan sering kali dapat dihitung dan dianalisis secara statistik.

1. Karakteristik Data Kuantitatif

- a. Numerik: Data kuantitatif terdiri dari angka atau bilangan yang dapat dihitung, diukur, dan dilakukan operasi matematika seperti penambahan, pengurangan, perkalian, dan pembagian.
- b. Terstruktur: Data kuantitatif sering kali diorganisir dalam kategori atau skala yang didefinisikan dengan jelas, seperti skala interval atau rasio.
- c. Objektif: Karena diukur dengan angka, data kuantitatif cenderung lebih objektif daripada data kualitatif, yang lebih subjektif karena didasarkan pada interpretasi dan persepsi individu.
- d. Statistik: Data kuantitatif memungkinkan analisis statistik yang mendalam, seperti perhitungan rata-rata, median, modus, standar deviasi, dan analisis regresi.

2. Jenis-jenis Data Kuantitatif

- a. Skala Nominal: Data yang dikelompokkan ke dalam kategori tanpa urutan atau nilai yang intrinsik (contoh: jenis kelamin, warna).
- b. Skala Ordinal: Data yang memiliki urutan atau ranking tertentu, tetapi jarak antara nilai tidak konsisten (contoh: tingkat kepuasan: rendah, sedang, tinggi).
- c. Skala Interval: Data dengan nilai yang memiliki jarak dan nol bukan merupakan nilai yang mutlak (contoh: suhu dalam derajat Celsius).
- d. Skala Rasio: Data dengan nilai yang memiliki nol absolut dan jarak antara nilai-nilai memiliki arti yang konsisten (contoh: pengukuran panjang, berat badan).

3. Teknik Pengumpulan Data Kuantitatif

- a. Survei: Pengumpulan data dari responden melalui kuesioner atau wawancara terstruktur.
- b. Pengamatan: Pengumpulan data dengan mengamati perilaku atau fenomena tertentu secara langsung.
- c. Eksperimen: Pengumpulan data melalui pengaturan kontrol yang ketat untuk mengukur dampak dari variabel independen terhadap variabel dependen.

4. Analisis Data Kuantitatif

- a. Deskripsi Data: Menggunakan statistik deskriptif untuk merangkum dan menggambarkan karakteristik data, seperti mean, median, modus, dan distribusi.
- b. Analisis Inferensial: Penggunaan statistik inferensial untuk membuat inferensi atau generalisasi dari sampel ke populasi lebih besar, misalnya dengan uji hipotesis dan interval kepercayaan.
- c. Analisis Regresi: Memeriksa hubungan antara variabel independen dan dependen untuk menentukan pengaruh yang ada.

5. Kelebihan dan Keterbatasan Data Kuantitatif

- a. Kelebihan: Memungkinkan analisis yang terstruktur dan objektif, memfasilitasi generalisasi yang lebih luas, dan memungkinkan pengujian hipotesis yang lebih kuat.
- b. Keterbatasan: Terkadang tidak dapat menangkap kompleksitas fenomena manusia secara menyeluruh, cenderung mengabaikan konteks dan nuansa yang lebih dalam.

6. Contoh Penerapan Data Kuantitatif

- a. Survei Opini Publik: Mengukur pendapat masyarakat terhadap isu-isu tertentu dengan angka.
- b. Penelitian Kesehatan: Menganalisis hubungan antara faktor risiko dan kejadian penyakit dalam populasi.
- c. Analisis Pasar: Memprediksi perilaku konsumen berdasarkan data penjualan dan tren pasar.

2.4 Data Primer dan Data sekunder

1. Data Primer

Data primer merujuk kepada informasi atau data yang dikumpulkan langsung oleh peneliti dari sumber pertama. Ini berarti data tersebut dibuat atau dikumpulkan secara spesifik untuk tujuan penelitian tertentu. Data primer diperoleh melalui metode pengumpulan data langsung seperti survei, wawancara, eksperimen, observasi, atau kuesioner. Peneliti bertanggung jawab atas proses pengumpulan, analisis, dan interpretasi data ini.

Karakteristik data primer:

- a. Konteks Khusus: Data primer dikumpulkan untuk menjawab pertanyaan penelitian tertentu.
- b. Kontrol Penuh: Peneliti memiliki kontrol penuh terhadap desain penelitian, metode pengumpulan data, dan analisis.
- c. Waktu dan Biaya: Memerlukan waktu dan biaya yang lebih besar karena proses pengumpulan yang intensif.

Contoh Penggunaan: Penelitian ilmiah eksperimental, survei pasar, studi kasus, atau studi kualitatif yang memerlukan data langsung dari subjek atau partisipan.

2. Data Sekunder

Data sekunder mengacu pada informasi atau data yang telah dikumpulkan sebelumnya oleh pihak lain atau untuk tujuan lain yang tidak langsung terkait dengan penelitian saat ini. Data sekunder dapat diperoleh dari berbagai sumber seperti publikasi akademis, basis data pemerintah, laporan industri, literatur ilmiah, atau data yang sudah ada dalam bentuk digital atau cetak.

Karakteristik data sekunder:

- a. Tidak Langsung: Data ini dikumpulkan untuk tujuan yang berbeda dari tujuan penelitian yang sedang dilakukan.
- b. Sudah Ada: Data ini sudah ada sebelum penelitian dimulai, sehingga peneliti tidak perlu menghabiskan waktu atau biaya untuk mengumpulkannya.
- c. Keterbatasan Kontrol: Peneliti memiliki keterbatasan dalam hal kontrol terhadap kualitas dan relevansi data, serta mungkin harus mengandalkan informasi yang tersedia.

Contoh Penggunaan: Analisis data historis, meta-analisis, penelitian lintas wilayah yang menggunakan data dari survei yang sudah ada, atau penelitian literatur.

2.5 Proses Pengumpulan Data

Proses pengumpulan data adalah langkah-langkah sistematis yang dilakukan untuk mengumpulkan informasi atau data yang relevan dan diperlukan dalam sebuah penelitian atau studi. Tahapan ini penting karena kualitas dan validitas hasil penelitian sangat dipengaruhi oleh cara data dikumpulkan.

2.5.1 Tahap- Tahap Pengumpulan Data

1. Perencanaan Penelitian:

- a. Definisi Tujuan: Menetapkan tujuan penelitian dan pertanyaan penelitian yang spesifik untuk membimbing pengumpulan data.

- b. Desain Penelitian: Memilih metode yang tepat, seperti survei, wawancara, eksperimen, observasi, atau analisis dokumen, berdasarkan tujuan penelitian dan karakteristik populasi atau sampel.

2. Pengembangan Instrumen Pengumpulan Data:

- a. Survei/Kuesioner: Membuat pertanyaan yang jelas dan relevan dengan tujuan penelitian, serta mengatur urutan pertanyaan secara logis.
- b. Wawancara: Menyusun panduan wawancara yang mengarahkan pembicaraan dan memastikan semua aspek penelitian tercakup.
- c. Observasi: Membuat daftar cek atau panduan observasi untuk memastikan pengamatan yang sistematis dan konsisten.

3. Pengumpulan Data:

- a. Implementasi Instrumen: Melaksanakan survei, wawancara, eksperimen, atau observasi sesuai dengan desain penelitian yang telah ditentukan.
- b. Pemantauan Kualitas: Memastikan kualitas data dengan memonitor proses pengumpulan, menjaga konsistensi, dan mengidentifikasi potensi bias atau kesalahan.

4. Pengolahan dan Analisis Data:

- a. Pengkodean: Mengubah data mentah menjadi format yang dapat dianalisis, misalnya dengan memberi kode pada jawaban survei atau wawancara.
- b. Entri dan Verifikasi Data: Memasukkan data ke dalam sistem dan memverifikasi keakuratannya untuk meminimalkan kesalahan input.

- c. Analisis Statistik: Menerapkan teknik analisis statistik yang sesuai untuk menjawab pertanyaan penelitian dan menguji hipotesis.

5. Interpretasi dan Pelaporan Hasil:

- a. Interpretasi: Menguraikan temuan berdasarkan analisis data, mengidentifikasi pola atau tren yang signifikan, serta menjelaskan implikasi temuan untuk teori atau praktik.
- b. Penyusunan Laporan: Menyusun laporan penelitian yang jelas dan sistematis, termasuk deskripsi metodologi, hasil analisis, dan kesimpulan yang didukung data.

2.5.2 Tantangan Dalam Pengumpulan Data

1. Bias dan Kesalahan: Risiko kesalahan pengumpulan data yang dapat mempengaruhi validitas dan reliabilitas hasil.
2. Keterbatasan Sumber Daya: Keterbatasan waktu, biaya, atau aksesibilitas untuk mengumpulkan data yang diperlukan.
3. Kompleksitas Pengambilan Keputusan: Memilih metode yang paling sesuai dengan tujuan penelitian dan karakteristik populasi atau sampel.

2.6 Metode Pengumpulan Data Kualitatif

2.6.1 Wawancara

Wawancara adalah metode pengumpulan data yang melibatkan interaksi langsung antara peneliti (pewawancara) dan responden (orang yang diwawancarai) untuk mendapatkan informasi secara mendalam tentang suatu topik.

Jenis-Jenis Wawancara

1. Wawancara Terstruktur

Wawancara terstruktur melibatkan serangkaian pertanyaan yang telah disusun sebelumnya dan diarahkan secara konsisten kepada setiap responden. Tujuannya adalah untuk memastikan semua topik yang relevan dibahas dalam wawancara.

Contoh:

- a. Penelitian pasar: Pewawancara menggunakan daftar pertanyaan yang sama untuk setiap responden untuk mengumpulkan data tentang preferensi produk.
- b. Penelitian kesehatan: Wawancara terstruktur dapat digunakan untuk mengumpulkan data tentang pola makan atau kebiasaan fisik dari partisipan.

2. Wawancara Tidak Terstruktur

Wawancara tidak terstruktur lebih fleksibel dan mengizinkan pewawancara untuk mengeksplorasi ide-ide baru atau mendalam di luar daftar pertanyaan yang telah disiapkan sebelumnya. Interaksi cenderung lebih bebas dan alamiah.

Contoh:

- a. Penelitian kualitatif: Peneliti dapat menggunakan wawancara tidak terstruktur untuk memahami pengalaman subjektif seseorang terkait dengan situasi atau kejadian tertentu.
- b. Antropologi: Wawancara tidak terstruktur sering digunakan untuk mendapatkan perspektif yang mendalam tentang budaya atau tradisi dari narasumber.

3. Wawancara Semi-Struktur

Wawancara semi-struktur menggabungkan elemen-elemen dari wawancara terstruktur dan tidak terstruktur. Terdapat daftar pertanyaan yang telah disiapkan sebelumnya, namun pewawancara memiliki kebebasan untuk mengeksplorasi topik secara lebih mendalam sesuai dengan tanggapan responden.

Contoh:

- a. Penelitian psikologi: Pewawancara dapat menggunakan panduan wawancara semi-struktur untuk mengeksplorasi lebih dalam perasaan dan emosi dari subjek penelitian.
- b. Penelitian sosial: Wawancara semi-struktur sering digunakan untuk memahami sikap dan nilai-nilai yang dianut oleh kelompok sosial tertentu.

4. Wawancara Telepon

Wawancara telepon dilakukan melalui panggilan telepon, di mana pewawancara bertanya kepada responden berdasarkan skrip wawancara yang telah disiapkan sebelumnya.

Contoh:

- a. Survei kepuasan pelanggan: Pewawancara menggunakan wawancara telepon untuk mengumpulkan data dari pelanggan terkait dengan pengalaman mereka menggunakan produk atau layanan tertentu.
- b. Penelitian politik: Wawancara telepon dapat digunakan untuk mendapatkan opini publik tentang isu-isu politik atau calon politik.

5. Wawancara Fokus (*Focus Group Interview*)

Wawancara fokus melibatkan sekelompok kecil orang (biasanya 6-10 orang) yang dipimpin oleh seorang moderator. Diskusi dilakukan untuk mendapatkan berbagai sudut pandang dan pendapat tentang topik tertentu.

Contoh:

- a. Penelitian pasar: Moderator memimpin diskusi kelompok tentang preferensi konsumen terhadap produk atau merek tertentu.
- b. Penelitian pemasaran: Wawancara fokus digunakan untuk memahami persepsi konsumen terhadap kampanye iklan atau strategi pemasaran baru.

6. Wawancara Naratif

Wawancara naratif fokus pada cerita atau narasi hidup individu, yang memungkinkan pewawancara untuk mendapatkan pemahaman mendalam tentang pengalaman hidup, nilai-nilai, dan peristiwa yang signifikan dalam kehidupan narasumber.

Contoh:

- a. Penelitian sejarah oral: Pewawancara menggunakan wawancara naratif untuk merekam sejarah oral dari generasi yang lebih tua tentang peristiwa historis atau kehidupan mereka di masa lalu.
- b. Antropologi: Wawancara naratif digunakan untuk memahami narasi budaya dan tradisi dari suatu komunitas atau kelompok etnis.

Teknik Pelaksanaan Wawancara

Teknik pelaksanaan wawancara adalah metode atau strategi yang digunakan untuk mengarahkan dan mengelola proses wawancara agar dapat menghasilkan informasi yang relevan dan bermakna dari responden. Teknik-teknik ini penting untuk memastikan bahwa wawancara dilakukan secara efektif dan dapat dipercaya. Berikut adalah beberapa teknik pelaksanaan wawancara beserta contohnya:

1. Pendekatan Terbuka (Open Approach)

Pendekatan ini memungkinkan narasumber untuk mengungkapkan pikiran, pendapat, dan pengalaman mereka dengan bebas tanpa banyak interupsi dari pewawancara.

Contoh: "Bisakah Anda menceritakan lebih banyak tentang pengalaman Anda dalam proyek ini?"

2. Pertanyaan Terbuka (Open Questions)

Pewawancara menggunakan pertanyaan yang memberi kebebasan narasumber untuk menjelaskan atau mengembangkan jawaban mereka.

Contoh: "Bagaimana Anda menangani tantangan yang dihadapi dalam pekerjaan ini?"

3. Pertanyaan Tertutup (Closed Questions)

Pewawancara menggunakan pertanyaan yang meminta jawaban singkat atau dengan pilihan jawaban yang terbatas.

Contoh: "Apakah Anda menggunakan metode A atau B dalam penelitian Anda?"

4. Pertanyaan Probing (Probing Questions)

Pewawancara menggunakan pertanyaan tambahan untuk mendapatkan informasi lebih dalam atau untuk mengklarifikasi jawaban dari narasumber.

Contoh: "Bisakah Anda memberikan contoh konkret tentang bagaimana Anda menyelesaikan masalah tersebut?"

5. Pertanyaan Reflektif (Reflective Questions)

Pewawancara menggunakan pertanyaan untuk mendorong narasumber untuk merenungkan atau mengkritisi pendapat mereka sendiri.

Contoh: "Bagaimana Anda merasa metode ini dapat ditingkatkan?"

6. Silent Probe (Silent Probe)

Pewawancara menggunakan teknik diam untuk memberikan narasumber kesempatan untuk mempertimbangkan jawaban mereka dengan lebih mendalam.

Contoh: Setelah narasumber memberikan jawaban, pewawancara tetap diam untuk beberapa saat sebelum melanjutkan.

7. Teknik Rapport (Rapport Building)

Pewawancara menggunakan teknik untuk membangun hubungan yang baik dengan narasumber untuk menciptakan lingkungan yang lebih terbuka dan nyaman.

Contoh: Memulai wawancara dengan percakapan informal atau menunjukkan minat yang tulus pada pandangan atau pengalaman narasumber.

8. Pengulangan (Repetition)

Pewawancara menggunakan teknik ini untuk mengonfirmasi pemahaman atau untuk meminta klarifikasi jika diperlukan.

Contoh: "Jadi, jika saya mengerti dengan benar, Anda mengatakan bahwa..."

9. Penegasan (Summarization)

Pewawancara menggunakan teknik ini untuk merangkum atau memperjelas apa yang telah dikemukakan narasumber selama wawancara.

Contoh: "Jadi, selama wawancara ini, Anda telah membahas A, B, dan C sebagai masalah utama yang Anda hadapi."

10. Sensitivitas (Sensitivity)

Pewawancara menggunakan kepekaan dan empati untuk menanggapi perasaan atau situasi narasumber secara anggun dan efektif.

Contoh: Mendengarkan dengan penuh perhatian saat narasumber berbicara tentang pengalaman yang mungkin sensitif atau emosional.

Analisis Data Wawancara

Analisis data wawancara merupakan proses menganalisis informasi yang diperoleh dari wawancara dengan responden atau subjek penelitian. Tujuan utamanya adalah untuk mengidentifikasi pola, tema, atau makna yang muncul dari percakapan dan menjadikan data tersebut sebagai dasar untuk menyusun temuan atau kesimpulan penelitian.

1. Transkripsi

Transkripsi adalah proses mentranskripsikan rekaman wawancara dari audio atau video menjadi teks tulisan.

Contoh: Seorang peneliti merekam wawancara dengan narasumber dan kemudian mentranskripsikan percakapan tersebut secara lengkap dan akurat untuk diproses lebih lanjut.

2. Koding Terbuka (Open Coding)

Koding terbuka adalah proses mengidentifikasi dan memberikan label pada ide atau konsep yang muncul secara bebas dari data wawancara.

Contoh: Seorang peneliti membaca teks transkripsi wawancara dan memberikan label atau kode pada setiap ide atau konsep yang berbeda, tanpa pembatasan terhadap kategori tertentu.

3. Koding Aksiomatis (Axiomatic Coding)

Koding aksiomatis melibatkan pengembangan aturan atau prinsip untuk mengklasifikasi dan menginterpretasikan data wawancara berdasarkan teori atau konsep tertentu.

Contoh: Seorang peneliti menggunakan teori tertentu (misalnya, teori feminis atau teori interaksi sosial) untuk mengembangkan kategori-kategori yang digunakan dalam menganalisis data wawancara.

4. Analisis Tematik (Thematic Analysis)

Analisis tematik melibatkan identifikasi, analisis, dan laporan pola tematik (tema atau pola penting) yang muncul dari data wawancara.

Contoh: Seorang peneliti mengidentifikasi tema-tema yang muncul secara berulang dalam wawancara dan mengorganisir temuan-temuan ini menjadi tematik yang mewakili pengalaman atau pandangan responden.

5. Analisis Naratif (Narrative Analysis)

Analisis naratif melibatkan penafsiran narasi atau cerita yang dibagikan oleh responden dan menggali makna dan implikasi di balik cerita tersebut.

Contoh: Seorang peneliti menganalisis bagaimana narasi individu dalam wawancara menggambarkan perubahan atau pengalaman hidup mereka, serta bagaimana cerita tersebut

berkontribusi terhadap pemahaman yang lebih dalam tentang topik penelitian.

6. Analisis Konversasi (Conversation Analysis)

Analisis konversasi mengeksplorasi struktur, pola, dan makna yang terkandung dalam interaksi verbal antara narasumber dan peneliti.

Contoh: Seorang peneliti menganalisis cara-cara di mana interaksi verbal antara narasumber (misalnya, pemilihan kata, jeda, atau intonasi) mencerminkan norma budaya atau dinamika sosial tertentu.

2.6.2 Observasi

Observasi adalah metode penelitian yang melibatkan pengamatan langsung terhadap orang, objek, atau fenomena di lingkungan alaminya. Tujuan utamanya adalah untuk memperoleh pemahaman yang mendalam tentang perilaku, interaksi, atau karakteristik yang diamati tanpa mempengaruhi atau mengubah kondisi yang sedang diamati.

A. Observasi Partisipatif dan Non-Partisipatif

Observasi partisipatif dan non-partisipatif adalah dua pendekatan yang berbeda dalam melaksanakan pengamatan dalam konteks penelitian atau studi lapangan. Keduanya memiliki karakteristik yang unik dan dapat digunakan tergantung pada tujuan penelitian, jenis fenomena yang diamati, serta hubungan antara peneliti dan subjek yang diamati.

1 Observasi Partisipatif

Observasi partisipatif melibatkan peneliti secara aktif terlibat dalam situasi atau kegiatan yang sedang diamati. Peneliti tidak hanya mengamati, tetapi juga berpartisipasi dalam kegiatan, berinteraksi dengan peserta, dan memahami pengalaman mereka dari dalam. Karakteristik observasi partisipatif adalah Peneliti menjadi bagian dari kelompok atau lingkungan yang

diamati, Observasi dilakukan dengan membangun hubungan dekat dengan peserta dan Peneliti dapat memahami konteks sosial, norma, nilai, dan perspektif yang mungkin tidak terungkap dalam observasi non-partisipatif.

Contoh Penerapan: Seorang antropolog melakukan observasi partisipatif di sebuah desa untuk memahami struktur sosial dan budaya lokal. Mereka tinggal bersama masyarakat, berpartisipasi dalam upacara adat, dan belajar bahasa setempat untuk memahami nilai-nilai dan norma yang mengatur kehidupan masyarakat.

2 Observasi Non-Partisipatif

Observasi non-partisipatif melibatkan peneliti sebagai pengamat yang tidak ikut serta dalam kegiatan yang sedang diamati. Peneliti membatasi diri untuk hanya mengamati dan mencatat apa yang terjadi tanpa intervensi langsung dalam situasi yang diamati. Karakteristik dari observasi non-partisipatif yaitu peneliti tetap di luar kegiatan yang diamati dan tidak berinteraksi secara langsung dengan peserta dan observasi dilakukan dengan pendekatan yang lebih objektif dan distanced.

Contoh Penerapan: Seorang peneliti sosial melakukan observasi non-partisipatif di sebuah tempat kerja untuk memahami dinamika komunikasi dan kebijakan perusahaan. Mereka mencatat interaksi antar karyawan, perubahan perilaku terkait dengan kebijakan baru, dan dinamika hierarki organisasi tanpa mengubah dinamika internal dengan berinteraksi langsung.

B. Teknik dan Alat Observasi

Teknik dan alat observasi adalah cara-cara dan instrumen yang digunakan untuk mengamati dan mencatat data secara sistematis dalam penelitian atau studi lapangan. Teknik dan alat ini membantu peneliti untuk memperoleh data yang akurat dan relevan terkait dengan fenomena yang diamati.

Teknik Observasi:

1. Teknik Checklist

Peneliti menggunakan daftar cek (checklist) untuk mencatat kehadiran atau frekuensi kejadian tertentu.

Contoh: Seorang guru menggunakan checklist untuk mencatat kehadiran siswa di kelas setiap hari.

2. Teknik Time Sampling

Observasi dilakukan pada interval waktu tertentu.

Contoh: Peneliti mengamati perilaku belajar siswa setiap 10 menit

selama satu jam di ruang kelas.

3. Teknik Event Sampling

Peneliti mencatat kejadian atau peristiwa yang terjadi pada saat yang ditentukan.

Contoh: Seorang psikolog mencatat setiap kali seorang anak bereaksi terhadap suara keras dalam lingkungan percobaan.

4. Teknik Continuous Recording

Semua kegiatan atau perilaku diamati dan dicatat secara terus menerus.

Contoh: Seorang ahli perilaku hewan mencatat setiap gerakan atau perilaku spesifik hewan dalam percobaan selama periode waktu tertentu.

5. Teknik Structured Observation

Observasi dilakukan berdasarkan struktur atau kerangka kerja tertentu yang telah ditentukan sebelumnya.

Contoh: Peneliti menggunakan daftar cek yang terstruktur untuk mengamati interaksi antara pengasuh dan anak-anak di pusat penitipan anak.

Alat Observasi

1. Catatan Lapangan (Field Notes)

Catatan tertulis tentang apa yang diamati, termasuk detail spesifik seperti perilaku, interaksi, atau lingkungan.

Contoh: Seorang antropolog mencatat obrolan, gestur, dan aktivitas yang terjadi dalam ritual adat masyarakat yang sedang diamati.

2. Kamera atau Rekaman Video

Kamera atau Rekaman Video digunakan untuk merekam visual dari kegiatan atau kejadian yang diamati.

Contoh: Seorang peneliti lingkungan menggunakan kamera untuk merekam perilaku hewan di alam liar.

3. Audio Recorder

Audio Recorder digunakan untuk merekam percakapan atau suara lingkungan yang relevan dengan penelitian.

Contoh: Seorang peneliti bahasa menggunakan perekam audio untuk merekam percakapan informal di komunitas bahasa yang sedang ditelitinya.

4. GPS Tracker

GPS Tracker digunakan untuk melacak lokasi atau pergerakan subjek yang diamati.

Contoh: Seorang ilmuwan lingkungan menggunakan GPS tracker untuk melacak migrasi satwa liar di habitat alami mereka.

5. Observation Schedule atau Jadwal Observasi

Daftar yang berisi waktu dan detail tentang kapan dan apa yang akan diamati.

Contoh: Seorang peneliti kesehatan masyarakat memiliki jadwal observasi yang terstruktur untuk mengamati perilaku makan anak-anak di kantin sekolah.

6. Sensor atau Alat Pemantauan Otomatis

Alat elektronik yang digunakan untuk mengukur atau merekam data secara otomatis.

Contoh: Seorang ilmuwan lingkungan menggunakan sensor suhu untuk merekam perubahan suhu dalam lingkungan tertentu.

C. Mencatat Hasil Observasi

Mencatat hasil observasi adalah langkah kunci dalam mengumpulkan data yang akurat dan terinci dari lapangan.

Pentingnya Mencatat Hasil Observasi

1. Memastikan Keteraturan: Dengan mencatat secara teratur dan sistematis, peneliti dapat memastikan bahwa tidak ada data yang terlewatkan atau terlupakan.
2. Mendukung Analisis: Catatan yang baik memudahkan analisis data setelahnya. Ini memungkinkan peneliti untuk mengidentifikasi pola, tren, atau temuan yang muncul dari data observasi.
3. Menjaga Objektivitas: Dengan mencatat secara langsung dari lapangan, peneliti dapat mengurangi bias yang mungkin muncul dari ingatan atau interpretasi subjektif.
4. Memfasilitasi Reproduksi Penelitian: Catatan yang terinci memungkinkan peneliti lain untuk memahami dan menggali kembali metodologi penelitian yang digunakan.

Cara Mencatat Hasil Observasi:

1. Catatan Lapangan (Field Notes)

Catatan tertulis tentang apa yang diamati selama proses observasi. Peneliti mencatat detail tentang kejadian, perilaku, interaksi, lingkungan, atau hal lain yang relevan dengan tujuan penelitian.

Contoh: Seorang antropolog mencatat perubahan dalam bahasa tubuh dan ekspresi wajah selama ritual adat, serta komentar-komentar yang diberikan oleh peserta.

2. Penggunaan Symbol atau Kode

Peneliti menggunakan simbol atau kode untuk mencatat informasi dengan cepat dan efisien. Misalnya, simbol "#" untuk menandai awal atau akhir kejadian penting, atau "!" untuk menunjukkan emosi yang kuat yang diamati.

Contoh: Seorang peneliti lingkungan mungkin menggunakan "P" untuk merekam pola perilaku hewan yang berulang.

3. Teknik Pencatatan Audio atau Video

Teknik ini menggunakan perekam suara atau kamera untuk merekam observasi secara langsung. Peneliti merekam percakapan, kegiatan fisik, atau situasi lain yang relevan dengan penelitian.

Contoh: Seorang psikolog dapat merekam interaksi antara pasangan dalam sesi terapi untuk analisis lebih lanjut.

4. Pencatatan Terstruktur

Pencatatan Terstruktur menggunakan format atau template untuk mencatat informasi dalam kerangka kerja yang telah ditentukan sebelumnya. Misalnya, menggunakan daftar cek atau formulir pencatatan yang mencakup kategori-kategori tertentu yang ingin diamati.

Contoh: Seorang peneliti kesehatan dapat menggunakan formulir pencatatan terstruktur untuk mencatat perilaku makan anak-anak di kantin sekolah.

5. Penggunaan Teknik Digital

Adalah dengan menggunakan teknologi digital seperti aplikasi pencatatan atau perangkat lunak untuk mencatat dan menyimpan data observasi. Menggunakan aplikasi seluler atau perangkat lunak komputer untuk mengambil catatan, memasukkan foto, atau merekam suara.

Contoh: Seorang peneliti lingkungan menggunakan aplikasi pencatatan untuk mencatat lokasi GPS dan kondisi cuaca selama pengamatan lapangan.

2.6.3 Focus Group Discussion (FGD)

a. Definisi dan Tujuan FGD

FGD adalah metode penelitian kualitatif yang melibatkan diskusi terstruktur antara sekelompok orang yang mewakili kelompok target tertentu. Diskusi ini bertujuan untuk menggali pemahaman mendalam tentang pandangan, sikap, keyakinan, dan pengalaman mereka terkait dengan topik tertentu. Biasanya, FGD melibatkan 6-12 peserta untuk memastikan adanya variasi pandangan yang cukup, namun jumlah ini dapat bervariasi tergantung pada tujuan dan konteks penelitian.

Tujuan FGD:

1. Mendapatkan Wawasan Mendalam

FGD digunakan untuk mendapatkan pemahaman yang lebih dalam tentang keyakinan, sikap, dan persepsi kelompok terkait dengan topik penelitian tertentu. Ini membantu dalam memahami kompleksitas pandangan yang mungkin tidak dapat diperoleh dari metode penelitian lain.

2. Mengidentifikasi Pola dan Variasi

Melalui FGD, peneliti dapat mengidentifikasi pola umum, perbedaan, atau variasi dalam pandangan kelompok terhadap topik yang dibahas. Diskusi kelompok dapat mengungkap perspektif yang saling bertentangan atau faktor-faktor yang mempengaruhi perbedaan pandangan.

3. Menggali Alasan dan Justifikasi

FGD memungkinkan peserta untuk menjelaskan alasan di balik pandangan atau sikap mereka. Ini memberikan wawasan tentang proses berpikir dan pemikiran mereka terhadap topik tersebut.

4. Mengembangkan Hipotesis Awal

Hasil dari FGD dapat digunakan untuk mengembangkan hipotesis awal atau pertanyaan penelitian yang lebih mendalam. Diskusi kelompok sering kali mengarah pada ide-ide baru atau pertimbangan yang mungkin belum dipikirkan sebelumnya.

5. Validasi Data Kualitatif

FGD dapat digunakan untuk memvalidasi temuan kualitatif dari metode penelitian lain, seperti wawancara mendalam atau observasi, dengan memperoleh sudut pandang dari beberapa orang dalam satu waktu.

b. Tahapan Pelaksanaan FGD

1. Perencanaan

- Tujuan: Tentukan tujuan yang jelas dari FGD, seperti topik yang akan dibahas dan informasi yang ingin dikumpulkan dari peserta.
- Pemilihan Peserta: Pilih peserta yang mewakili kelompok target atau yang memiliki pengalaman atau pengetahuan yang relevan dengan topik penelitian.

- Pengaturan Tempat dan Waktu: Tentukan tempat yang nyaman dan waktu yang cocok bagi peserta.
- Panduan Diskusi: Siapkan panduan diskusi yang berisi pertanyaan terstruktur atau topik-topik yang akan dibahas selama FGD.

2. Pelaksanaan

- Moderator: Seorang moderator memfasilitasi FGD dengan memandu diskusi sesuai dengan panduan yang telah disiapkan.
- Aturan Diskusi: Tentukan aturan diskusi seperti waktu yang dialokasikan untuk setiap topik, kebijakan mengenai interaksi antara peserta, dan pentingnya mendengarkan pendapat setiap peserta.
- Rekam Diskusi: Disarankan untuk merekam audio atau video diskusi untuk memudahkan transkripsi dan analisis data lebih lanjut.

3. Analisis Sementara (Debriefing):

Setelah FGD selesai, moderator dapat melakukan sesi debriefing dengan tim peneliti untuk membahas kesan awal, temuan utama, atau pengamatan penting yang muncul selama diskusi.

c. Analisis Data FGD

1. Transkripsi

Audio atau rekaman video dari FGD ditranskripsi ke dalam teks tertulis. Transkripsi dilakukan secara verbatim, mencakup semua percakapan, tanggapan, dan interaksi antara peserta.

2. Kode dan Kategorisasi

Data transkripsi dianalisis dengan menerapkan teknik pengkodean dan kategorisasi. Peneliti mengidentifikasi tema

atau kategori utama berdasarkan diskusi, dan memberikan kode pada unit data yang relevan dengan tema-tema ini.

3. Identifikasi Tema dan Pola

Tema atau pola utama dari data ditemukan dan dianalisis. Peneliti mengelompokkan kode-kode menjadi tema atau pola yang lebih luas yang mencerminkan pemahaman atau pandangan yang muncul dari peserta dalam FGD.

4. Interpretasi dan Penyajian Temuan

Data dianalisis untuk mengeksplorasi implikasi temuan terhadap tujuan penelitian. Peneliti menginterpretasikan temuan berdasarkan analisis data untuk menyediakan wawasan yang mendalam dan konteks terhadap topik penelitian.

5. Validasi dan Kesimpulan

Hasil analisis data FGD divalidasi untuk memastikan keabsahan temuan dan kesimpulan yang dihasilkan. Diskusi kelompok, tinjauan oleh sesama peneliti, atau triangulasi dengan data dari metode penelitian lain digunakan untuk memvalidasi temuan.

2.6.4 Studi Kasus

a. Pengertian studi kasus

Studi kasus adalah metode penelitian kualitatif yang mendalam dan komprehensif yang digunakan untuk menyelidiki fenomena tertentu dalam konteks nyata, seperti individu, kelompok, peristiwa, proses, atau organisasi. Tujuan utama dari studi kasus adalah untuk memahami fenomena yang sedang diteliti secara mendalam melalui pengumpulan data yang kaya dan detail.

b. Proses pelaksanaan studi kasus

1. Perencanaan

- Tujuan Studi Kasus: Tentukan tujuan penelitian dan apa yang ingin dicapai dari studi kasus tersebut.
- Seleksi Kasus: Pilih kasus yang relevan dan representatif yang akan diteliti. Kasus dapat dipilih berdasarkan keunikan, kompleksitas, atau relevansi dengan pertanyaan penelitian.
- Desain Studi: Rancang pendekatan penelitian yang sesuai dengan tujuan dan konteks kasus yang dipilih.

2. Pengumpulan Data

- Metode Pengumpulan Data: Gunakan berbagai teknik pengumpulan data seperti wawancara, observasi, analisis dokumen, atau catatan lapangan untuk mengumpulkan informasi yang relevan tentang kasus.
- Pengamatan Mendalam: Lakukan observasi mendalam terhadap kasus dan interaksinya dengan lingkungan sekitarnya.

3. Analisis Data

- Pengkodean dan Kategorisasi: Analisis data meliputi pengkodean, kategorisasi, dan pengelompokan data berdasarkan tema atau pola yang muncul dari studi kasus.
- Interpretasi: Interpretasikan data untuk mengidentifikasi pola, tren, dan temuan utama yang muncul dari analisis.

4. Pelaporan

- Penyajian Temuan: Sajikan hasil studi kasus dalam format yang sistematis dan terstruktur, termasuk latar

belakang kasus, metodologi penelitian, temuan utama, dan interpretasi.

- Diskusi dan Kesimpulan: Diskusikan implikasi temuan untuk teori, praktik, atau kebijakan yang relevan dengan kasus yang diteliti.

c. Kelebihan dan kekurangan metode studi kasus

Kelebihan Metode Studi Kasus

1. Mendalam dan Komprehensif: Studi kasus memungkinkan peneliti untuk memeriksa fenomena dengan detail yang tinggi dan dalam konteks nyata.
2. Konteks yang Kaya: Penelitian dilakukan dalam konteks yang nyata, yang memungkinkan peneliti untuk memahami kompleksitas dan faktor-faktor lingkungan yang mempengaruhi kasus.
3. Memungkinkan Analisis Kasus yang Kompleks: Cocok untuk kasus-kasus unik atau yang kompleks yang sulit dipelajari dengan metode penelitian lain.

Kekurangan Metode Studi Kasus

1. Generalisasi Terbatas: Temuan dari studi kasus mungkin sulit untuk digeneralisasi ke populasi lebih luas karena fokus yang mendalam pada kasus tunggal atau beberapa kasus.
2. Potensial untuk Bias: Keterlibatan peneliti dalam studi kasus dapat menghasilkan bias subjektif dalam pengumpulan dan analisis data.
3. Waktu dan Biaya: Memerlukan waktu yang lebih lama dan biaya yang lebih besar untuk mengumpulkan data dan menganalisisnya dibandingkan dengan metode penelitian kuantitatif yang lebih ringkas.

2.7 Metode Pengumpulan Data Kuantitatif

2.7.1 Survei

Survei adalah metode pengumpulan data yang sistematis dan terstruktur untuk mendapatkan informasi dari responden tentang berbagai topik atau fenomena. Survei sering kali dilakukan dengan menggunakan kuesioner atau wawancara terstruktur, dan tujuannya adalah untuk mendapatkan pemahaman yang representatif tentang pendapat, sikap, perilaku, atau karakteristik populasi tertentu.

a. Desain dan Distribusi Kuesioner

Desain Kuesioner

- Tujuan: Tentukan tujuan survei dan informasi apa yang ingin dikumpulkan dari responden.
- Pemilihan Pertanyaan: Pilih jenis pertanyaan yang sesuai dengan tujuan penelitian dan karakteristik responden.
- Susunan Pertanyaan: Atur urutan pertanyaan dengan logis dan alir yang baik untuk memfasilitasi pemahaman dan respons yang akurat.
- Bahasa dan Gaya: Gunakan bahasa yang jelas dan sederhana, hindari penggunaan kata-kata ambigu atau bermakna ganda.
- Uji Coba (Pilot Testing): Uji coba kuesioner pada sejumlah kecil responden untuk mengidentifikasi masalah potensial dan memperbaiki desain sebelum implementasi.

Distribusi Kuesioner

- Metode Distribusi: Pilih metode distribusi yang sesuai dengan karakteristik populasi target, seperti survei online, survei kertas, atau wawancara langsung.

- Rekrutmen Responden: Tentukan cara untuk merekrut responden yang representatif dari populasi yang dituju.
- Pengumpulan Data: Pastikan bahwa pengumpulan data dilakukan dengan cara yang sesuai, termasuk pengaturan waktu yang tepat untuk pengumpulan kuesioner yang dikirimkan kembali.

b. Jenis-jenis Pertanyaan Survei

1. Pertanyaan Tertutup

Pilihan Ganda: Responden memilih satu atau lebih dari satu jawaban dari pilihan yang disediakan. Contoh: "Pilihlah salah satu: Suka, Netral, Tidak suka."

2. Skala Likert

Responden menilai sejauh mana mereka setuju atau tidak setuju dengan pernyataan yang diberikan. Contoh: "Saya merasa puas dengan pelayanan pelanggan: Sangat Puas, Puas, Netral, Tidak Puas, Sangat Tidak Puas."

3. Skala Interval

Responden memberikan rating pada variabel dengan skala interval tertentu. Contoh: "Berapa kali Anda berolahraga dalam seminggu? 1 kali, 2 kali, 3 kali, lebih dari 3 kali."

4. Pertanyaan Terbuka

Pendapat atau Tanggapan Bebas: Responden memberikan jawaban mereka dalam bentuk naratif atau deskriptif. Contoh: "Apa yang menurut Anda faktor utama yang mempengaruhi kepuasan kerja?"

5. Pertanyaan Semi-Tertutup

Skala Semantik Diferensial: Responden menilai item menggunakan skala dua ujung yang berlawanan.

Contoh: "Produk ini mudah digunakan versus sulit digunakan."

c. Teknik Sampling dalam Survei

1. Random Sampling

- Random Sampling Sederhana: Setiap anggota populasi memiliki kesempatan yang sama untuk dipilih sebagai sampel.
- Stratified Sampling: Populasi dibagi menjadi lapisan (strata) berdasarkan karakteristik tertentu, dan sampel diambil secara acak dari setiap strata.
- Cluster Sampling: Populasi dibagi menjadi kelompok (cluster), dan beberapa cluster dipilih secara acak untuk dijadikan sampel.

2. Non-Random Sampling

- Convenience Sampling: Memilih responden berdasarkan ketersediaan dan kemudahan akses.
- Purposive Sampling: Memilih responden berdasarkan karakteristik tertentu yang relevan dengan penelitian.
- Snowball Sampling: Menggunakan responden yang merekomendasikan atau mengarahkan peneliti kepada responden lain yang memiliki karakteristik yang diinginkan.

2.7.2 Eksperimen

Eksperimen adalah metode penelitian ilmiah yang digunakan untuk menguji hipotesis atau menyelidiki hubungan sebab-akibat antara variabel-variabel tertentu. Dalam eksperimen, peneliti mengendalikan satu atau lebih variabel independen untuk mengamati efeknya terhadap variabel dependen.

a. Desain Eksperimen

Desain eksperimen merujuk pada struktur atau rencana yang digunakan oleh peneliti untuk mengatur dan mengelola kondisi percobaan. Ada beberapa jenis desain eksperimen yang umum digunakan:

- **Pre-Experimental Designs**

One-shot case study: Peneliti memberikan perlakuan dan mengamati hasilnya tanpa kelompok kontrol atau pengukuran sebelum perlakuan.

One-group pretest-posttest: Peneliti mengukur variabel dependen sebelum dan sesudah pemberian perlakuan pada satu kelompok tanpa kelompok kontrol.

- **True Experimental Designs**

Randomized Controlled Trial (RCT): Peserta acak dibagi menjadi kelompok kontrol dan eksperimen. Hanya kelompok eksperimen yang menerima perlakuan. Ini adalah desain yang paling kuat untuk menentukan sebab-akibat.

Posttest-only Control Group: Peserta acak dibagi menjadi kelompok kontrol dan eksperimen, dengan pengukuran variabel dependen setelah perlakuan diberikan pada kelompok eksperimen.

- **Quasi-Experimental Designs**

Non-equivalent Control Group: Ada kelompok kontrol, tetapi peserta tidak acak ditempatkan.

Time Series Design: Variabel dependen diukur sebelum dan setelah perlakuan pada beberapa titik waktu.

b. Kelompok Kontrol dan Eksperimen

Kelompok kontrol adalah kelompok dalam eksperimen yang tidak menerima perlakuan atau kondisi yang diuji, tetapi tetap diamati dan diukur untuk membandingkan dengan kelompok eksperimen yang menerima perlakuan.

Kelompok eksperimen adalah kelompok dalam eksperimen yang menerima perlakuan atau kondisi yang diuji oleh peneliti untuk melihat efeknya terhadap variabel dependen.

c. Validitas dan Reliabilitas Eksperimen

Validitas dalam konteks eksperimen mengacu pada sejauh mana eksperimen tersebut mengukur apa yang seharusnya diukur dan menghasilkan kesimpulan yang valid tentang hubungan sebab-akibat antara variabel-variabel yang diteliti. Jenis-jenis validitas dalam eksperimen meliputi:

- Validitas Internal: Sejauh mana eksperimen mengisolasi hubungan sebab-akibat antara variabel independen dan dependen tanpa pengaruh variabel-variabel lain.
- Validitas Eksternal: Sejauh mana hasil eksperimen dapat umumkan atau diterapkan pada populasi atau situasi lain di luar lingkungan eksperimen.

Reliabilitas mengacu pada sejauh mana hasil eksperimen dapat diandalkan dan konsisten jika diulang dengan cara yang sama atau di lingkungan yang berbeda. Reliabilitas menjamin bahwa hasil eksperimen dapat dipercaya sebagai hasil dari variabel yang dimanipulasi dan bukan dari kesalahan pengukuran atau faktor-faktor lain.

2.8 Pengukuran dan Skala

Pada dasarnya, skala pengukuran mengacu pada cara atau metode yang digunakan untuk mengukur variabel dalam penelitian.

a. Jenis-Jenis Skala Pengukuran

1. Skala Nominal

Skala yang digunakan untuk mengkategorikan data tanpa adanya urutan atau tingkatan. Contoh: Jenis kelamin (laki-laki, perempuan), status perkawinan (belum menikah, menikah, cerai).

2. Skala Ordinal

Skala yang menunjukkan urutan atau ranking antara nilai-nilai yang diukur, tetapi interval antar nilai tidak konsisten. Contoh: Tingkat pendidikan (SD, SMP, SMA, Sarjana, Magister, Doktor), rating kepuasan (sangat puas, puas, netral, tidak puas, sangat tidak puas).

3. Skala Interval

Skala yang memiliki interval antara titik pengukuran yang tetap, namun tidak ada titik nol mutlak. Contoh: Skala Celsius, skor tes IQ.

4. Skala Rasio

Skala yang memiliki titik nol mutlak dan interval antara titik-titik pengukuran yang tetap. Contoh: Usia, berat badan, pendapatan.

b. Penyusunan Instrumen Pengukuran

Penyusunan instrumen pengukuran melibatkan proses merancang dan mengembangkan alat atau kuesioner yang digunakan untuk mengumpulkan data dalam penelitian. Beberapa langkah yang penting dalam penyusunan instrumen pengukuran meliputi:

1. Definisi Variabel: Tentukan dengan jelas variabel apa yang akan diukur dan bagaimana mereka akan dioperasionalisasikan.
2. Pemilihan Skala: Pilih jenis skala yang tepat berdasarkan sifat variabel yang diukur.
3. Kembangkan Item Pertanyaan: Susun pertanyaan atau pernyataan yang sesuai dengan tujuan penelitian dan teori yang ada.
4. Uji Coba Instrumen: Lakukan uji coba (pilot testing) untuk mengidentifikasi dan memperbaiki masalah potensial dalam instrumen.
5. Reliabilitas dan Validitas: Pastikan instrumen memiliki reliabilitas yang tinggi (konsistensi) dan validitas yang baik (mengukur apa yang seharusnya diukur).

c. Validasi Instrumen

Validasi instrumen adalah proses untuk menilai apakah instrumen pengukuran tersebut memang mengukur apa yang seharusnya diukur dan apakah hasilnya dapat dipercaya. Beberapa langkah dalam validasi instrumen meliputi:

1. Validitas Isi: Pastikan bahwa instrumen mencakup semua aspek variabel yang diukur.
2. Validitas Konstruksi: Periksa apakah instrumen mengukur konsep atau variabel yang dimaksud secara akurat.
3. Validitas Fasial: Evaluasi apakah instrumen terlihat cukup untuk mengukur variabel tersebut.
4. Reliabilitas: Pastikan bahwa instrumen memberikan hasil yang konsisten jika digunakan secara berulang-ulang.

2.9 Sumber Data Sekunder

2.9.1 Pengertian Data Sekunder

Data sekunder merujuk pada data yang telah dikumpulkan oleh pihak lain sebelumnya dan tersedia untuk digunakan kembali oleh peneliti untuk tujuan penelitian mereka sendiri. Data ini tidak diperoleh langsung oleh peneliti melalui penelitian lapangan atau eksperimen, tetapi didapatkan dari sumber yang sudah ada.

2.9.2 Sumber-Sumber Data Sekunder

1. Literatur

Referensi dari buku, jurnal ilmiah, artikel, laporan penelitian, dan tulisan-tulisan ilmiah lainnya yang berisi data atau informasi yang relevan dengan topik penelitian.

2. Dokumen

Dokumen administratif seperti laporan tahunan perusahaan, laporan keuangan, dokumen kebijakan pemerintah, dan dokumen organisasi lainnya yang mengandung data yang dapat digunakan untuk analisis.

3. Database

Kumpulan data yang disimpan secara sistematis dalam bentuk elektronik, seperti basis data statistik pemerintah, basis data industri, atau basis data yang dikelola oleh lembaga riset atau organisasi lain.

2.9.3 Teknik Analisis Data Sekunder

Analisis data sekunder melibatkan proses mengumpulkan, mengevaluasi, dan menginterpretasikan data yang sudah ada untuk menjawab pertanyaan penelitian. Berikut adalah beberapa teknik analisis data sekunder yang umum digunakan:

1. Identifikasi dan Pemilihan Data: Identifikasi data yang relevan dan sesuai dengan tujuan penelitian dari sumber-sumber yang tersedia.
2. Verifikasi Kualitas Data: Evaluasi keandalan, keakuratan, dan relevansi data sekunder yang digunakan dalam penelitian.
3. Transformasi dan Penyusunan Ulang Data: Jika diperlukan, data sekunder dapat diubah bentuknya atau diintegrasikan dengan data lain untuk analisis lebih lanjut.
4. Analisis Statistik: Penggunaan metode statistik untuk menganalisis data sekunder, seperti analisis regresi, analisis faktor, atau teknik clustering untuk mengidentifikasi pola atau hubungan dalam data.
5. Interpretasi Hasil: Menafsirkan temuan dari analisis data sekunder dan mengaitkannya dengan teori atau literatur yang relevan.
6. Validasi Temuan: Memastikan bahwa temuan dari analisis data sekunder konsisten dengan tujuan penelitian dan mendukung hipotesis yang diajukan.

2.10 Etika Pengumpulan Data

2.10.1 Pentingnya Etika Dalam Pengumpulan Data

Etika dalam pengumpulan data merupakan prinsip-prinsip moral yang harus diterapkan oleh para peneliti dalam semua tahapan penelitian mereka. Hal ini penting karena:

1. Perlindungan Partisipan: Etika memastikan bahwa partisipan atau subjek penelitian dilindungi dari risiko atau bahaya yang tidak diinginkan. Ini termasuk keamanan fisik, psikologis, dan emosional mereka selama penelitian.
2. Kepercayaan Publik: Prinsip-prinsip etika yang diikuti dengan ketat membantu membangun kepercayaan masyarakat terhadap penelitian ilmiah. Ini sangat penting

untuk menjaga reputasi ilmu profesionalisme peneliti.

3. Integritas dan Kualitas Data: Etika memastikan bahwa data yang dikumpulkan adalah akurat, relevan, dan dapat dipercaya. Ini mencegah manipulasi atau distorsi data yang dapat merugikan hasil penelitian.
4. Keadilan dan Kepatuhan Hukum: Prinsip etika juga memastikan bahwa penelitian dilakukan dengan cara yang adil dan sesuai dengan hukum yang berlaku, termasuk peraturan terkait privasi dan hak-hak partisipan.

2.10.2 Prinsip-prinsip Etika Penelitian

Beberapa prinsip etika yang harus diikuti oleh peneliti dalam pengumpulan data meliputi:

1. Persetujuan Informed (Informed Consent)

Partisipan harus sepenuhnya mengetahui tujuan, prosedur, dan risiko potensial dari penelitian sebelum mereka setuju untuk berpartisipasi. Mereka harus diberi informasi yang cukup untuk membuat keputusan yang berdasarkan pengetahuan.

2. Kerahasiaan dan Anonimitas

Data individu harus dijaga kerahasiaannya dan partisipan harus dijamin bahwa identitas mereka tidak akan diungkapkan tanpa persetujuan mereka, kecuali dalam keadaan yang diwajibkan oleh hukum.

3. Keadilan

Penelitian harus adil terhadap semua partisipan dan tidak membedakan berdasarkan ras, gender, orientasi seksual, agama, atau karakteristik lainnya. Peneliti harus menghindari potensi konflik kepentingan yang dapat mempengaruhi integritas penelitian.

4. Kesejahteraan Partisipan

Keamanan, kesehatan, dan kesejahteraan partisipan harus menjadi prioritas utama selama dan setelah penelitian. Risiko harus diminimalkan dan manfaat potensial harus melebihi risiko yang terlibat.

5. Transparansi dan Akuntabilitas

Peneliti harus jujur dan transparan tentang metode, prosedur, dan temuan penelitian mereka. Mereka harus siap untuk mempertanggungjawabkan tindakan dan keputusan mereka kepada masyarakat ilmiah dan public.

Bab 3

Teknik Analisis Data dengan Software Statistik

3.1 Analisis Data

Pada saat penelitian setiap data yang dikumpulkan perlu dianalisis untuk memperoleh fakta sebenarnya. Analisis data adalah proses sistematis dalam mengolah, menafsirkan dan mengkomunikasikan informasi yang relevan dari kumpulan data yang diperoleh dengan menggunakan berbagai metode (Assiddiq et al., 2023; Baidi, 2022; Lestari et al., 2022). Agar proses analisis data dilakukan dengan tepat, mudah dan cepat, maka diperlukan alat bantu analisis berupa software atau perangkat lunak. Software atau perangkat lunak statistik adalah alat yang umum digunakan untuk analisis data yang memungkinkan peneliti mengumpulkan, memanipulasi dan memvisualisasikan data secara efisien (Nurviana, 2020). Banyak software statistik saat ini digunakan untuk menganalisis data baik yang berbayar maupun yang *open access* diantaranya:

- 1) SPSS
- 2) Microsoft Excel
- 3) Minitab
- 4) Stata
- 5) SAS

- 6) R
- 7) Python

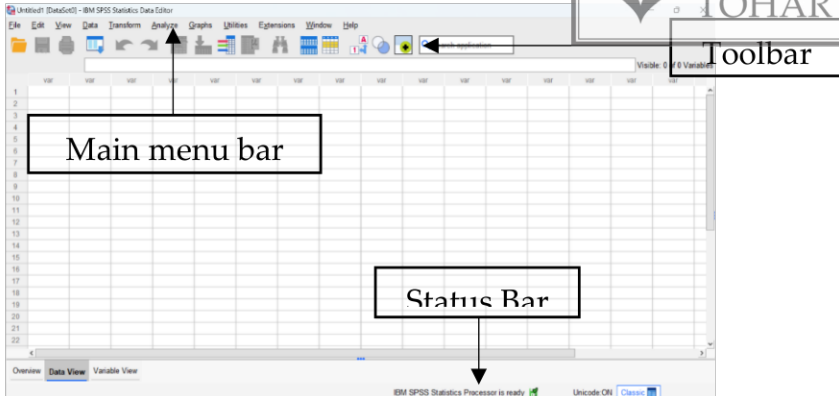
Berikut kami akan memperkenalkan salah satu software yang umum digunakan oleh para peneliti dalam menganalisis dan memvisualisasikan data mereka yaitu SPSS.

3.2 Mengenal SPSS

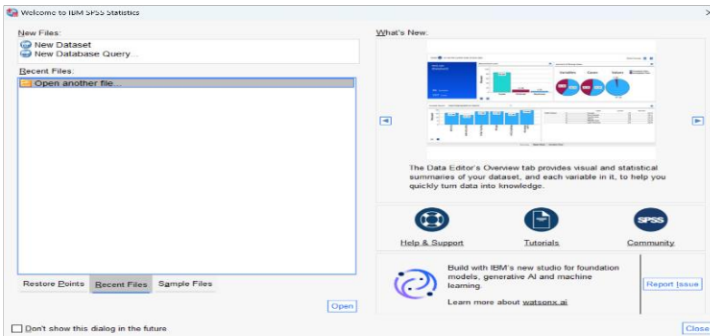
Statistical Product and Service Solution (SPSS) merupakan salah satu software statistik yang paling populer dari banyak paket statistik yang saat ini tersedia. Penampilan sedikit berbeda pada versi Windows, Mac OS dan Linux (Field, 2018). Menurut Pallant (2020) Ada beberapa cara berbeda untuk memulai software ini yaitu:

- 1) Cara yang paling simpel adalah dengan mencari ikon IBM SPSS statistik yang muncul di desktop atau taskbar PC atau Laptop, memulainya dengan cara mengklik dua kali pada ikon tersebut;
- 2) Melalui menu **start**, periksa seluruh program yang terinstal apakah ada folder IBM SPSS statistik, jika ada buka folder tersebut dan klik dua kali pada IBM SPSS Statistik;
- 3) Jika Anda sudah memiliki file dengan format 'sav.', maka anda dapat memulai dengan mengklik dua kali pada file tersebut.

Ketika anda menjalankan software dan terbuka dengan jendela data kosong sebagaimana Gambar 3.1 berarti data siap untuk di input. Sebaliknya jika muncul jendela data kosong dan layar sampul seperti Gambar 3.2, disarankan untuk menutupnya untuk memunculkan jendela data kosong saja.



Gambar 3.1 Jendela Data Editor SPSS



Gambar 3.2 Jendela Layar Sampul

Pada jendela data editor (Gambar 3.1), terdapat beberapa bar meliputi menu bar, toolbar hingga status bar yang berfungsi sebagai pusat navigasi. Berikut beberapa menu utama pada main menu bar (Grotenhuis and Matthijssen, 2016):

- 1) **File**. Menu ini berisi berbagai opsi, yang sering dipakai adalah:
 - Open: membuka file yang sudah ada
 - Save: menyimpan file data maupun output yang sudah dikerjakan
 - Import dan Export: mengimport data dari sumber lain misalnya excel atau mengekspor data ke format lain

- 2) **Edit.** Menu ini berisi berbagai opsi diantaranya:
- Undo/Redo: membatalkan atau mengulangi tindakan terakhir
 - Find/Replace: mencari dan mengganti teks dalam dataset
 - Options: mengatur preferensi pengguna untuk berbagai fungsi dalam SPSS.
- 3) **View.** Pada menu ini pengguna dapat mengubah pengaturan layar. Opsi yang digunakan biasanya digunakan adalah *value labels* untuk menampilkan atau menyembunyikan label nilai dalam data, *grid lines* untuk menampilkan atau menyembunyikan garis grid di lembar data dan *status bar* menampilkan atau menyembunyikan status bar di bagian bawah jendela SPSS.
- 4) **Data.** Menu ini berisi pilihan untuk mendeskripsikan dan memodifikasi variabel serta memilih dan mengurutkan data. Opsi yang umum digunakan adalah *sort case* untuk mengurutkan data, *select case* untuk memilih subset data dan *split file* untuk memisahkan file berdasarkan kategori tertentu
- 5) **Transform.** Menu ini biasanya digunakan pengguna untuk membuat variabel baru. Biasanya jika data tidak normal maka data ini dapat dimanfaatkan untuk menormalkan data. Opsi yang biasa digunakan adalah *compute variabel* digunakan untuk membuat variabel baru, *record into same/different variabel* untuk mengkode ulang nilai variabel dan *rank cases* untuk mengurutkan dan memberikan peringkat data.
- 6) **Analyze.** Ini adalah menu utama dari SPSS statistik terdiri dari berbagai opsi yang digunakan untuk menganalisis data yang Anda miliki. Opsi yang dapat Anda gunakan untuk analisis statistik dasar meliputi *Descriptiv statistic* dengan luaran berupa mean, median, modus dan standar deviasi. *Compare means and proportions* untuk membandingkan rata-rata dua atau lebih kelompok. Regression untuk memahami

pengaruh antar variabel. Correlation hubungan antara variabel.

- 7) **Graphs.** Menu ini memberikan digunakan untuk membuat berbagai bagan dan grafik seperti histogram, diagram batang, dan scatterplot. Opsinya yang umum digunakan meliputi *chart builder* untuk membuat grafik, *legacy dialogs* untuk menyediakan pilihan grafik yang lebih tua dan *interactive* untuk membuat grafik interaktif yang dapat dimodifikasikan langsung.
- 8) **Utilitas.** Menu ini memungkinkan pengguna untuk mengambil detail dari data mereka dengan cepat. Opsi yang umum digunakan adalah *variables* untuk melihat dan mengedit atribut variabel pada dataset. *File info* untuk menampilkan tentang file data yang sedang digunakan dan *use dataset* untuk berpindah dari dataset yang terbuka.
- 9) **Extension.** Menu ini bersisi tautan ke opsi tambahan di luar paket SPSS misalnya untuk menambahkan python atau R dan menjalankannya pada SPSS.
- 10) **Window.** Menu ini memungkinkan pengguna untuk menentukan jendela mana yang aktif.
- 11) **Help.** Menu bantuan yang membantu pengguna dalam memilih analisis yang tepat serta informasi tentang versi SPSS yang digunakan.

Secara default pada toolbar tersedia beberapa menu yang sebagian telah aktif meskipun belum dimasukkan data, namun untuk sebagian lainnya aktif jika data sudah terinput. Jika kursor diarahkan pada menu-menu tersebut, menu yang tersorot akan mengeluarkan deskripsi singkat mengenai fungsinya. Menu status bar memberikan rincian tentang proses yang dijalankan oleh SPSS saat itu. Pada program sedang dijalankan akan munculkan pesan pada status bar "IBM SPSS Statistics Processor is ready" dengan demikian data siap untuk dimasukkan.

3.3 Menyiapkan Data File di SPSS



Misalkan sebuah survei dilakukan kepada keluarga nelayan di suatu kawasan pesisir untuk melihat apakah mereka setuju dengan kebijakan pemerintah terkait dengan kenaikan harga BBM. Data yang diperoleh dari hasil Survei dapat dilihat pada Tabel berikut.

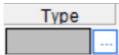
Tabel 3.1 Buku Kode

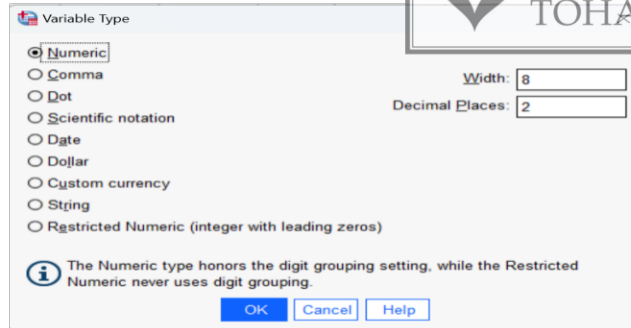
Variabel	Nama Variabel di SPSS	Kode
Jenis Kelamin	Jenis_Kelamin	1 Laki-Laki 2 Perempuan
Umur	Umur	Umur dalam Tahun
Kenaikan BBM	Kenaikan_BBM	1 Sangat tidak setuju 2 Tidak setuju 3 Hampir tidak setuju 4 Setuju 5 Sangat setuju

Tabel 3.2 Data Hasil Survei


Jenis Kelamin	Umur	BBM
1	55	1
1	53	2
2	38	1
1	40	3
1	45	2

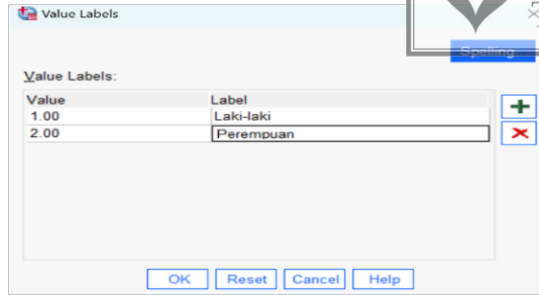
Setelah data penelitian terkumpul, maka ada beberapa langkah yang harus dilakukan untuk menyiapkan data file (Ho, 2006), diantaranya:

- 1) Jalankan program SPSS, lalu muncul jendela data editor sebagaimana Gambar 1.
- 2) Sebelum memasukkan data hasil penelitian, variable-variabel data harus diberi nama dan didefinisikan. Untuk mengisinya klik **Variabel View** pada bagian bawah jendela data editor. Pengisian variabel view meliputi:
 - a. **Name** digunakan untuk memberikan nama variabel. Nama variabel diawali dengan huruf misalnya "Kebijakan1" tidak boleh "1kebijakan" dan tidak boleh menggunakan spasi misalnya "Jenis kelamin" harusnya "Jenis_kelamin" (Cronk, 2020).
 - b. **Type** digunakan untuk mendefinisikan variabel seperti teks, numerik, string dan lain-lain. Klik  untuk memunculkan Gambar variabel type



Gambar 3.3 Variabel Type

- c. **Width** mendefinisikan jumlah karakter yang dapat ditampung oleh kolom yang menampung variabel
- d. **Decimal** mendefinisikan jumlah desimal yang akan muncul di tampilan data
- e. **Label** digunakan untuk mendefinisikan label untuk variabel
- f. **Values** digunakan untuk mendefinisikan label yang sesuai nilai numerik tertentu. Sebagaimana data yang digunakan sebagai contoh pada tabel dan buku kode, maka data jenis kelamin dan kenaikan BBM, memerlukan pendefinisian label. Untuk membuatnya dengan Klik pada bagian  maka muncul jendela value label seperti gambar berikut

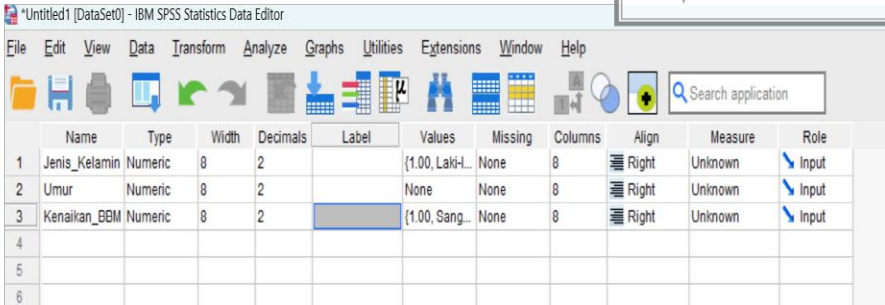


Gambar 3.4 Jendela Value Label

Ketika Anda ingin memasukkan label nilai klik tanda plus (+) setelah setiap entri data. Ketika semua label sudah diinput maka klik ok untuk kembali ke jendela variabel view. Lakukan hal yang sama pada variabel Kenaikan BBM (Cronk, 2020).

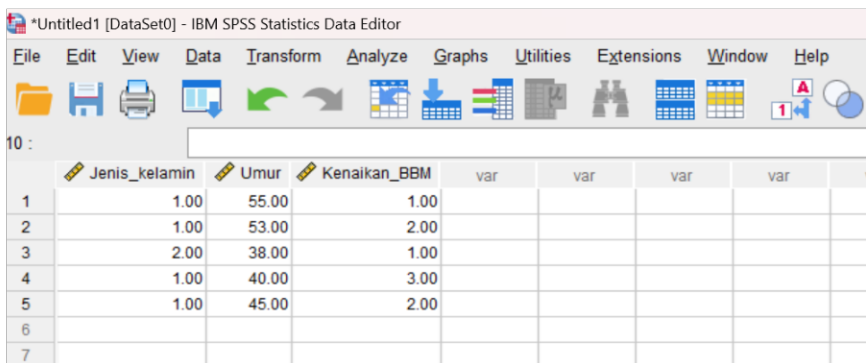
- g. **Missing** mengindikasikan bagaimana data yang hilang ditangani
- h. **Column** mendefinisikan jumlah ruang yang dialokasikan untuk variabel di jendela data view
- i. **Align** mendefinisikan bagaimana data akan muncul dalam sel seperti rata kiri, kanan atau tengah.
- j. **Measure** mendefinisikan skala pengukuran yang paling sesuai dengan variabel seperti nominal, ordinal atau interval.
- k. **Role** digunakan untuk menjalankan analisis secara otomatis terhadap suatu variabel (Field, 2018)

Buku kode menjadi panduan untuk mengisi, menamai dan mendefinisikan variabel. Berdasarkan buku kode yang digunakan, setelah pegisian data dan definisi dilakukan variabel view terlihat seperti Gambar berikut.



Gambar 3.5 Tampilan Data Pada Variabel View

- 3) Setelah entri data pada variabel view selesai, alihkan tampilan dari variabel view ke data view pada sudut bawa jendela editor untuk melakukan input data. Hasil input data terlihat sebagaimana Gambar berikut.



Gambar 3.6 Tampilan Data Pada Data View

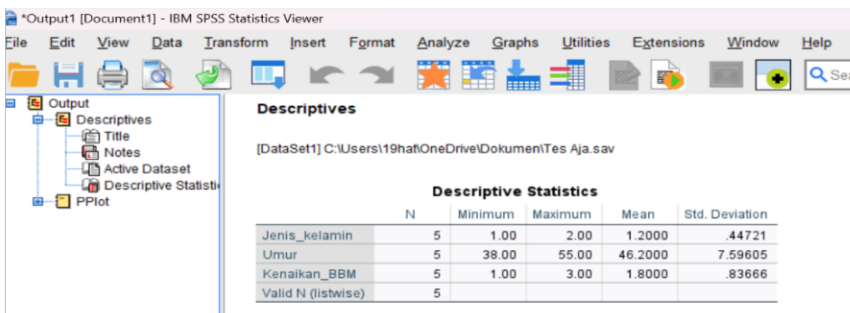
- 4) Selanjutnya, file data dapat disimpan pada komputer atau laptop masing-masing. Penyimpanan dilakukan dengan klik file pada menu bar, lalu klik save as. Tulis nama file anda dan tentukan lokasi penyimpanan pada folder, lalu klik save. File SPSS biasanya tersimpan dengan akhiran .sav (Misalnya PenlitainQ.sav).

File data penelitian juga biasanya diarsipkan pada perangkat lunak lain seperti excel, SAS dan lain-lain. Ketika file tersebut ingin digunakan di SPSS pengguna boleh mengimpor file data tersebut dengan memanfaatkan main menu bar. **Klik**

File > Import Data > Excel/SAS/Stata, cari file folder tempat data tersebut disimpan. Jika sudah ditemukan filenya **Klik Open > OK**.

3.4. Pengenalan Output pada SPSS

Penampilan layar output berbeda dengan layar data view maupun variable view. Pada bagian kiri jendela terdapat diagram pohon, menyediakan bagian-bagian output yang mempermudah anda mengakses jika analisis yang dilakukan bervariasi. Setiap analisis yang dilakukan akan dicantumkan sebagai judul utama di bagian kiri jendela output. Pada bagian kanan terdapat ruang yang besar untuk menampilkan output analisis yang diinginkan baik berupa table, gambar, bahkan pesan kesalahan. Pada Gambar 3.7 merupakan output hasil analisis deskriptif menggunakan data buku kode sebelumnya. Jika output hasil analisis ini ingin dibagi ke orang lain yang tidak menggunakan SPSS, maka harus menyalin dulu ke perangkat lain seperti microsoft word atau PDF. Namun bisa juga menyuruh mereka menginstal smartreader dari web IBM SPSS, agar output dapat terbaca jika belum menginstal SPSS.



The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the following data:

[DataSet1] C:\Users\19hat\OneDrive\Dokumen\Tes Aja.sav

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Jenis_kelamin	5	1.00	2.00	1.2000	.44721
Umur	5	38.00	55.00	46.2000	7.59605
Kenaikan_BBM	5	1.00	3.00	1.8000	.83666
Valid N (listwise)	5				

Gambar 3.7 Output Hasil Analisis dengan SPSS

SPSS sebagai salah satu software analisis data statistik yang banyak digunakan saat ini, untuk itu perlu dikenali dan dipelajari dengan baik. Saat ini sudah mulai dikembangkan versi 30, sedangkan yang digunakan dalam buku ini adalah IBM SPSS versi 29. Terdapat beberapa fitur tambahan yang disesuaikan

dengan kebutuhan analisis saat ini. Tampilan outputnya juga lebih smooth dibandingkan dengan versi yang lebih lawas.

Bab 4

Studi Univariat

4.1 Univariat

Adalah pendekatan untuk analisis data di mana pengamatan dan penjelasan hanya satu variabel data diprioritaskan. "Univariat" berasal dari kata "uni", yang berarti "satu", dan "variat", yang berarti "variabel".

Sebenarnya, tujuan analisis adalah untuk menyederhanakan atau meringkas kelompok informasi hasil pengukuran sehingga dapat diganti menjadi data yang bermanfaat. Peringkasan ini terdiri dari dimensi statistik, tabel, dan grafik. Secara teknis, ringkasan ini adalah meringkas kumpulan informasi menjadi dimensi tengah dan perubahan. Selanjutnya, tujuan analisis adalah untuk membandingkan ringkasan ini dengan kelompok subyek tertentu.

4.2 Peringkasan

Peringkasan informasi ditetapkan oleh tipe informasi secara numerik atau katagorik. Wujud peringkasan sebagai berikut:

1. Ringkasan Data Numerik

Ukuran Rata-rata

Di sisi lain, dimensi tengah menunjukkan konsentrasi nilai-nilai yang dihasilkan dari pengukuran. Untuk mencerminkan dimensi tengah, berbagai dimensi

dibesarkan. Mean, median, dan mode/modus adalah yang paling umum.

1. Rata-rata

Dikenal juga sebagai rata-rata, mean adalah dimensi yang dapat dihitung dengan membagi jumlah total nilai pengukuran oleh banyaknya pengukuran. Menghitung nilai mean, Anda dapat menggunakan rumus berikut:

$$X = \Sigma Xi/n$$

Karena menghitung semua data, mudah untuk menghitung keuntungan dari nilai mean. Namun, nilai ekstrim tinggi dan rendah sangat memengaruhi nilai mean. Oleh karena itu, tidak mungkin untuk menunjukkan kumpulan nilai pengamatan rata-rata pada kelompok data dengan nilai ekstrimnya juga dikenal sebagai "distribusi data yang menceng/miring". Sebagai contoh, ada 6 pasien yang dirawat selama 3 hari, 5 hari, 6 hari, 8 hari, 12 hari, dan 51 hari. Waktu rawat rata-rata adalah $(3+5+6+8+12+51)/6 = 10,6$ hari.

Hasil penghitungan menunjukkan bahwa lama rawat rata-rata 10,6 hari. Ini jelas tidak dapat diterima karena data secara visual sebagian besar kurang dari 6 hari. Ini mungkin karena nilai ekstrim dari kumpulan data di atas.

2. Medium

Penghitungan median hanya melihat urutan nilai hasil daripada perbandingan nilai yang signifikan; median adalah nilai di mana separuh banyaknya pengamatan memiliki nilai di atas dan separuh lagi memiliki nilai di bawahnya. Nilai ekstrim tidak mempengaruhi median karena besarnya perbandingan. Penghitungan median terdiri dari langkah-langkah berikut:

- a) Data diurutkan dari nilai terendah ke nilai paling tinggi
- b) Gunakan rumus $(n+1)/2$ untuk menemukan posisi median.
- c) Cari nilai median.

Metode penghitungan median terdiri dari langkah-langkah berikut:

- a) Data diurutkan dari nilai terkecil ke nilai terbesar.
- b) Cari letak median menggunakan rumus $(n+1)/2$.
- c) Temukan hasil angka median.

Sebagai contoh, ada enam ibu hamil yang berusia 22 tahun, 25 tahun, 28 tahun, 30 tahun, dan 37 tahun.

Data diurutkan: Posisi = $(6+1)/2 = 3,5$

Data yang digunakan sebagai media berurutan ke 3,5, yaitu $(28 + 30)/2 = 29$.

Oleh karena itu, setengah dari ibu hamil yang berusia di bawah 29 tahun dan setengah dari ibu hamil yang berusia di atas 29 tahun

3. Mode atau modus

Nilai pengamatan yang memiliki frekuensi atau jumlah tertinggi disebut mode. Contoh mode data usia karyawan adalah 23 tahun, 28 tahun, 30 tahun, 32 tahun, 35 tahun, dan 37 tahun, yang menyuguhkan bahwa modusnya adalah 30.

4.3 Bentuk Data yang Didistribusikan

Bentuk distribusi data akan ditentukan oleh hubungan nilai mean, median, dan mode:

- Wujud distribusi data normal jika nilai mode, median, dan mean sama.
- Wujud distribusi data menceng/miring ke kanan jika nilai mean lebih besar dari median dan mode.
- Wujud distribusi data mencengangkan ketika nilai $\text{mean} < \text{median} < \text{mode}$, miring ke kanan

1. Tingkat Variasi

Hasil pengamatan cenderung berbeda atau bervariasi satu sama lain. Untuk mengetahui seberapa jauh data berbeda, ukuran variasi seperti standar deviasi, jarak linier kuartil, dan range digunakan.

a. Interval atau jangkauan

Alat yang paling sederhana untuk mengukur variasi adalah interval, yang dapat dihitung dengan menghitung perbedaan nilai terbesar dan terkecil. Kelemahan range dipengaruhi oleh nilai ekstrim. Penghitungan dapat menghasilkan keuntungan dalam waktu singkat.

b. Jarak antara dua kuartil

Nilai-nilai yang dilihat disusun dari nilai kecil ke nilai besar, kemudian ke kuartil bawah dan atas. Kuartil adalah pembagian data menjadi empat bagian yang dibatasi oleh tiga ukuran kuartil: kuartil I, kuartil II, dan kuartil III. Kuartil I mencakup 25% data yang berada di bawah dan 75% data yang berada di atasnya.

Kuartil II (median) mencakup setengah dari data di bawahnya, sedangkan Kuartil III mencakup 75% dan 25% dari data di bawahnya.

Kuartil III dan I berbeda dengan jarak antar mereka, jarak ini lebih besar ketika banyak pengamatan dan tersebar luas.

c. Standar variasi

Untuk mengetahui variasi data, kita harus menghitung penyimpangan atau deviasi dari nilai-nilai pengamatan dibandingkan dengan nilai mean-nya. Variasi ini dihitung sebagai rata-rata dari kuadrat deviasi dibandingkan dengan mean, yang rumusnya adalah;

$$\text{Variasi} = \frac{\sum (X_i - \bar{X})^2}{n}$$

Karena satuan nilai pengamatan tidak sama dengan satuan varian (kuadrat), variasi akan meningkat seiring dengan nilai varian. Akibatnya, standar deviasi diciptakan untuk mengukur variasi.

Variasi berasal dari standar deviasi:

Standard deviasi S atau SD = $\sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$

Variasi meningkat dengan SD. Untuk data numerik, nilai mean (rata-rata), median, standar deviasi, rentang interquartile, minimal, dan maksimal digunakan. Dalam kasus di mana tidak ada variasi, SD=0.

Jika data yang dikumpulkan tidak menunjukkan nilai ekstrim dalam distribusi normal, nilai median dan rentang interkuartil (IQR) lebih tepat daripada nilai mean. Sebaliknya, jika data yang dikumpulkan menunjukkan nilai ekstrim dalam distribusi normal, perhitungan nilai median dan standar deviasi adalah teknik analisis univariat yang tepat.

2. Ringkasan Data Kategorik

Untuk memberikan peringkasan dan informasi penting tentang data katagorik, distribusi frekuensi dan ukuran persentase atau proporsi adalah satu-satunya cara untuk memberikan informasi jumlah dan persentase. Variasi maksimum pada data katagorik terjadi ketika jumlah antar katagori sama.

Contohnya:

Grup C: 50 pelajar laki laki dan 50 pelajar wanita

Grup D: 80 pelajar laki laki dan 20 pelajar wanita

Grup C bervariasi (heterogen), dengan 50% pelajar laki laki dan 50% pelajara wanita. Grup D tidak bervariasi (homogen pada pria), dengan 80% pelajar laki laki dan 20% pelajar wanita.

3. Metode Penyampaian Data

Seseorang dapat menyajikan analisis univariat dalam tabel atau grafik. Namun, perlu diingat bahwa saat menyampaikan data atau variabel hanya menggunakan satu bentuk, bukan keduanya.

Sebagai contoh, menyajikan analisis deskriptif

a. Informasi Numerik

Tabel 1

Rata rata dan nilai Tengah Usia Pasien dan ALOS di RSUD Y di jakarta pada tahun 2025

Variabel	Mean Median	SD	Minimal- Maksimal
1. Usia	35,3 36,1	10,1	20 – 55 Tahun
2. Average length of Stay	9,2 8,5	7,8	3 – 50 Hari

b. Informasi Katagorik

Tabel 2

Penyaluran Responden Menurut Tingkat
Pendidikan Pasien

Rumah sakit X XX Tahun 2025

Pendidikan	Jumlah	Persentase (%)
SMU	55	55,0
D3	20	20,0
S1	25	25,0
Total	100	100,0

Bagaimana Anda melihat tabel di atas?

Kelompok mana yang memiliki konsentrasi/jumlah data tertinggi?"

Saat bersamaan, analisis univariat juga dapat untuk memeriksa variabel yang dapat membantu dalam diagnosis asumsi statistik lanjut; ini terutama berlaku untuk variabel jenis numerik, seperti apakah variasinya homogen atau heterogen atau distribusinya normal. Nilai ekstrim atau outlier juga dapat ditemukan melalui eksplorasi data. Nilai ekstrim sangat menentukan apakah nilai akan diturunkan dalam analisis bivariat berikutnya.

Selain itu, analisis univariat juga untuk mengaplikasikan dalam penilaian variabel yang dapat membantu dalam mengidentifikasi asumsi statistik lebih mendalam. Ini berlaku terpenting untuk variabel jenis numerik, seperti variabel yang memiliki distribusi normal atau homogen. Selain itu, eksplorasi data memungkinkan penemuan nilai ekstrim atau outlier.

Nilai ekstrim sangat berarti buat memastikan apakah nilai hend
ak diturunkan dalam analisis bivariat selanjutnya

Pertimbangkan Distribusi Frekuensi:

Analisis univariat membantu menganalisis distribusi frekuensi responden berdasarkan variabel tertentu, seperti tingkat pengetahuan, kepatuhan menjalankan diet, dan kepatuhan mengonsumsi obat diabetes melitus. Sebagai contoh, analisis univariat dapat menunjukkan bahwa 53,7% responden tidak patuh mengonsumsi obat diabetes melitus oral.

Beberapa pendekatan dapat digunakan dalam analisis univariat, termasuk:

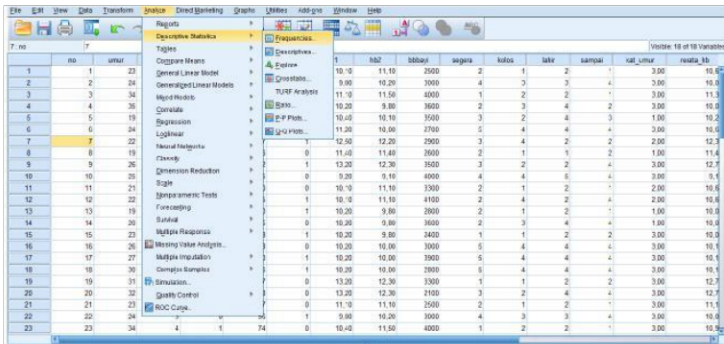
1. Uji normalitas digunakan untuk menentukan apakah data berdistribusi normal. Misalnya, uji Kolmogorov-Smirnov dapat digunakan untuk menentukan apakah variabel model regresi, variabel bebas/independent/(X), dan variabel terikat/dependen/(Y), memiliki distribusi normal. Uji normalitas dilakukan jika ada kurang dari 30 observasi untuk mengetahui apakah error term mendekati distribusi normal. Jika ada lebih dari 30 observasi, uji normalitas tidak diperlukan karena distribusi sampling error term telah mendekati normal. Jika ada 285 observasi, uji normalitas dapat diabaikan. Uji normalitas dapat dilakukan dengan histogram, plot normal P, uji Chi Square, Skewness, dan Kurtosis, atau uji Kolmogorov Smirnov dan Shapiro-Wilk.
2. Rumus persentase digunakan untuk menghitung persentase jawaban responden berdasarkan hasil pengisian kuisioner; contohnya, rumus persentase dapat digunakan untuk menghitung persentase jawaban yang benar.

4.4 Teori Univariat

Memberikan penjelasan tentang proporsi masing-masing kelompok adalah tujuan dari analisis univariat kategorik. Sebagai contoh, analisis univariat data kategorik tentang tingkat

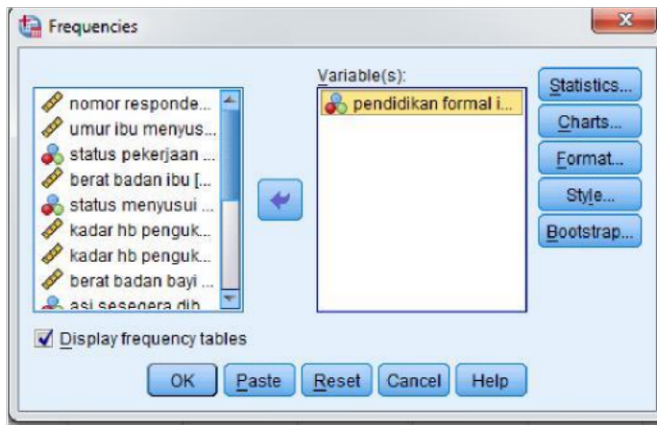
pendidikan formal ibu ditunjukkan di sini, termasuk berapa banyak ibu yang berpendidikan SMA, berapa banyak ibu yang obesitas, dan sebagainya. Apapun versi SPSS yang Anda gunakan, analisis akan sama.

- Klik menu Analisa / *Analyze*, di tampilan awal / *Home Screen*
- Lalu pilih Frekwensi / *Frequencies*;



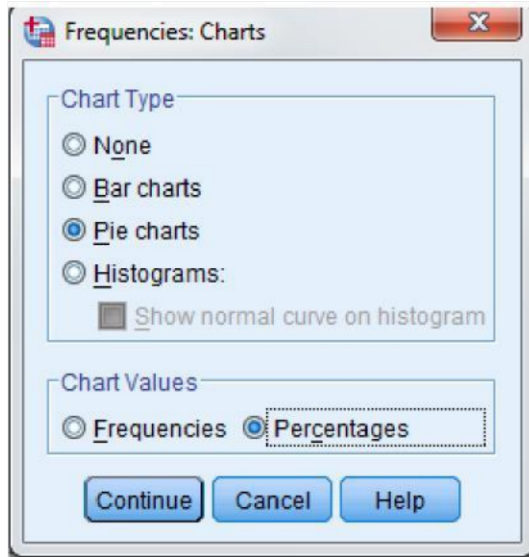
Gambar 4.1. Halaman Awal Uji Univariat Data Kategorik

- input variabel pendidikan formal ke dalam kolom variabel yang muncul. Selama data berbentuk kategorik, kolom ini dapat dipenuhi dengan jumlah informasi sebanyak sesuai dengan kebutuhan yang di inginkan.



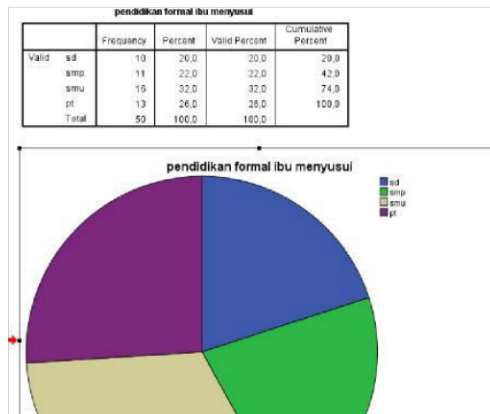
Gambar 4.2 Menunjukkan Kolom Frekuensi.

- d. Untuk menampilkan diagram, seperti diagram *kuil*, tekan menu grafik / *charts*. akan tampil kolom frekuensi / *frequencies: grafik / chart*;

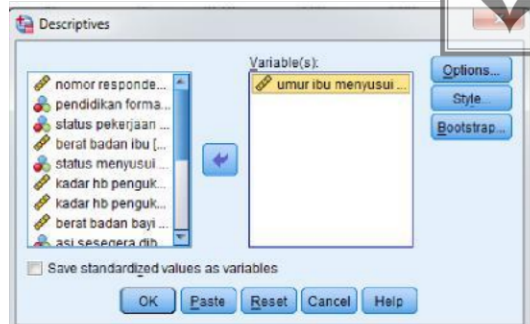


Gambar 4.3. Kolom Frekwensi / *Frequencies* : Grafik / *chart*

- e. Tekan tombol Lanjut / *Continue*
 f. Tekan Ya / *Ok*



Gambar 4.4 Menunjukkan Hasil Dari Analisis Univariat Data Kategorik.



Gambar 4.6 Kolom *Descriptive*.

- d. Selanjutnya, pilih menu opsi / *options*;
- e. Klik/Pilih beberapa analisis, misalnya untuk mendapatkan nilai minimum, maksimum, rata-rata, standar deviasi, dan nilai deviasi;



Gambar 4.7. Kolom Deskripsi / *Descriptive*: Pilihan / *Options*

- f. Selanjutnya, pilih menu Pilihan / *Options*;
- g. Klik/Pilih beberapa analisis. Misalnya, nilai minimum, maksimum, standar deviasi, dan rata-rata.
- h. Klik tombol Lanjut / *Continue*.
- i. Klik Ya

Descriptive Statistics

	N	Minumum	Maximum	Mean	Std. Deviation
Usia ibu Hb rendah Valid N	40 40	20	36	24,50	4,250

Descriptives

Gambar 4.8 Menunjukkan Output Dari Analisis Univariat Data Numerik.

Hasil uji menunjukkan usia ibu Hb rendah dengan minimum 20 tahun, maksimum 36 tahun, dan rata-rata 24,50 tahun dengan standar deviasi 4,250. Tabelkan data tersebut agar lebih menarik.

Bab 5

Analisis Bivariat

5.1. Bivariat

Sikap ilmiah mendorong para peneliti untuk berpikir tentang penelitian yang akan dilakukan dengan menggunakan metode ilmiah yang sistematis. Strategi yang dilakukan dalam penelitian dengan menguji hipotesis, pemilihan subjek, data yang diperoleh, teknik yang digunakan dengan prosedur yang diambil dalam pemilihan metode pada statistik penelitian.

Statistik berasal dari kata Latin, yang berarti "kapasitas", yang berarti kondisi pada kebijakan, dan bahasa Italia, yang berarti "situasi" ("negarawan" atau "politikus"). Pada awalnya, istilah "statistika" digunakan untuk data seperti data sensus, data militer, data fiskal, dan data politik. Namun, istilah "statistika" sering disalahgunakan, sehingga pengertian dari statistika merupakan bidang ilmu mengeksplorasi bagaimana suatu objek dapat diidentifikasi, dideskripsikan dan dianalisis dari berbagai aspek terkait yang menghasilkan output penelitian dan sebagai pendukung pengambilan keputusan.

Untuk tujuan penelitian, metode pengumpulan, pengolahan, penyajian, dan analisis hipotesis, serta metode peramalan/peramalan interval, semuanya merupakan bagian dari hubungan antara statistik dan metode ilmiah statistik. Alat dan metode statistik yang tersedia dalam perkembangan dan pemahaman pada bidang penelitian ataupun pada pekerjaan

yang menentukan aktivitas dalam bidang sosial, fisik, sumber daya alam serta informasi dan teknologi.

Statistik menunjukkan keterbatasan masing-masing desain penelitian dengan menunjukkan cara untuk mengurangi keterbatasan variabel tersebut. Hal ini dapat dicapai dengan mengontrol variabel terpilih atau dengan memanfaatkan desain penelitian dengan mempertimbangkan pengaruh lebih banyak variabel. Kesalahan dapat diminimalkan dengan meningkatkannya.

Perkembangan statistik yang telah menghasilkan metode untuk menghitung beragam aspek tertentu, seperti ukuran *tendency central, dispersion, and error*, yang sangat membantu para peneliti dalam mengklasifikasikan dan menyajikan data. Dengan menggunakan teknik yang dikembangkan dalam statistika untuk menyajikan data, seperti tabel atau grafik, mereka membuat data mudah dipahami oleh semua orang.

Secara harfiah, penelitian yang memiliki beragam instrumen yang akan diteliti, membutuhkan data pertanyaan dan uji hipotesis yang menghasilkan rancangan penelitian dan tidak menggambarkan data empiris yang menyimpang dari hasil penelitian.

Kualitas data yang dihasilkan oleh peneliti bergantung pada reliabilitas dan validitas pada instrumen atau alat yang digunakan untuk menggabungkan berbagai data. Jika instrumennya valid dan reliabel, data yang dikumpulkan juga akan cukup valid dan reliabel. Namun, jika pengambil data kurang atau tidak memahami instrumennya, data yang dikumpulkan mungkin tidak valid dan tidak reliabel.

Pada teknik maupun metode pengambilan data biasanya disertai dengan tahapan pelaksanaan, yang harus dipahami oleh peneliti maupun surveyor yang diberikan oleh peneliti untuk mengumpulkan data lapangan. Selain itu, tahap penelitian dilakukan secara sistematis dan terukur dalam proses penelitian.

Bentuk analisis data yang dilakukan bila proses pengolahan data dilakukan dengan data dan informasi yang lengkap maka proses pengolahan analisis akan berjalan dengan baik. sesuai dengan jumlah pasangan variabel yang dianalisis, analisis univariat, analisis bivariat, dan analisis multivariat. Jika hanya ada satu variabel yang dianalisis, itu disebut analisis univariat. analisis bivariat merupakan analisis yang bertujuan menguji hipotesis penelitian yang dibuat oleh peneliti dengan dua macam variabel, sedangkan analisis multivariat merupakan analisis yang memiliki lebih dari dua variabel dengan analisis determinan maupun analisis faktorial.

5.2. Pengertian Analisis Bivariat

Definisi dari analisis bivariat merupakan analisis memiliki koneksi antara dua variabel yang tertuang dalam tabel silang. Untuk menghasilkan tabel silang ini, peneliti sebaiknya menentukan apakah hubungan terkandung asimetris, simetris, atau resiprokal. Ini berarti menggunakan persentase seperti baris, kolom, atau total..

Analisis bivariat berfungsi untuk memperhitungkan substansi sistem dengan membandingkan nilai $p < 0,05$, dengan dua variabel independen dan variabel dependen (H_0 ditolak). Begitupun sebaliknya bila tidak ada koneksi/hubungan maka (H_0 diterima) bila nilai p lebih besar dari $0,05$ ($p > 0,05$), dengan hasil uji analisis data menunjukkan bahwa nilai p lebih kecil dari $0,05$ ($p < 0,05$) yakni nilai p sama dengan $0,00$ dan nilai r sama dengan $0,559$, hal ini menghasilkan bahwa ada keterkaitan pada hubungan yang bermanfaat antara variabel dependen dengan variabel independen.

Hubungan asimetris menggunakan persen baris dan kolom berdasarkan posisi variabel terikat dalam tabel yang ditafsirkan diikuti oleh posisi variabel bebas. Hubungan timbal balik atau simetris menggunakan persen total.

Dalam pengukuran statistik yang dilakukan berdasarkan pada tingkat pengukuran dengan nilai lebih minimal. Dimana, pengukuran yang nilainya lebih tinggi mempunyai karakter tingkat pengukuran memiliki nilai lebih rendah. Apabila peneliti menggunakan instrumen penyusun data, maka akan menciptakan beragam jenis tolok ukur analisis yang akan dipergunakan dalam penelitian.

5.3. Hubungan asimetris pada Analisis Bivariat

Hubungan asimetris ini menjelaskan pengaruh suatu variabel berdampak pada variabel lainnya. Terdapat enam tipe hubungan asimetris, diantaranya :

1. Hubungan antara peristiwa dengan respon

Pada hubungan tersebut digambarkan variabel pada objek peristiwa yang mempengaruhi variabel respon, yang selanjutnya mempengaruhi reaksi terhadap stimulus. Para ilmuwan yang melakukan penelitian kuantitatif dalam berbagai bidang keilmuan, termasuk ilmu kealaman dan ilmu kesosialan dengan melihat hubungan ini sebagai bagian dari hubungan sebab akibat. Sebagai contoh, Para peneliti menyelidiki hubungan antara tayangan negatif di media massa dan perilaku sikap remaja di perkotaan terhadap tayangan tersebut, selain itu pada penelitian berkaitan dengan bagaimana ekonomi wilayah dan kota yang setelah mendapatkan bantuan cicilan UKM pada wilayah tersebut.

2. Hubungan antara disposisi dengan respon

Di sini, disposisi berarti keinginan untuk menunjukkan respons khusus dalam posisi dan kondisi tertentu. Namun, tingkah laku individu menentukan respons. Misalnya, menonton TV, memilih merek mobil, merokok, membolos, dan hal lainnya. Kemudian pada perilaku masyarakat terhadap sektor informal dan preferensi bekerja di sektor informal adalah contoh sederhana dari hubungan disposisi

dengan respon umum. Dalam beberapa kasus, sektor informal tampaknya sulit diatur dan tidak menjamin pendapatan yang memadai. Akibatnya, seseorang mungkin bersikap sinis terhadap sektor ini. Namun, ketika seseorang tidak dapat mendapatkan uang disektor tersebut dengan usaha yang lebih baik, mereka akan menetapkan sektor informal sebagai cara sumber penghasilan.

3. Hubungan antar ciri individu dengan disposisi

Dalam hubungan ini berupa stigma individu dengan karakteristik pribadi maupun individu yang tidak dapat berubah meskipun lingkungannya berpengaruh. Faktor-faktor seperti jenis kelamin, suku, ras, kebangsaan, dan sebagainya dapat menjadi ciri khas seseorang. Misalnya, ada karakteristik pada ciri khas budaya orang Madura, dengan karakter pekerja ulet, terlihat keras, dan jujur. Orang Jawa juga memiliki ciri-ciri budaya, seperti keinginan kuat untuk mempertahankan hubungan, kesulitan memilih, budaya ewuh pekewuh, dan terkesan lamban karena mereka sering menggunakan kekuatan batin saat menghadapi permasalahan.

4. Hubungan antara prakondisi yang dibutuhkan dan akibat yang diinginkan

Jika ada variabel lain, ada variabel itu; jika muncul, hanya jika variabel yang lain muncul lebih dahulu. Misalnya, masyarakat lebih aktif dalam melaporkan kasus korupsi bila dalam kondisi tersebut akan terjamin keselamatan diri serta memastikan bahwa kejadian tersebut akan ditangani oleh pihak berkepentingan. Dalam masyarakat demokratis yang sudah mapan, orang-orang akan memanfaatkan kebebasan mereka untuk menentukan wakil rakyat dalam pemilu jika mereka yakin bahwa pilihan mereka dengan mengupayakan hak-hak rakyat di badan legislatif.

5. Hubungan yang imanensi antara dua variabel

Hubungan imanensi berarti bahwa perubahan bentuk variabel tertentu diikuti oleh perubahan bentuk variabel lainnya. Salah satu contoh hubungan imanensi adalah jumlah bentuk barang yang beredar di lingkungan pasar baik dalam keadaan yang berlebihan ($S > D$), maka ketika hal tersebut terjadi harga barang akan turun. Sebaliknya bila jumlah barang mengalami penurunan ($S < D$) maka harga barang akan naik. Selain itu, ada contoh lainnya yakni jumlah lulusan sekolah semakin bertambah, namun tidak sesuai dengan jumlah penambahan sekolah yang baru, akan menyebabkan lebih banyak pengangguran para siswa.

6. Hubungan antar variabel tujuan dengan variabel cara

Ketika seseorang menginginkan sesuatu, mereka biasanya menemukan tujuan dan cara untuk mencapainya. Namun, tidak jarang tujuan dan metode berbeda dalam berbagai kondisi. Contohnya adalah hubungan antara janji kesejahteraan masyarakat yang ditawarkan oleh partai politik dalam kampanye politik dan kondisi partai politik mengontrol dalam mengontrol sistem pemerintahan. Selain itu misalnya pada hubungan antara tingkat kesembuhan pasien dengan jumlah obat yang dikonsumsi, hubungan antara kasih sayang orang tua terhadap anaknya dan bagaimana mendidiknya hingga sukses.

5.4. Jenis Pengukuran Penelitian Analisis Bivariat

Adapun beberapa ukuran terkait analisis bivariat diantaranya sebagai berikut :

1. Chi-Square (χ^2)

Hipotesis pada analisis komparatif bivariate yang tidak berpasangan diuji pada pengukuran skala data berupa nominal.

Dalam studi kasus, jika ada dua variabel, salah satunya adalah variabel yang memiliki skala nominal, kemudian pada uji chi-square dilakukan dan disarankan untuk menggunakan uji di tingkat terendah.

Uji Chi-square adalah uji non parametris yang sering digunakan. Namun, Anda harus tahu bahwa uji ini dapat digunakan untuk beberapa syarat, yaitu frekuensi responden atau sampel yang besar. Uji Chi-square dapat digunakan untuk beberapa syarat, seperti:

- a) Tidak ada sel yang memiliki nilai frekuensi kenyataan atau disebut dengan jumlah kenyataan/ *actual count* (F_0) yang sama dengan 0 (nol);
- b) Jika tabel kontijensi berukuran 2×2 , maka hasilnya tidak boleh ada satu sel pun dengan mendapatkan nilai frekuensi harapan maupun disebut juga jumlah yang diharapkan / *expected count* (" F_n ") dengan nilai kurang dari 5;
- c) Jika bentuk tabel lebih dari ukuran 2×2 , misalnya pada ukuran 2×3 , maka jumlah sel dengan frekuensi harapan memiliki nilai kurang dari 5 tidak boleh lebih dari nilai 20%.

Dengan demikian tahapan uji Chi square dengan tabel kontingensi 2×2 harus diperhatikan bahwa tidak boleh ada data menghasilkan nilainya dibawah 5 baik itu pada nilai Nol maupun dengan tidak ada nilainya. Sehingga, analisis pada Uji Chi Square baik keseluruhan sel dalam tabel tidak boleh ada kosong dan memiliki nilai dibawah sebesar 5, sehingga semua sel pada tabel kontinjensi 2×2 yang terisi keseluruhannya.

Dengan demikian pada hasil analisis tersebut, Nilai P dihasilkan oleh Asymp.Sig (dua sisi) pada *Continuity Correction* dengan didapatkan Nilai P (Asymp.Sig) sama

dengan 0,000. Dengan demikian membuktikan pada analisis, bahwa nilai P kurang dari 0,05 ($p < 0,05$) dengan hasil H_0 ditolak. Sehingga pada rumus uji Chi-square dapat menghasilkan hipotesis tidak hanya ada satu. Jika pada tabel kontijensi 2 X 2 maka rumus yang digunakan adalah *koreksi kontinuitas*. Jika pada tabel kontingensi 2 X 2, dengan hasil tidak memberikan syarat dalam uji Chi-square maka rumus yang diperlukan yaitu *Fisher Exact Test*. Sedangkan apabila tabel kontijensi lebih dari 2 X 2 seperti pada tabel kontigensi dengan 2 X 3 maka rumus yang diperlukan yakni *Pearson Chi-square* (Supranto, 2000). Uji Chi Square dapat diinterpretasikan yakni sebagai berikut :

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

dimana,

X^2 = Distribusi Chi-square

O_i = Nilai Observasi (pengamatan) ke – i

E_i = Nilai ekspetasi ke – i

Adapun langkah-langkah dalam pengujian Chi Square, yaitu :

1. Merumuskan dan mendeskripsikan hasil uji hipotesis dengan H_0 dan H_1 , dengan ketentuan
 - H_0 : tidak memiliki pengaruh signifikan
 - H_1 : pengaruh signifikan
2. Mencari nilai frekuensi harapan (E_i)
3. Membagi pada distribusi *chi-square*
4. Menetapkan hasil dari nilai taraf signifikansi α
5. Menetapkan nilai dari X^2
 - a. Taraf signifikansi (α) = 0,05

b. $d.f = (\text{Jumlah baris} - 1) (\text{Jumlah kolom} - 1)$

6. Memastikan kriteria pengujian

Jika $X^2_{hitung} < X^2_{tabel}$, maka H_0	Diterima
Jika $X^2_{hitung} > X^2_{tabel}$, maka H_0	Ditolak
Jika $Sig. \geq 0,05$ maka H_0	Diterima
Jika $Sig. < 0,05$ maka H_0	Ditolak

7. Membandingkan X^2 hitung dengan X^2 tabel atau Sig. dengan α Keputusan H_0 ditolak atau diterima

8. Membuat kesimpulan, ada tidaknya pengaruh antar variabel.

2. Lamda (λ)

Lambda (λ) adalah metrik yang menunjukkan penurunan proporsional dalam kesalahan. Dengan mempertimbangkan ukuran ini, arti pergaulan menjadi lebih jelas. Metode pengukuran ini bergantung pada rasio kesalahan pengukuran yang digunakan untuk memprediksi nilai suatu variabel hanya berdasarkan variabel itu sendiri, serta kesalahan pengukuran yang sama yang digunakan untuk memprediksi nilai variabel lainnya.

Nilai lambda (λ) selalu berada antara 0 dan 1. Nilai 0 menunjukkan bahwa variabel bebas tidak dapat memprediksi variabel terikat, sedangkan nilai 1 menunjukkan bahwa variabel bebas dengan jelas menentukan variabel terikat. Untuk variabel dengan tingkat pengukuran nominal dan arah hubungan asimetris, uji lambda (λ) digunakan.

Lamda (λ) adalah parameter penentu. Jika nilainya = 1, ruang jarak antara dua variabel sama, ruang jarak antara dua variabel sama, dan ruang jarak tak terhingga.

3. Kendall Tau (τ)

Jika pada data berskala ordinal, digunakan untuk mengidentifikasi hubungan antara dua variabel atau lebih, atau untuk menguji signifikansi hipotesis asosiatif. Uji Korelasi Kendal Tau (τ) ini berbeda dari Korelasi Spearman Rank karena dapat diperkirakan untuk sampel yang memiliki bobot tinggi, lebih dari 30 akan menghasilkan nilai hubungan antar variabel yang berbeda. Gagasan dari hasil hipotesis untuk menerima atau menolak H_0 adalah:

a) Berdasarkan Perbandingan antara nilai X hitung dengan nilai X pada tabel:

- ✓ Jika nilai X dihitung lebih besar dari X tabel maka H_0 ditolak atau dalam hal ini berarti H_a diterima, sehingga dapat disimpulkan bahwa ada hubungan antar variabel yang diuji; dan kebalikkannya,
- ✓ Jika nilai X dihitung lebih kecil dari X tabel, maka H_0 diterima atau dalam hal ini berarti H_a ditolak, dengan demikian dapat diketahui bahwa tidak ada hubungan antar variabel yang diuji.

b) Berdasarkan perbandingan antara nilai P dengan α (0,05)
:

- ✓ Jika Nilai p lebih kecil dari 0,05 ($p < 0,05$), maka uji hipotesis pada H_0 ditolak dan H_a diterima, maka dalam hal ini ada hubungan antar variabel yang diuji; dan kebalikkannya,
- ✓ Jika Nilai p lebih besar dari 0,05 ($p > 0,05$), maka uji hipotesis pada H_0 diterima dan H_a ditolak, dalam hal ini tidak ada hubungan antar Variabel yang diuji.

4. Somers'd

Kekuatan hubungan diukur dengan parameter analisis ini pada tingkat pengukuran ordinal dengan arah hubungan

asimetris dan simetris. Somers'd menciptakan korelasi Somers pada tahun 1962. Untuk melihat hubungan di antara dua variabel dengan skala data ordinal, korelasi ini adalah korelasi nonparametrik yang tepat. Variabel X dan Y adalah dua variabel yang membentuk korelasi Somers'd (Nugroho, Akbar, Vusvitasari, 2008).

Jika variabel X dianggap sebagai variabel independen dan variabel Y dianggap sebagai variabel dependen, maka parameter populasi di antara kedua variabel tersebut adalah Δ_{YX} . Sebaliknya, jika peran masing-masing variabel dibalik, atau jika variabel X dianggap sebagai variabel dependen dan variabel Y dianggap sebagai variabel independen, maka parameter populasinya adalah Δ_{YX} . Koefisien korelasi Somers'd dalam sampel ditunjukkan dengan d_{YX} dan d_{XY} .

Untuk mendapatkan koefisien korelasi Somers'd, nilai konkordan (C) dan diskordan (D) dari masing-masing variabel harus dihitung terlebih dahulu.

Jika nilai pada $D = 0$ dan setiap baris mempunyai setidaknya satu sel yang tidak bernilai sama dengan nol, maka koefisien korelasi Somers'd $d_{YX} = 1$. Dalam bentuk tabel kontingensi, sel yang memiliki nilai sebesar nol akan turun dari atas ke bawah seperti tangga, dan sebaliknya, koefisien dari korelasi Somers'd $d_{YX} = -1$ jika sel tersebut memiliki nilai nol yang tersusun dari bawah ke atas.

Jika variabel dalam sampel adalah independen, koefisien korelasi Somers'd $d_{YX} = 0$. Namun, jika koefisien korelasi Somers'd $d_{YX} = 0$, tidak ada bukti langsung bahwa variabel tersebut adalah independen (Siegel dan Castellan, 1988).

5.5. Metode Pengujian Bivariat Untuk Hipotesis Penelitian

Untuk dapat memperkirakan hasil analisis dan memastikan teknik analisis yang tepat untuk pengujian hipotesis bivariat, dengan memahami tahapan-tahapan yang diperlukan untuk

menentukan metode uji statistik yang tepat. Dalam hal ini, peneliti harus melakukan hal-hal selanjutnya:

- a. Menentukan capaian pada tujuan dengan jenis hipotesis dibuat, seperti hipotesis komparatif atau hipotesis korelatif;
- b. Menentukan klasifikasi skala variabel maupun kelompok data yang didapatkan, apakah sudah termasuk dalam kategori (nominal dan ordinal) atau numerik; dan
- c. Dalam hipotesis komparatif, terdiri jenis data yang akan disampel atau kelompok data yang diperkirakan, apakah berpasangan ataupun tidak berpasangan.
- d. Identifikasi distribusi data adalah normal atau tidak normal untuk skala data numerik. Ini untuk menentukan apakah metode uji statistik parametric masih dapat digunakan atau apakah lebih baik menggunakan metode non-parametrik.

Setelah proses penelitian diamati dengan sistematis, peneliti dapat memanfaatkan bentuk saran secara umum untuk memastikan dan memastikan jenis uji statistik yang akan diperhitungkan untuk menganalisis hipotesis bivariat.

Terdapat tiga indikator yang harus dipahami yakni bentuk hipotesis, skala data dan kelompok data. Menurut ketiga indikator tersebut, maka terumusnya tabel untuk Panduan Untuk Menentukan Analisis Uji Statistik Dengan Pengujian Hipotesis Secara Bivariat (2 Sampel), sebagai berikut: (Sugiyono, 2009; Dahlan, 2014; Stang, 2018).

Tabel 5.1. Panduan Untuk Menentukan Analisis Uji Statistik Dengan Pengujian Hipotesis Secara Bivariat

Skala data/Skala Pengukuran Variabel	Bentuk Hipotesis		
	Komparatif		korelatif
	berpasangan	Tidak berpasangan	
Nominal	Mc. Nemar	Chi Square	Koefisien Kontingensi
Ordinal	Wilcoxon Matched Pairs	Mann-Whilney	Spearman Rank/ Kendal Tau
Interval/rasio	Paired Samples T-Test") (Related T-Test)	Independent Samples T-T') (T-Test Independent)	Pearson Product Moment")

*) statistisk parametrik (diperlukan uji normalitas data)

Bab 6

Analisis Multivariat

6.1 Multivariat

Analisis multivariat adalah cabang statistik yang berkaitan dengan pengamatan dan analisis lebih dari satu variabel statistik sekaligus. Berbeda dengan analisis univariat (satu variabel) atau bivariat (dua variabel), analisis multivariat menangani situasi di mana peneliti tertarik pada interaksi dan hubungan antara banyak variabel. Tujuan utama dari analisis multivariat adalah untuk memahami struktur dan pola yang mendasari data yang kompleks dan untuk membuat prediksi yang lebih akurat berdasarkan hubungan ini (Timm, 2002; Chatfield, 2018).

Beberapa tujuan utama analisis multivariat meliputi:

- Mengidentifikasi pola dan struktur data: Memahami bagaimana variabel-variabel dalam dataset berhubungan satu sama lain.
- Mengurangi dimensionalitas data: Menyederhanakan data yang kompleks dengan mengurangi jumlah variabel tanpa kehilangan informasi penting.
- Membuat prediksi: Menggunakan hubungan antar variabel untuk memprediksi nilai dari variabel lain.
- Mengelompokkan data: Mengidentifikasi kelompok atau kluster dalam data berdasarkan karakteristik yang serupa.
-

6.1.1. Ruang Lingkup Analisis Multivariat

Ruang lingkup analisis multivariat sangat luas dan mencakup berbagai teknik dan metode yang dapat diterapkan pada berbagai jenis data dan bidang penelitian. Berikut adalah beberapa teknik dan metode utama dalam analisis multivariat. Beberapa metode yang dirasa paling umum digunakan akan disampaikan pada Bab ini. Pembaca disarankan untuk merujuk pada referensi yang lebih lengkap untuk beberapa metode analisis multivariat yang tidak dibahas secara merinci pada Bab ini.

1. Analisis Faktor (*Factor Analysis*): Bertujuan untuk mengidentifikasi struktur yang mendasari variabel yang diamati dengan mengelompokkan variabel yang berkorelasi tinggi menjadi faktor-faktor yang lebih sedikit. Aplikasi: Psikologi (menganalisis konstruksi psikologis), pemasaran (memahami preferensi konsumen).
2. Analisis Komponen Utama (*Principal Component Analysis, PCA*): Bertujuan untuk mengurangi dimensionalitas data dengan mengubah variabel asli menjadi sekumpulan variabel baru (komponen utama) yang tidak berkorelasi dan memaksimalkan varians. Aplikasi: Pengolahan citra, genomika.
3. Analisis Diskriminan (*Discriminant Analysis*): Bertujuan untuk mengklasifikasikan data ke dalam kelompok berdasarkan variabel prediktor. Aplikasi: Biomedis (mendiagnosis penyakit berdasarkan gejala), keuangan (memprediksi kebangkrutan perusahaan).
4. Regresi Berganda (*Multiple Regression*): Bertujuan untuk memodelkan hubungan antara satu variabel dependen dengan beberapa variabel independen untuk membuat prediksi. Aplikasi: Ekonomi (memodelkan faktor-faktor yang mempengaruhi pendapatan), ilmu sosial (menganalisis faktor-faktor yang mempengaruhi perilaku).

5. Analisis Klaster (*Cluster Analysis*): mengelompokkan data menjadi klaster-klaster berdasarkan kemiripan antar observasi. Aplikasi: Pemasaran (segmentasi pasar), biologi (mengelompokkan spesies berdasarkan karakteristik genetik).
6. Analisis Korespondensi (*Correspondence Analysis*): Bertujuan untuk menganalisis hubungan antara dua variabel kategori dengan mereduksi dimensionalitas data kategorikal. Aplikasi: Ilmu sosial (menganalisis preferensi politik), pemasaran (analisis preferensi produk).
7. Model Persamaan Struktural (*Structural Equation Modeling, SEM*): Bertujuan untuk memodelkan hubungan kompleks antara variabel-variabel laten dan ter observasi melalui persamaan struktural. Aplikasi: Psikologi (memodelkan konstruksi teoretis), pendidikan (menganalisis faktor-faktor yang mempengaruhi kinerja siswa).

6.1.2. Contoh Penerapan Analisis Multivariat

Misalkan seorang peneliti ingin memahami faktor-faktor yang mempengaruhi kepuasan pelanggan di sebuah perusahaan ritel. Peneliti dapat menggunakan analisis faktor untuk mengidentifikasi faktor-faktor utama (seperti harga, kualitas produk, dan layanan pelanggan) yang mempengaruhi kepuasan. Selanjutnya, analisis regresi berganda dapat digunakan untuk memodelkan pengaruh masing-masing faktor terhadap tingkat kepuasan pelanggan.

6.2 Dasar-dasar Statistik untuk Analisis Multivariat

Analisis multivariat melibatkan beberapa konsep statistik yang fundamental. Memahami dasar-dasar ini sangat penting untuk melakukan dan menginterpretasikan analisis multivariat dengan benar. Penguasaan terhadap beberapa dasar-dasar yang berkaitan dengan analisis multivariat merupakan bagian penting untuk mampu melakukan analisis multivariat, bukan hanya

pada hal yang bersifat deskriptif, mau pun juga pada hal yang bersifat inferensial. Statistik dengan multi-variabel diharapkan tidak hanya mampu memberikan deskripsi data tertentu terhadap data-data yang disajikan, namun juga diharapkan mampu memprediksi terhadap kemungkinan-kemungkinan kejadian tertentu dengan tingkat kepercayaan yang signifikan. Meskipun demikian, baik mendeskripsikan maupun memprediksi dengan sesuai berdasarkan data-data yang disajikan, teknik ini tetap memiliki keterbatasan tertentu yang berkaitan dengan kemungkinan-kemungkinan batas galat (*error*), anomali, serta ketidakpastian (Ismayani, 2019; Sahir, 2021). Oleh sebab itu, deskripsi dan prediksi yang dilakukan oleh statistik multi-variabel harus dilakukan dengan hati-hati dan memberikan informasi yang didapatkan secara keseluruhan, demi menghindari beberapa kesalahan-kesalahan pengambilan keputusan dikarenakan kekurangan-kekurangan tertentu dari hasil olah dan analisis data yang sudah dilakukan.

6.2.1. Pengertian variabel dan jenisnya

Variabel merupakan entitas yang dapat mengambil berbagai nilai. Dalam analisis multivariat, kita biasanya bekerja dengan beberapa variabel secara bersamaan. Dalam statistik setidaknya terdapat dua jenis variabel, yaitu variabel independen (*predictor*) dan variabel dependen (*responses*). Variabel independen merupakan variabel yang digunakan untuk memprediksi, sementara variabel dependen merupakan variabel yang diprediksi. Dalam kasus tertentu, meskipun tidak sebagaimana sering variabel independen dan dependen, terdapat pula istilah variabel interdependen, di mana variabel ini saling mempengaruhi satu sama lain, tanpa hierarki yang jelas antara *predictor* atau *response*.

Dalam kasus sosial, variabel dapat berupa metode pemasaran (variabel independen) – kinerja pemasaran (variabel dependen), yaitu pada suatu penelitian yang melibatkan kedua

variabel tersebut, metode pemasaran memprediksi kinerja pemasaran suatu produk. Dalam kasus lain di bidang teknik, misalnya, dapat berupa parameter-parameter. Misalnya, kuat arus (variabel independen) – kekuatan las (variabel dependen), yang mana dalam penelitian tersebut, seorang peneliti menggunakan kuat arus pada proses pengelasan untuk memprediksi kekuatan sambungan dua buah bahan yang digabungkan dengan proses pengelasan. Terdapat berbagai contoh judul artikel yang biasanya dalam banyak kasus menggambarkan kedua jenis variabel tersebut dan hubungan masing-masing. Misalnya, pada judul penelitian “Pengaruh Variasi Komposisi Tempurung Kelapa dan Cangkang Kemiri terhadap Kualitas Briket Alternatif” (dapat diakses di tautan: <https://doi.org/10.24176/crankshaft.v7i1.11996>) variasi komposisi antara tempurung kelapa dan cangkang kemiri merupakan variabel independen, sedangkan kualitas briket merupakan variabel dependen.

6.2.2. Korelasi dan Kovariansi

Korelasi merupakan ukuran statistik yang menunjukkan seberapa kuat hubungan linear antara dua variabel (Roflin & Riana, 2022). Nilai korelasi (r) berkisar antara -1 hingga 1. Dalam hal nilai korelasi, jika nilai korelasi > 0 , maka di antara variabel tersebut memiliki korelasi positif, yaitu jika pada suatu variabel independen terjadi peningkatan, maka peningkatan juga terjadi pada variabel dependen. Sementara itu, jika nilai korelasi < 0 , maka korelasi bernilai negatif, yaitu jika pada suatu variabel independen terjadi peningkatan, maka variabel dependen akan mengalami penurunan. Sementara itu, jika nilai korelasi $= 0$, maka hal ini menunjukkan bahwa tidak ada hubungan linier antara variabel independen dan dependen.

Sementara itu, kovariansi merupakan ukuran statistik yang menunjukkan arah hubungan antara dua variabel. Metrik ini mengevaluasi seberapa banyak – sejauh mana – variabel

berubah bersama-sama. Dapat juga dikatakan bahwa nilai kovarian ini mengukur variansi antara dua variabel. Meskipun demikian, kovariansi tidak dapat menilai ketergantungan antara dua variabel. Nilai kovarian antara dua variabel x dan y dinyatakan sebagai:

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \tag{1}$$

Nilai kovariansi positif menunjukkan bahwa kedua buah variabel bergerak ke arah yang sama. Sedangkan, nilai kovariansi negatif mengungkapkan bahwa dua variabel cenderung bergerak ke arah yang berlawanan.

6.2.3 Teori Matriks

Dalam banyak kasus, analisis multivariat juga melibatkan variabel-variabel yang disusun dan dioperasikan dalam bentuk matriks. Oleh sebab itu, pemahaman mengenai beberapa aspek pada matriks dianggap perlu sebagai dasar untuk melakukan analisis multivariat.

Matriks merupakan susunan bilangan dalam bentuk baris dan kolom, yang biasanya berukuran tertentu. Misalnya, suatu matriks A memiliki ukuran $m \times n$, di mana m adalah baris dan n adalah kolom, seperti:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{2}$$

Dalam suatu matriks dikenal dengan konsep determinan. Determinan merupakan suatu nilai skalar yang dihubungkan oleh suatu matriks berbentuk persegi, misalnya nilai $m = n = 2$, yaitu matriks dengan jumlah 2 baris dan 2 kolom. Dalam kasus tersebut, determinan matriksnya dapat dihitung dengan:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{3}$$

Konsep invers matriks juga merupakan konsep yang banyak digunakan. Invers matriks adalah suatu matriks yang jika

dikalikan dengan matriks asli maka akan menghasilkan matriks identitas, atau dapat ditulis sebagai berikut:

$$AA^{-1} = I \quad (4)$$

di mana I adalah invers matriks.

Satu konsep matriks lainnya yang digunakan dalam analisis multivariat yaitu *eigenvalue* dan *eigenvector*. Jika A adalah matriks $n \times n$, maka skalar λ dan vektor bukan nol v yang memenuhi persamaan:

$$Av = \lambda v \quad (5)$$

disebut masing-masing *eigenvalue* dan *eigenvetor* dari A.

6.2.4 Analisis Asumsi Dasar dan Persiapan Data

Terdapat dua hal yang biasa dilakukan untuk menganalisis data-data yang akan digunakan untuk analisis multi-variabel. Hal seperti transformasi dan pembersihan data serta normalisasi data merupakan hal yang lumrah untuk dilakukan. Meskipun bukan merupakan suatu keharusan, langkah ini dianggap penting, utamanya ketika melakukan pemrosesan data yang bervariasi dan dalam jumlah yang besar.

Dalam pembersihan dan transformasi data hal-hal seperti normalitas multivariat, homoskedastisitas, dan linearitas merupakan hal yang perlu diketahui nilai-nilainya, untuk memastikan bahwa analisis multivariat yang dilakukan memiliki tingkat kepercayaan data yang tinggi, di luar dari validitas dan reliabilitas data. Normalitas multivariat asumsi bahwa kombinasi linier dari variabel-variabel dalam dataset mengikuti distribusi normal multivariat. Hal ini berarti distribusi gabungan dari semua variabel dalam dataset berbentuk lonceng yang simetris di sekitar nilai rerata. Sedangkan, homoskedastisitas berasumsi bahwa variansi residual (kesalahan prediksi) adalah konstan di seluruh rentang nilai variabel independen pada data yang tersedia. Dalam konteks data multivariat, hal ini berarti

bahwa variansi residual dari semua kombinasi linier variabel independen adalah konstan. Yang terakhir, linearitas adalah asumsi bahwa hubungan antara variabel dependen dan setiap variabel independen dapat dijelaskan dengan baik oleh model linier, misalnya regresi linier. Dalam konteks multivariat, hal ini berarti bahwa hubungan antara variabel dependen dan kombinasi linier dari variabel independen adalah linier.

Dalam analisis multivariat, merupakan hal yang penting untuk melakukan pengujian asumsi klasik. Pengujian asumsi klasik merujuk pada pengujian-pengujian terhadap data yang digunakan untuk analisis multivariat, di mana data-data tersebut diharuskan memenuhi kriteria tertentu, juga terhindar dari beberapa kriteria lainnya (Sugiyono, 2013). Beberapa asumsi tersebut meliputi:

- Uji normalitas, yaitu variabel-variabel harus mengikuti distribusi normal multivariat.
- Uji linearitas, yaitu hubungan antar variabel harus linear.
- Uji homoskedastisitas, yaitu variansi dari residual harus konstan.
- Uji multikolinearitas, yaitu variabel independen tidak boleh sangat berkorelasi satu sama lain.

Data yang nantinya digunakan untuk analisis multivariat mesti melalui pengujian yang teliti melalui pengujian asumsi klasik. Beberapa aspek pengujian asumsi klasik juga masuk pada analisis data multivariat, namun pengujian asumsi klasik dinilai merupakan pengujian yang lebih luas dan *robust*. Pengujian seperti ini bertujuan agar kesimpulan yang didapatkan dari analisis multivariat memiliki nilai akurasi dan reliabilitas yang tinggi, serta memberikan tingkat kepercayaan yang tinggi pada validitas pengujian.

6.3 Analisis Faktor (*Factor Analysis*)

6.3.1. Teori Dasar

Analisis faktor adalah teknik statistik multivariat yang digunakan untuk mengidentifikasi struktur dasar dalam satu set data variabel yang besar (Kline, 2014). Tujuan utama analisis faktor adalah untuk menemukan variabel yang tidak dapat diamati (dikenal sebagai faktor) yang dapat menjelaskan pola korelasi atau kovariansi di antara variabel yang diamati. Analisis faktor didasarkan pada asumsi bahwa sejumlah variabel dapat dipengaruhi oleh faktor-faktor yang mendasarinya. Dengan kata lain, variasi yang diamati dalam variabel dapat dijelaskan oleh sejumlah kecil faktor. Analisis faktor adalah alat statistik yang kuat untuk mengidentifikasi struktur laten dalam data multivariat. Dengan memahami dasar teori, proses analisis, dan interpretasi hasil, peneliti dapat mengaplikasikan analisis faktor untuk mengurangi dimensionalitas data, mengidentifikasi faktor-faktor yang mendasari variabel yang diamati, dan menghasilkan wawasan yang lebih dalam tentang struktur data tersebut.

Beberapa komponen utama dalam suatu analisis faktor yaitu:

1. Faktor (*Factor*) adalah variabel laten yang tidak dapat diamati secara langsung tetapi mempengaruhi beberapa variabel yang diamati. Faktor biasanya diwakili oleh huruf Yunani seperti ξ atau η .
2. Loading Faktor (*Factor Loading*) menunjukkan seberapa kuat variabel teramati berkorelasi dengan faktor laten, biasanya dinyatakan dalam matriks loading L .
3. Unik (*Unique Factor*) adalah komponen dari variabel yang diamati yang tidak dapat dijelaskan oleh faktor laten. Variabel unik biasanya diwakili oleh huruf ϵ dan disebut juga sebagai kesalahan pengukuran.

Dalam menggunakan analisis faktor pada analisis data multivariat, beberapa asumsi dasar harus terpenuhi terlebih dahulu. Beberapa asumsi tersebut meliputi:

1. Linearitas yang menyatakan hubungan antara variabel yang diamati dan faktor laten diasumsikan bersifat linier.
2. Normalitas multivariat, yaitu data variabel yang diamati diasumsikan mengikuti distribusi normal multivariat.
3. Independensi kesalahan pengukuran, yang berhubungan dengan kesalahan pengukuran, di mana kesalahan pengukuran diasumsikan tidak berkorelasi dengan satu sama lain dan dengan faktor laten.

6.3.2. Proses Analisis Faktor

Jika seorang peneliti bermaksud menggunakan analisis faktor untuk mendapatkan wawasan mengenai bagaimana struktur data multivariat pada suatu data yang besar, dan apakah terdapat faktor-faktor tertentu yang mempengaruhi secara mendasar pada sejumlah variabel, maka perlu dilakukan suatu analisis faktor.

1. Ekstraksi Faktor:

- Metode Analisis Komponen Utama (*Principal Component Analysis, PCA*): Metode ini mengekstraksi faktor dengan cara memaksimalkan varian total yang dijelaskan oleh faktor-faktor utama (akan dijelaskan kemudian).
- Metode Analisis Faktor (*Factor Analysis, FA*): Metode ini mengekstraksi faktor dengan cara memaksimalkan varian bersama (*shared variance*) yang dijelaskan oleh faktor-faktor.

2. Rotasi Faktor:

- Setelah faktor diekstraksi, faktor tersebut sering kali dirotasi untuk mendapatkan struktur faktor yang lebih mudah diinterpretasikan.

- Rotasi *Orthogonal* (misalnya, *Varimax*): Menjaga faktor tetap ortogonal (tidak berkorelasi).
- Rotasi *Oblique* (misalnya, *Promax*): Memungkinkan faktor berkorelasi satu sama lain.

3. Interpretasi Faktor:

- Loading faktor yang signifikan digunakan untuk menafsirkan makna dari masing-masing faktor.
- Faktor biasanya diberi label berdasarkan variabel-variabel yang memiliki nilai *loading* tertinggi pada faktor tersebut.

6.3.3. Persamaan dalam Analisis Faktor

Secara umum, model analisis faktor dapat dinyatakan sebagai:

$$\mathbf{X} = \mathbf{LF} + \mathbf{U} \quad (6)$$

di mana:

- \mathbf{X} adalah vektor variabel yang diamati.
- \mathbf{L} adalah matriks *loading* faktor.
- \mathbf{F} adalah vektor faktor laten.
- \mathbf{U} adalah vektor kesalahan pengukuran.

6.3.4. Contoh Penerapan Analisis Faktor

Misalkan kita memiliki dataset dari kuesioner psikologi yang mengukur lima variabel: Kepuasan Kerja (X_1), Stres Kerja (X_2), Keterlibatan Kerja (X_3), Dukungan Sosial (X_4), dan Kepuasan Hidup (X_5). Kita ingin mengetahui apakah ada faktor-faktor yang mendasari pola korelasi antara variabel-variabel ini.

Langkah-langkah Analisis Faktor:

- Persiapan Data
- Matriks Korelasi
- Ekstraksi Faktor

- Rotasi Faktor
- Interpretasi Faktor

1. Persiapan Data

Kita mulai dengan data mentah dari kuesioner:

Tabel 6.1. Contoh tabel hasil kuesioner.

Responden	X1	X2	X3	X4	X5
1	4.2	3.8	4.5	3.1	4.6
2	3.9	4.0	4.2	3.0	4.3
3	4.5	3.7	4.8	3.2	4.7
4	4.0	3.9	4.3	3.1	4.5
5	4.3	3.6	4.6	3.3	4.6

2. Matriks Korelasi

Langkah pertama adalah menghitung matriks korelasi antar variabel.

$$R = \begin{pmatrix} 1.00 & -0.10 & 0.85 & 0.50 & 0.90 \\ -0.10 & 1.00 & -0.15 & 0.70 & -0.20 \\ 0.85 & -0.15 & 1.00 & 0.55 & 0.95 \\ 0.50 & 0.70 & 0.55 & 1.00 & 0.60 \\ 0.90 & -0.20 & 0.95 & 0.60 & 1.00 \end{pmatrix}$$

3. Ekstraksi Faktor

Menggunakan metode Analisis Komponen Utama (PCA), kita menghitung *eigenvalues* dan *eigenvectors* dari matriks korelasi untuk mengekstraksi faktor-faktor. Katakanlah kita menemukan dua faktor utama yang memiliki *eigenvalues* lebih besar dari 1.

$$L = \begin{pmatrix} 0.80 & 0.40 \\ -0.30 & 0.85 \\ 0.90 & 0.20 \\ 0.60 & 0.70 \\ 0.95 & 0.10 \end{pmatrix}$$

4. Rotasi Faktor

Untuk mendapatkan hasil yang lebih interpretatif, kita melakukan rotasi *orthogonal* menggunakan metode *Varimax*.

$$L_{\text{rotated}} = \begin{pmatrix} 0.75 & 0.45 \\ -0.25 & 0.80 \\ 0.88 & 0.15 \\ 0.55 & 0.75 \\ 0.92 & 0.05 \end{pmatrix}$$

5. Interpretasi Faktor

Dari matriks *loading* faktor yang telah dirotasi, kita dapat melihat pola hubungan antara variabel yang diamati dan faktor laten. Misalnya:

- Faktor 1 (Kesejahteraan Umum): Variabel yang memiliki *loading* tinggi pada faktor ini adalah Kepuasan Kerja (0.75), Keterlibatan Kerja (0.88), dan Kepuasan Hidup (0.92).
- Faktor 2 (Stres dan Dukungan)**: Variabel yang memiliki *loading* tinggi pada faktor ini adalah Stres Kerja (0.80) dan Dukungan Sosial (0.75).

Dengan menggunakan analisis faktor, kita dapat mengidentifikasi dua faktor laten yang mendasari pola korelasi antara lima variabel psikologi kerja. Interpretasi faktor ini membantu kita memahami dimensi yang mendasari data, memungkinkan kita untuk merancang intervensi yang lebih efektif untuk meningkatkan kesejahteraan karyawan dan mengelola stres kerja. Analisis faktor ini juga membantu dalam reduksi dimensionalitas, membuat data yang kompleks lebih mudah dipahami dan dianalisis

Contoh Interpretasinya adalah sebagai berikut:

- Kesejahteraan Umum: Faktor ini mencerminkan aspek-aspek positif dari kesejahteraan karyawan, termasuk kepuasan kerja, keterlibatan kerja, dan kepuasan hidup. *Loading* faktor yang tinggi menunjukkan bahwa variabel-variabel ini sangat berkaitan satu sama lain dan mungkin dipengaruhi oleh faktor yang sama.
- Stres dan Dukungan: Faktor ini mencerminkan aspek-aspek negatif dan dukungan dalam lingkungan kerja. Stres kerja dan dukungan sosial memiliki *loading* faktor yang signifikan pada faktor ini, menunjukkan bahwa variabel-variabel ini saling berkaitan dan dapat dipengaruhi oleh faktor umum terkait stres dan dukungan di tempat kerja.

6.3.5. Kelebihan dan Kekurangan Analisis Faktor

Kelebihan analisis faktor di antaranya:

- Reduksi Dimensionalitas: Mengurangi jumlah variabel menjadi beberapa faktor yang lebih mudah diinterpretasikan.
- Identifikasi Struktur Laten: Mengidentifikasi faktor-faktor yang mendasari variabel-variabel yang diamati.

Sementara itu, keterbatasan dalam penggunaan analisis faktor adalah sebagai berikut:

- Asumsi Linearitas dan Normalitas: Analisis faktor bergantung pada asumsi linearitas dan normalitas yang mungkin tidak selalu terpenuhi.
- Kesulitan dalam Interpretasi: Interpretasi faktor dapat subjektif dan terkadang sulit, terutama jika *loading* faktor tidak jelas.

6.4 Analisis Komponen Utama (*Principal Component Analysis* – PCA)

6.4.1. Pengantar

Analisis Komponen Utama (*Principal Component Analysis*, disingkat PCA) adalah teknik statistik yang digunakan untuk mereduksi dimensionalitas data, sambil mempertahankan sebanyak mungkin variabilitas dalam data (Abdi & Williams, 2010; Greeacre, dkk., 2022; Wold, dkk., 1987). PCA mengubah variabel-variabel asli yang saling berkorelasi menjadi sekumpulan variabel baru yang tidak berkorelasi yang disebut sebagai komponen utama. Teknik ini sangat berguna dalam menganalisis dataset dalam jumlah besar dan kompleks dalam berbagai bidang seperti genetika, ilmu lingkungan, dan pengenalan pola (*pattern recognition*). PCA merupakan teknik statistik yang dinilai *robust* untuk mereduksi dimensionalitas data dan mengidentifikasi struktur laten dalam data multivariat. Dengan memahami proses PCA, dari standardisasi data hingga interpretasi komponen utama, peneliti dapat menerapkan teknik ini untuk berbagai aplikasi dalam sains dan teknologi. PCA tidak hanya membantu dalam mengelola data dalam jumlah yang besar dan kompleks tetapi juga memberikan wawasan yang lebih dalam tentang pola dan hubungan yang mendasari data tersebut.

PCA didasarkan pada konsep bahwa data yang mengandung banyak variabel mungkin memiliki struktur internal yang dapat diekspresikan dengan lebih sedikit dimensi tanpa kehilangan informasi yang signifikan. Proses PCA melibatkan transformasi linear dari variabel-variabel asli ke variabel baru yang disebut komponen utama. Komponen-komponen utama ini diurutkan berdasarkan variansi data, di mana komponen pertama menjelaskan varian terbesar, komponen kedua menjelaskan varian terbesar berikutnya, dan seterusnya.

6.4.2. Proses PCA

Dalam penggunaannya, PC memiliki beberapa tahap proses yang perlu dilakukan, yaitu meliputi standardisasi data, analisis matriks kovarians dan korelasi, ekstraksi komponen utama, serta pembentukan komponen utama.

1. *Standardisasi Data*

Langkah pertama dalam PCA adalah standardisasi variabel, terutama jika data yang diukur berada pada skala yang berbeda. Proses standardisasi data dilakukan dengan mengurangi rerata dan membagi dengan standar deviasi untuk setiap variabel, sehingga menghasilkan variabel dengan rerata nol dan standar deviasi satu. Standardisasi ini dianggap merupakan tahap yang penting, dikarenakan dalam beberapa kasus, perbedaan rentang antara satu variabel komponen dengan yang lainnya dapat mempengaruhi pembobotan pada komponen utama. Misalnya, variabel X_1, X_2, \dots, X_p distandardisasi menjadi Z_1, Z_2, \dots, Z_3 .

2. *Matriks Kovarians atau Korelasi*

Setelah data distandardisasi, langkah berikutnya adalah menghitung matriks kovarians atau matriks korelasi antar variabel. Matriks ini mengukur seberapa besar variabel-variabel berkorelasi satu sama lain. Dapat dilihat, bahwa data yang tidak dilakukan standardisasi dengan baik akan mengalami keterbatasan-keterbatasan tertentu dalam proses penghitungan matriks kovarian. Misalnya, menghitung matriks kovarian (\mathbf{S}) atau matriks korelasi (\mathbf{R}) dari variabel-variabel yang sudah dibuat standardisasinya

3. *Ekstraksi Komponen Utama*

Komponen utama diekstraksi dengan menghitung *eigenvalues* dan *eigenvectors* dari matriks kovarians atau korelasi. *Eigenvectors* menunjukkan arah komponen utama,

sementara *eigenvalues* menunjukkan varian yang dijelaskan oleh setiap komponen utama. Dua nilai ini sangat menentukan proses PCA, sehingga perlu dilakukan dengan sangat hati-hati. Misalnya, kita bisa menemukan nilai *eigenvalues* $(\lambda_1, \lambda_2, \dots, \lambda_p)$ dan *eigenvectors* $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p)$ dari matriks kovarians atau korelasi. Harap diperhatikan, sampai di sini, bahwa proses PCA perlu dilakukan berurutan untuk menghasilkan analisis yang tepat.

4. Pembentukan Komponen Utama

Komponen utama dibentuk sebagai kombinasi linear dari variabel asli, menggunakan *eigenvectors* sebagai koefisien. Komponen utama pertama (PC_1) adalah kombinasi linear yang memaksimalkan varian total dalam data. Komponen utama kedua (PC_2) adalah kombinasi linear yang memaksimalkan varian yang tersisa, dengan syarat ortogonal terhadap (PC_1), dan seterusnya. Komponen utama pertama $PC_1 = \mathbf{Z} \cdot \mathbf{e}_1$, sedangkan komponen utama kedua $PC_2 = \mathbf{Z} \cdot \mathbf{e}_2$, dan seterusnya.

6.4.3. Persamaan Matematis dan Contoh Kasus

1. Persamaan matematis

Jika \mathbf{X} adalah matriks data yang telah distandarisasi dengan dimensi $n \times p$ (di mana n adalah jumlah pengamatan dan p adalah jumlah variabel), maka PCA dapat dinyatakan sebagai:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \tag{7}$$

di mana:

- \mathbf{T} adalah matriks skor komponen utama (dimensi $n \times p$)
- \mathbf{P} adalah matriks *loading* komponen utama (dimensi $n \times p$)

2. *Contoh kasus*

Misalkan kita memiliki dataset psikologi dengan lima variabel: Kepuasan Kerja (X_1), Stres Kerja (X_2), Keterlibatan Kerja (X_3), Dukungan Sosial (X_4), dan Kepuasan Hidup (X_5). Setelah dilakukan 4 proses tersebut, misalnya didapatkan data tertentu. Misalkan hasil PCA menunjukkan bahwa dua komponen utama pertama menjelaskan 85% dari total variansi dalam data. Matriks faktor *loading* yang diperoleh setelah rotasi (misalnya metode *Varimax*) dapat terlihat seperti berikut:

Tabel 6.2. Contoh nilai matriks faktor *loading*.

Variabel	Komponen 1	Komponen 2
1	4.2	3.8
2	3.9	4.0
3	4.5	3.7
4	4.0	3.9
5	4.3	3.6

3. Interpretasi:

- Komponen 1 (Kesejahteraan Psikologis): Variabel yang memiliki *loading* tinggi adalah Kepuasan Kerja (0.80), Keterlibatan Kerja (0.85), dan Kepuasan Hidup (0.90).
- Komponen 2 (Stres dan Dukungan Sosial): Variabel yang memiliki *loading* tinggi adalah Stres Kerja (0.70) dan Dukungan Sosial (0.75).

Dengan demikian, kita sudah dapat melakukan analisis data multivariat dengan teknik PCA, di mana beberapa faktor-faktor tertentu berhasil dianalisis dan diinterpretasikan berdasarkan komponen-komponen utamanya.

6.4.4. Kelebihan dan Keterbatasan PCA

1. Kelebihan

- Reduksi Dimensionalitas: PCA secara efektif mengurangi jumlah variabel tanpa kehilangan banyak informasi.
- Identifikasi Struktur Data: PCA membantu mengidentifikasi pola dan struktur dalam data yang kompleks.
- Visualisasi: PCA memungkinkan visualisasi data dalam dua atau tiga dimensi.

2. Kekurangan/keterbatasan

- Asumsi Linearitas: PCA mengasumsikan bahwa hubungan antar variabel adalah linier, sehingga analisis hubungan antar variabel yang non-linier akan mengalami kesulitan atau bahkan tidak mungkin dengan menggunakan metode ini. Dalam banyak kasus, beberapa metode tambahan perlu dilakukan dan disesuaikan jika memang dirasa perlu menggunakan metode PCA.
- Interpretasi Faktor: Interpretasi komponen utama mungkin tidak selalu jelas atau mudah. Dalam banyak kasus, dimungkinkan juga interpretasi komponennya terkesan kabur. Oleh sebab itu, beberapa metode alternatif lainnya dirasa diperlukan untuk menguji reliabilitas dan validitas data.
- Sensitivitas terhadap Skala: Variabel dengan skala yang berbeda harus dilakukan standarisasi sebelum melakukan PCA. Hal ini akan memperpanjang proses analisis, yang mana terhitung tidak praktis.

6.5 Regresi Berganda (*Multiple Regression*)

6.5.1. Pengantar

Regresi berganda adalah teknik statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen (Kelley & Bolin, 2013). Tujuannya adalah untuk memahami bagaimana variabel dependen berubah sebagai respons terhadap perubahan dalam variabel-variabel independen, serta untuk memprediksi nilai variabel dependen berdasarkan nilai variabel-variabel independen. Regresi berganda adalah alat statistik yang dinilai sangat *robust* untuk menganalisis dan memprediksi hubungan antara satu variabel dependen dan beberapa variabel independen. Melalui proses *fitting* model, evaluasi, dan interpretasi hasil, peneliti dapat memahami lebih dalam tentang pengaruh variabel-variabel independen terhadap variabel dependen. Dengan memperhatikan asumsi-asumsi yang mendasari regresi berganda, analisis ini dapat digunakan untuk membuat keputusan yang lebih informatif dan akurat dalam berbagai bidang seperti ekonomi, ilmu sosial, dan ilmu alam.

6.5.2. Dasar Teori

Regresi berganda memperluas konsep regresi linear sederhana, di mana hanya ada satu variabel independen, menjadi kasus di mana ada beberapa variabel independen. Model regresi berganda dinyatakan dengan persamaan linear sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (8)$$

di mana:

- Y adalah variabel dependen.
- X_1, X_2, \dots, X_p adalah variabel independen.
- β_0 adalah *intercept* (titik potong dengan sumbu Y).

- $\beta_1, \beta_2, \dots, \beta_p$ adalah koefisien regresi yang menunjukkan pengaruh masing-masing variabel independen terhadap variabel dependen.
- ϵ adalah *error terms* yang mencakup variabilitas yang tidak dapat dijelaskan oleh model.

Sama seperti PCA, regresi berganda juga memiliki beberapa asumsi tertentu, di antaranya:

1. *Linearitas*: Hubungan antara variabel dependen dan variabel independen adalah linear.
2. *Independensi*: Observasi-observasi dalam dataset bersifat independen satu sama lain.
3. *Homoskedastisitas*: Varians dari *error terms* adalah konstan untuk semua pengamatan.
4. *Normalitas*: *Error terms* mengikuti distribusi normal.
5. *Tidak Ada Multikolinearitas*: Variabel independen tidak berkorelasi tinggi satu sama lain.

Dalam penggunaannya, regresi berganda dilakukan dengan melalui beberapa tahap, yaitu pengumpulan data, eksplorasi data, *fitting* model, evaluasi model, dan interpretasi hasil. Secara lebih terperinci, proses-proses tersebut dapat dijabarkan sebagai berikut.

1. *Pengumpulan Data*: Kumpulkan data yang relevan untuk variabel dependen dan independen. Pastikan data berkualitas dan bebas dari kesalahan. Data dapat berasal dari data primer maupun sekunder. Kedua jenis data tersebut tidak berpengaruh terhadap kualitas data. Namun demikian, untuk mendapatkan interpretasi yang baik, yang mana data tersebut mampu menggambarkan keadaan populasi, maka data tersebut mesti dikumpulkan metode sampling data yang tepat.

2. Eksplorasi Data: Eksplorasi data melibatkan deskriptif statistik, visualisasi, dan pengecekan asumsi-asumsi regresi. Ini termasuk plot *scatter*, korelasi, dan uji multikolinearitas menggunakan *Variance Inflation Factor* (VIF). Eksplorasi data memastikan bahwa data yang dikumpulkan dapat digunakan untuk proses lebih lanjut dalam pengembangan model, yang biasanya akan berkaitan erat dengan validitas dan reliabilitas data.
3. *Fitting Model*: *Fitting* model regresi berganda melibatkan estimasi parameter β dengan menggunakan metode kuadrat terkecil (*Least Squares Method*), yang meminimalkan jumlah kuadrat dari residuals (selisih antara nilai yang diamati dan nilai yang diprediksi).
4. Evaluasi Model: Model dievaluasi menggunakan berbagai metrik seperti koefisien determinasi (R^2), adjusted (R^2), uji F untuk keseluruhan model, dan uji t untuk koefisien individu. Residual plots juga digunakan untuk memeriksa asumsi-asumsi berdasarkan model yang sudah tersusun dari data tersebut.
5. Interpretasi Hasil: Koefisien regresi diinterpretasikan untuk memahami pengaruh masing-masing variabel independen terhadap variabel dependen. Signifikansi statistik dari koefisien diuji untuk memastikan bahwa pengaruh yang diamati bukan kebetulan, yaitu terdapat konsistensi dan akurasi dari hasil yang didapatkan.

6.5.3. Regresi Berganda pada Data Penjualan

Misalkan kita ingin memodelkan penjualan (Y) berdasarkan anggaran iklan di TV (X_1), radio (X_2), dan surat kabar (X_3). Data untuk 10 pengamatan pertama adalah sebagai berikut:

Tabel 6.3. Contoh data hasil penjualan berdasarkan anggaran iklan.

ID	TV (X1)	Radio (X2)	Surat Kabar (X3)	Penjualan (Y)
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9
6	8.7	48.9	75	7.2
7	57.5	32.8	23.5	11.8
8	120.2	19.6	11.6	13.2
9	8.6	2.1	1	4.8
10	199.8	2.6	21.2	10.6

Model regresi berganda yang dapat dibuat dengan data tersebut, yaitu sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \tag{9}$$

Menggunakan metode kuadrat terkecil, kita estimasi parameter β :

$$Y = 2.937 + 0.045 \times X_1 + 0.188 \times X_2 - 0.001 \times X_3$$

Dengan model tersebut, terdapat beberapa hal yang dapat dievaluasi, yaitu meliputi:

1. Koefisien Determinasi (R^2): Mengukur proporsi variasi dalam penjualan yang dapat dijelaskan oleh variabel-variabel independen. Misalkan, $R^2 = 0.897$, berarti 89.7% variasi dalam penjualan dijelaskan oleh model.
2. *Adjusted R²* : Mengoreksi R^2 dengan mempertimbangkan jumlah variabel independen. *Adjusted R²* biasanya sedikit lebih rendah dari nilai R^2 awal.
3. Uji F: Menilai signifikansi keseluruhan model. Misalkan, hasil uji F menunjukkan nilai $p < 0.05$, berarti model signifikan secara statistik.
4. Uji t : Menguji signifikansi masing-masing koefisien. Misalkan, β_1 dan β_2 signifikan ($p < 0.05$), tetapi β_3 tidak signifikan ($p > 0.05$).

Berdasarkan evaluasi tersebut, maka dapat diinterpretasikan beberapa hal, yaitu:

1. *Intercept* ($\beta_0 = 2.937$): Ketika semua variabel independen bernilai nol, nilai penjualan diperkirakan sebesar 2.937.
2. Koefisien $\beta_1 = 0.045$: Setiap peningkatan \$1 dalam anggaran TV terkait dengan peningkatan 0.045 unit dalam penjualan, dengan asumsi variabel lainnya konstan.
3. Koefisien $\beta_2 = 0.188$: Setiap peningkatan \$1 dalam anggaran radio terkait dengan peningkatan 0.188 unit dalam penjualan, dengan asumsi variabel lainnya konstan.
4. Koefisien $\beta_3 = -0.001$: Setiap peningkatan \$1 dalam anggaran surat kabar terkait dengan penurunan 0.001 unit dalam penjualan, tetapi efek ini tidak signifikan secara statistik.

6.6 Model Persamaan Struktural (*Structural Equation Modeling* - SEM)

6.6.1. Pengantar

Structural Equation Modeling (SEM) adalah teknik statistik multivariat yang menggabungkan aspek regresi berganda, analisis faktor, dan analisis jalur untuk memodelkan hubungan antara variabel-variabel yang diamati dan laten (tidak terukur secara langsung). SEM memungkinkan peneliti untuk menguji dan memperkirakan hubungan kompleks antar variabel serta memvalidasi model teoritis yang melibatkan beberapa variabel dependen dan independen sekaligus (Mueller & Hancock, 2018). SEM adalah metode yang dinilai sangat *robust* dan fleksibel untuk memodelkan hubungan yang kompleks antara variabel independen dan variabel dependen yang diamati. Gabungan elemen-elemen dari regresi, analisis faktor, dan analisis jalur pada SEM memungkinkan peneliti untuk menguji model teoritis secara komprehensif. Penerapan SEM dalam berbagai bidang penelitian telah memberikan wawasan yang lebih dalam dan validasi yang lebih kuat terhadap hipotesis penelitian yang diajukan oleh peneliti.

6.6.2. Dasar Teori

SEM mencakup dua komponen utama: model pengukuran dan model struktural. Model pengukuran (*Measurement Model*) mencakup hubungan antara variabel-variabel laten (variabel independen) dan indikator-indikatornya (variabel dependen yang diamati). Model pengukuran digunakan untuk menilai validitas dan reliabilitas konstruk-konstruk laten. Sedangkan, model struktural (*Structural Model*) mencakup hubungan kausal antara variabel-variabel laten. Model struktural menggambarkan arah dan kekuatan pengaruh antara konstruk-konstruk laten dalam model.

Terdapat tiga komponen SEM, yaitu meliputi:

1. Variabel Laten (*Latent Variables*): Variabel yang tidak dapat diukur secara langsung, tetapi diperkirakan melalui beberapa indikator yang diamati.
2. Variabel yang Diamati (*Observed Variables*): Variabel yang dapat diukur secara langsung dan digunakan sebagai indikator untuk variabel laten.
3. Jalur (*Paths*): Representasi dari hubungan kausal antara variabel-variabel dalam model.

Model pengukuran untuk variabel laten eksogen (ξ) dan endogen (η) dapat dinyatakan sebagai:

$$\mathbf{X} = \Lambda_x \xi + \delta \quad (10)$$

$$\mathbf{Y} = \Lambda_y \eta + \varepsilon \quad (11)$$

di mana

- \mathbf{X} adalah vektor indikator untuk variabel eksogen.
- \mathbf{Y} adalah vektor indikator untuk variabel endogen.
- Λ_x dan Λ_y adalah matriks *loading factor*.
- δ dan ε adalah *error terms*.

Sedangkan, model struktural dapat dinyatakan sebagai:

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta \quad (12)$$

di mana

- Γ adalah matriks koefisien yang menghubungkan variabel eksogen dengan variabel endogen.
- \mathbf{B} adalah matriks koefisien yang menghubungkan variabel endogen dengan variabel endogen
- ζ adalah *error terms*.

6.6.3. Proses SEM

Terdapat beberapa tahap dalam melakukan SEM, yang meliputi beberapa hal sebagai berikut:

1. Pengembangan Model Teoritis: Langkah pertama adalah mengembangkan model teoritis berdasarkan literatur dan hipotesis penelitian. Model ini mengidentifikasi variabel laten, variabel yang diamati, dan hubungan antar variabel. Harap berhati-hati dalam pengembangan model teoretis, dikarenakan kesalahan dalam mengidentifikasi variabel-variabel, serta hubungannya, maka akan membuat analisis selanjutnya mengalami kesulitan atau bahkan terjadi kesalahan.
2. Spesifikasi Model: Spesifikasi model melibatkan penentuan variabel laten dan indikatornya, serta jalur kausal yang menghubungkan variabel-variabel tersebut. Hal ini dinyatakan dalam bentuk diagram jalur (*path diagram*).
3. Identifikasi Model: Model harus diidentifikasi, yang berarti jumlah parameter yang akan diestimasi harus kurang dari atau sama dengan jumlah informasi yang tersedia. Identifikasi biasanya diperiksa dengan derajat kebebasan (*degrees of freedom*).
4. Estimasi Model: Parameter model diestimasi menggunakan metode seperti *Maximum Likelihood Estimation* (MLE). Metode ini melibatkan perhitungan matriks kovarian yang diharapkan dan membandingkannya dengan matriks kovarian yang diamati.
5. Evaluasi Model: Model dievaluasi menggunakan berbagai indeks kecocokan (*fit indices*) seperti:
 - a. Chi-square (χ^2)
 - b. *Root Mean Square Error of Approximation* (RMSEA)
 - c. *Comparative Fit Index* (CFI)

- d. *Tucker-Lewis Index (TLI)*
- e. *Standardized Root Mean Square Residual (SRMR)*

6.6.4. Contoh penggunaan SEM

Misalkan kita ingin memodelkan pengaruh kepuasan kerja (*job satisfaction*) terhadap kinerja karyawan (*employee performance*) dengan variabel mediasi motivasi (*motivation*). Model teoritis yang diajukan adalah:

- Kepuasan kerja (*latent variable: JS*) diukur oleh tiga indikator: gaji (X_1), lingkungan kerja (X_2), dan hubungan dengan atasan (X_3).
- Motivasi (*latent variable: M*) diukur oleh dua indikator: motivasi intrinsik (Y_1) dan motivasi ekstrinsik (Y_2).
- Kinerja karyawan (*latent variable: EP*) diukur oleh dua indikator: produktivitas (Y_3) dan kualitas kerja (Y_4).

Model struktural mengasumsikan bahwa kepuasan kerja mempengaruhi motivasi, yang pada gilirannya mempengaruhi kinerja karyawan.

Model dapat divisualisasikan dalam diagram jalur sebagai berikut:

$$JS \rightarrow M \rightarrow EP \quad (13)$$

Model Pengukuran

$$X_1, X_2, X_3 \rightarrow JS$$

$$Y_1, Y_2 \rightarrow M$$

$$Y_3, Y_4 \rightarrow EP$$

Model Struktural

$$M = \gamma JS + \zeta_1$$

$$EP = \beta JS + \zeta_2$$

Estimasi dan Evaluasi

Parameter model diestimasi menggunakan software SEM (misalnya, AMOS, LISREL, atau Mplus). Hasil estimasi parameter dan fit indices dievaluasi untuk menentukan kecocokan model.

- a. χ^2 : misalkan hasilnya tidak signifikan ($p > 0.05$), menunjukkan model fit.
- b. RMSEA: nilai di bawah 0.06 menunjukkan fit yang baik.
- c. CFI dan TLI, Nilai di atas 0.95 menunjukkan fit yang baik.
- d. SRMR: nilai di bawah 0.08 menunjukkan fit yang baik.

Interpretasi Hasil

- Kepuasan Kerja \rightarrow Motivasi (γ_1): Jika koefisiennya signifikan dan positif, berarti peningkatan kepuasan kerja meningkatkan motivasi.
- Motivasi \rightarrow Kinerja Karyawan (β_1): Jika koefisiennya signifikan dan positif, berarti peningkatan motivasi meningkatkan kinerja karyawan.

Dengan demikian, maka kita sudah berhasil menggunakan SEM untuk keperluan analisis data multivariat, yang mulai dari pengembangan model teoretis hingga interpretasi hasil.

Analisis multivariat adalah alat yang dinilai *robust* untuk memahami data yang kompleks dan beragam, di mana variabel yang terlibat yaitu lebih dari dua variabel. Dengan menggabungkan berbagai teknik statistik, peneliti dapat mengungkap pola tersembunyi, membuat prediksi yang lebih akurat, dan mengambil keputusan yang lebih baik berdasarkan data. Ruang lingkup analisis multivariat mencakup banyak bidang aplikasi, mulai dari ilmu sosial hingga ilmu alam, yang menunjukkan betapa pentingnya teknik ini dalam penelitian dan praktik profesional. Dalam kasus lain, saat ini sudah terdapat berbagai teknik pengenalan data dengan menggunakan

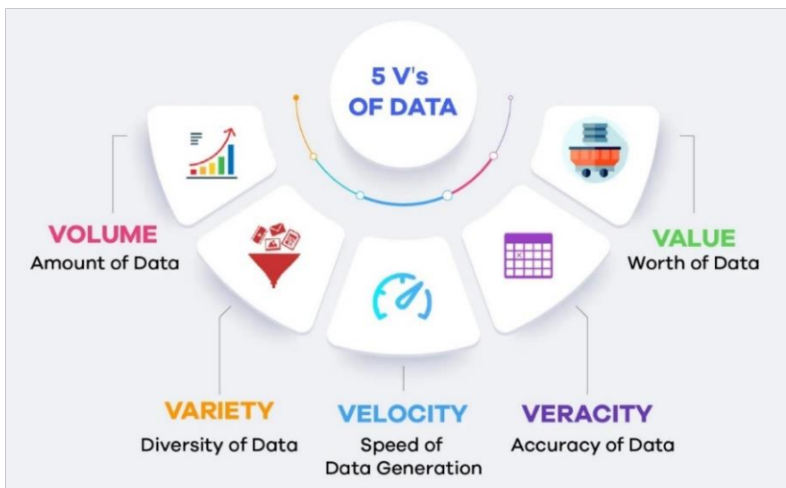
algoritma *machine learning*, seperti *support vector machine* (SVM), *neural network* (NN), *random forest* (RF), dan lain sebagainya (El Naqa & Murphy, 2015). Meskipun terkesan berbeda, pada dasarnya, teknik-teknik tersebut menggunakan persamaan yang berdasarkan pada perhitungan matematis aljabar linier (Dietterich, 1990). Oleh sebab itu, saat ini algoritma tersebut sudah termasuk banyak digunakan dan masuk dalam teknik analisis data dengan tingkat ketepatan yang beragam, dengan *tools* aplikasi yang lumayan tersebar luas, mulai dari bahasa pemrograman seperti Python, hingga piranti lunak yang sudah terintegrasi seperti *Orange*. Sementara itu, aplikasi-aplikasi tertentu untuk menghitung dengan fungsi-fungsi statistik seperti regresi berganda bisa didapatkan melalui aplikasi seperti IBM SPSS®, bahasa pemrograman R, Matlab®, dan masih banyak lagi. Dalam banyak kasus, piranti lunak tersebut memiliki kemudahan dan kesulitan tersendiri dalam penggunaan, serta terdapat piranti lunak yang berbayar seperti IBM SPSS® dan Matlab®, sedangkan piranti lunak yang gratis seperti bahasa pemrograman R dan Python. Memilih platform untuk analisis data multivariat juga dianggap cukup penting, dikarenakan penggunaannya yang dimungkinkan terus berulang, sehingga satu platform yang baik harus mengakomodasi kemudahan penggunaan dan dimungkinkannya pada pengolahan data dalam skala yang besar.

Bab 7

Penggunaan *Big Data* dan Analitik Lanjutan

7.1 Pengantar *Big Data*

Big data merujuk pada kumpulan data yang sangat besar dan kompleks, yang tidak dapat dikelola, dianalisis, maupun diproses menggunakan alat dan metode tradisional. *Big data* memiliki karakteristik utama yang dikenal sebagai 5V: *volume*, *variety*, *velocity*, *veracity*, dan *value*.



Gambar 7.1. Karakteristik *Big Data*

https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1

- **Volume:** mengacu pada sekumpulan data dalam jumlah dan volume yang sangat besar. Contohnya *feed social media* dan alur klik *user* dari halaman web. Arus data-data tersebut mencapai *terabyte*, *petabyte*, atau lebih.
- **Variety:** mengacu pada beragam jenis data, termasuk data terstruktur, semi-terstruktur, dan tidak terstruktur. Analisis terhadap data yang tidak terstruktur akan memerlukan algoritma yang agak berbeda, seperti data teks, gambar, suara, dan video.
- **Velocity:** mengacu pada kecepatan aliran data masuk sehingga dapat langsung diproses. Data dapat berasal dari berbagai sumber secara *real-time*. Contohnya: sistem operasi *online* berbasis Microsoft Silverlight dan aplikasi perkantoran (*office*) berbasis web, seperti Office365 dan *cloud storage* (Dropbox dan GDrive).
- **Veracity:** mengacu pada kualitas dan akurasi data. Data besar sering kali mengandung ketidakpastian dan inkonsistensi, sehingga memerlukan kedalaman untuk menganalisis *big data* agar bisa menghasilkan keputusan yang tepat.
- **Value:** mengacu pada potensi nilai yang dapat diperoleh dari analisis data tersebut. Data harus diolah untuk menghasilkan wawasan yang berguna. Contohnya, biodata karyawan suatu perusahaan penjualan bahan baku makanan tidak akan bernilai untuk kepentingan analisis prediksi penjualan bahan baku ke *customer*, namun sangat bernilai untuk hal lain.

Big data telah menjadi alat yang sangat penting dalam berbagai bidang penelitian, diantaranya:

- **Pengambilan Keputusan Berbasis Data:** peneliti dapat membuat keputusan yang lebih baik dan lebih tepat berdasarkan analisis data yang besar.

- **Inovasi dan Penemuan:** *big data* memungkinkan penemuan pola, tren, dan hubungan baru yang sebelumnya tidak terlihat.
- **Efisiensi Operasional:** analisis *big data* dapat meningkatkan efisiensi operasional melalui optimasi dan otomatisasi proses.
- **Personalisasi:** dalam bidang seperti kesehatan, pemasaran, dan pendidikan, *big data* memungkinkan personalisasi layanan yang lebih baik berdasarkan analisis data individu.

Big data telah digunakan pada berbagai bidang penelitian, seperti kesehatan, ekonomi, ilmu sosial, dan lingkungan.

- **Kesehatan:** analisis data genomik, catatan medis elektronik, dan data sensor kesehatan untuk diagnosis dan perawatan yang lebih baik.
- **Ekonomi:** analisis data transaksi untuk memahami tren ekonomi, perilaku konsumen, dan untuk merancang kebijakan ekonomi yang lebih efektif.
- **Ilmu Sosial:** studi tentang perilaku manusia melalui analisis data media sosial, survei besar, dan catatan publik.
- **Lingkungan:** pemantauan dan analisis data sensor untuk penelitian perubahan iklim dan pengelolaan sumber daya alam.

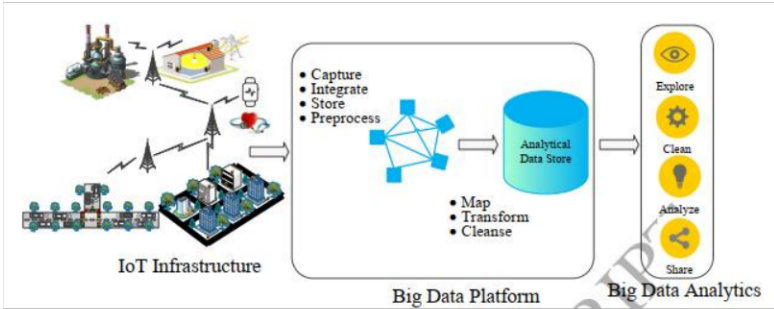
7.2. Sumber dan Jenis Data

Big data dihasilkan dari berbagai sumber yang semakin berkembang seiring dengan kemajuan teknologi. Mengetahui sumber-sumber data ini penting agar peneliti dapat memilih dan mengumpulkan data yang relevan untuk penelitian. Data dapat berasal dari berbagai sumber, diantaranya:

- **Media Sosial:** *platform* seperti Facebook, Twitter, Instagram, dan LinkedIn menghasilkan sejumlah besar data yang mencerminkan perilaku, preferensi, dan opini pengguna.

Data ini sering digunakan dalam penelitian pemasaran, sosial, dan perilaku.

- **Internet of Things (IoT):** perangkat yang terhubung seperti sensor, alat rumah pintar, dan perangkat *wearable* menghasilkan data secara *real-time* yang dapat digunakan dalam penelitian kesehatan, lingkungan, dan teknik.



Gambar 7.2. Big data pada IoT

<https://www.softobotics.com/blogs/the-power-of-big-data-analytics-in-industrial-iot/>

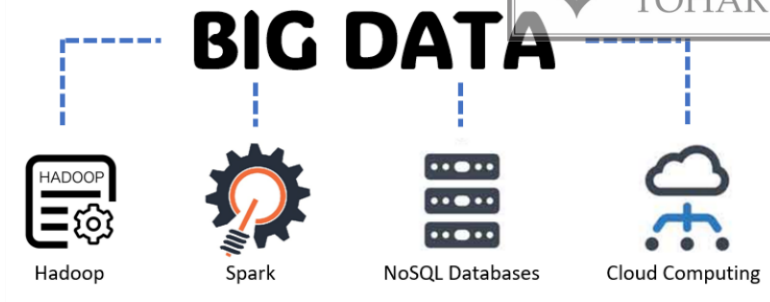
- **Transaksi Keuangan:** data dari transaksi perbankan, pembelian kartu kredit, dan aktivitas *e-commerce* memberikan wawasan berharga untuk penelitian ekonomi, pemasaran, dan manajemen risiko.
- **Log Data:** data yang dihasilkan oleh *server web*, aplikasi, dan perangkat jaringan memberikan informasi tentang penggunaan, kinerja, dan keamanan sistem.
- **Data Genomik:** data yang dihasilkan dari penelitian genom dan proteom memberikan wawasan penting dalam penelitian biologi, kedokteran, dan farmasi.
- **Catatan Medis Elektronik (EMR):** data dari catatan medis pasien digunakan dalam penelitian kesehatan untuk analisis diagnosis, perawatan, dan hasil klinis.

Big data terdiri dari berbagai jenis data yang dapat dikategorikan menjadi tiga kelompok utama: data terstruktur, semi-terstruktur, dan tidak terstruktur.

- **Data Terstruktur:** data yang disimpan dalam format yang terorganisir dan mudah diproses. Contoh: data penjualan, data pelanggan, dan data keuangan. Data terstruktur mudah diindeks dan dicari. Data jenis ini seringkali berbentuk angka atau teks.
- **Data Semi-Terstruktur:** data yang tidak mengikuti skema model data yang kaku tetapi memiliki *tag* atau penanda untuk memisahkan elemen data. Contoh: sistem pengolahan data XML/JSON. Data semi-terstruktur lebih fleksibel daripada data terstruktur, tetapi tetap memiliki beberapa tingkat organisasi.
- **Data Tidak Terstruktur:** data yang tidak memiliki format atau struktur yang telah ditentukan sebelumnya. Contoh: teks dokumen, gambar, video, dan audio. Data tidak terstruktur membutuhkan teknik analisis lanjutan untuk ekstraksi informasi, seperti *text mining*, *computer vision*, dan *speech recognition*.

7.3 Teknologi dan Infrastruktur *Big Data*

Big Data memerlukan teknologi dan infrastruktur khusus untuk mengumpulkan, menyimpan, memproses, dan menganalisis *volume* data yang sangat besar dan beragam. Teknologi ini memungkinkan peneliti untuk menangani data dengan cara yang efisien dan efektif, sehingga peneliti dapat mengekstraksi wawasan yang berguna. Teknologi dan infrastruktur *big data* meliputi *hadoop*, *spark*, *NoSQL databases*, dan *cloud computing*.



Gambar 7.3. Teknologi dan Infrastruktur *Big Data*

8.3.1 *Hadoop*

Hadoop adalah kerangka kerja *open-source* yang memungkinkan pemrosesan data dalam skala besar secara terdistribusi. *Hadoop* menggunakan HDFS (*Hadoop Distributed File System*) untuk penyimpanan data dan *MapReduce* untuk pemrosesan data. Kelebihan *Hadoop* diantaranya: skalabilitas tinggi, biaya rendah, dapat menangani data terstruktur dan tidak terstruktur. *Hadoop* digunakan dalam berbagai industri untuk analisis data besar, termasuk *e-commerce*, kesehatan, dan telekomunikasi.

8.3.2. *Spark*

Apache Spark adalah mesin pemrosesan data cepat yang dibangun di atas *Hadoop*. *Spark* menawarkan kecepatan lebih tinggi karena pemrosesan *in-memory* dan mendukung berbagai model pemrograman seperti *batch processing*, *stream processing*, dan *machine learning*. Kelebihan *spark* diantaranya: kecepatan tinggi, dukungan untuk pemrosesan *real-time*, dan ekosistem yang luas. *Spark* cocok untuk aplikasi yang memerlukan analisis *real-time* dan *machine learning*.

8.3.3. *NoSQL Databases*

NoSQL Databases adalah basis data non-relasional yang dirancang untuk menangani *volume* data besar dan beragam. Contoh: *MongoDB*, *Cassandra*, *Couchbase*. *NoSQL databases* memiliki kelebihan, diantaranya: skalabilitas horizontal, fleksibilitas dalam penyimpanan data terstruktur, semi-terstruktur, dan tidak terstruktur. *NoSQL satabases* digunakan dalam aplikasi yang memerlukan skalabilitas tinggi dan kinerja cepat, seperti aplikasi web, analisis data sensor, dan aplikasi *mobile*.

8.3.4. *Cloud Computing*

Platform cloud menyediakan infrastruktur skalabel dan fleksibel untuk penyimpanan dan pemrosesan data besar. Contoh: *Amazon Web Services (AWS)*, *Google Cloud Platform (GCP)*, dan *Microsoft Azure*. Kelebihan *platform cloud* diantaranya: skalabilitas dinamis, biaya sesuai penggunaan, dan integrasi alat *big data* yang mudah. *Platform cloud* digunakan oleh organisasi yang memerlukan solusi penyimpanan dan pemrosesan data yang skalabel tanpa investasi besar dalam infrastruktur fisik.

7.4 Metodologi Analitik Lanjutan

Metodologi analitik lanjutan mencakup berbagai teknik dan pendekatan untuk mengeksplorasi, menganalisis, dan memodelkan data. Dengan menggunakan teknik ini, peneliti dapat mengekstrak wawasan yang lebih dalam dan membuat prediksi yang lebih akurat. Beberapa metodologi analitik lanjutan yaitu *data mining*, *machine learning*, *deep learning*, serta analisis prediktif dan prespektif.

7.3.1 Data Mining

Data mining merupakan proses menemukan pola, korelasi, dan anomali dalam kumpulan data besar melalui teknik otomatis atau semi-otomatis. Data mining biasanya diaplikasikan untuk segmentasi pasar, analisis keranjang belanja, dan deteksi penipuan. Beberapa teknik untuk melakukan data mining adalah sebagai berikut:

- ✚ **Clustering**, adalah teknik mengelompokkan data ke dalam grup berdasarkan kemiripan. Contoh algoritma clustering yaitu *K-Mean* dan *Hierarchical Clustering*.
- ✚ **Association rules**, adalah teknik menemukan hubungan antar variabel dalam dataset. Contoh algoritma *association rules* adalah *Apriori*, *Eclat*, dan *FP-Growth*.
- ✚ **Classification**, adalah teknik mengelompokkan data ke dalam kategori yang telah ditentukan. Contoh algoritma *classification* adalah *Decision Trees*, *Naive Bayes*, *Support Vector Machines*.

7.3.2 Machine Learning

Machine learning merupakan cabang kecerdasan buatan yang berfokus pada pengembangan algoritma yang dapat belajar dan membuat prediksi data. *Machine learning* biasanya diaplikasikan untuk prediksi penjualan, klasifikasi email spam, maupun rekomendasi produk. Jenis pembelajaran pada *machine learning* terdiri dari:

- ✚ **Supervised learning**, dimana algoritma dilatih dengan data berlabel untuk membuat prediksi. Contoh algoritma *supervised learning* adalah *Linear Regression*, *Support Vector Machines (SVM)*, *Random Forest*, dan *Neural Networks*.
- ✚ **Unsupervised learning**, dimana algoritma mencoba menemukan struktur dalam data yang tidak berlabel. Contoh

algoritma *unsupervised learning* adalah *PCA* (*Principal Component Analysis*) dan *K-Means Clustering*.

- ✚ **Reinforcement learning**, dimana algoritma belajar melalui trial and error dengan menerima umpan balik dari tindakannya. Contoh algoritma *reinforcement* adalah *Q-Learning*, *Deep Q-Networks* (DQN).

7.3.3. Deep Learning

Deep learning merupakan subset dari *machine learning* yang menggunakan jaringan saraf tiruan dengan banyak lapisan (*deep neural network*) untuk pemodelan data kompleks dan besar. *Deep learning* dapat diaplikasikan untuk pengenalan suara, pengenalan gambar, maupun pemrosesan bahasa alami. Beberapa jenis *deep learning* adalah sebagai berikut:

- ✚ **Neural Network**, merupakan struktur jaringan yang meniru cara kerja otak manusia, terdiri dari lapisan input, lapisan tersembunyi, dan lapisan output.
- ✚ **Convolutional Neural Networks** (CNNs), digunakan untuk pengolahan data citra. Contoh aplikasi CNNs untuk pengenalan gambar, maupun analisis video.
- ✚ **Recurrent Neural Networks** (RNNs), digunakan untuk data deret waktu atau data yang memiliki urutan. Contoh aplikasi RNNs untuk prediksi saham dan analisis teks.

7.3.4 Analisis Prediktif dan Prespektif

✚ Analisis Prediktif

Analisis prediktif menggunakan data historis dan teknik statistik serta *machine learning* untuk memprediksi kejadian masa depan. Analisis prediktif dapat dilakukan menggunakan teknik *Regression analysis*, *time series analysis*, *machine learning models*. Analisis prediktif biasanya digunakan untuk prediksi penjualan, peramalan cuaca, dan analisis risiko kredit.

✚ Analisis Preskriptif

Analisis preskriptif menggunakan hasil analisis prediktif untuk merekomendasikan tindakan yang optimal. Analisis preskriptif dapat dilakukan menggunakan *optimization models*, *simulation models*, dan *decision analysis*. Analisis preskriptif biasanya digunakan untuk penentuan harga optimal, manajemen rantai pasok, dan perencanaan produksi.

7.5 Proses dan Tahapan Analisis Big Data

Analisis *big data* adalah serangkaian tahapan yang sistematis untuk mengubah data mentah menjadi wawasan yang berharga. Setiap tahapan memiliki tujuan dan teknik spesifik yang digunakan untuk memastikan data yang diolah akurat, relevan, dan siap untuk dianalisis. Proses dan analisis *big data* terdiri dari:

7.5.1 Pengumpulan Data

Proses pengumpulan data dilakukan dengan identifikasi sumber data, yaitu menentukan berbagai sumber data yang relevan dengan tujuan penelitian. Sumber data dapat berupa data internal organisasi, data eksternal dari publikasi, media sosial, sensor IoT, log server, transaksi bisnis, dan lainnya. Beberapa teknik pengumpulan data adalah sebagai berikut:

- ✚ **Web Scraping**, yaitu menggunakan alat seperti *Beautiful Soup* atau *Scrapy* untuk mengumpulkan data dari situs web.
- ✚ **API (Application Programming Interface)**, yaitu mengakses data melalui API yang disediakan oleh *platform* seperti *Twitter*, *Google Maps*, dan lain-lain.
- ✚ **Streaming Data**, yaitu menggunakan alat seperti *Apache Kafka* untuk mengumpulkan data *real-time* dari perangkat atau aplikasi.

7.5.2 Penyimpanan dan Manajemen Data

Penyimpanan dan manajemen data dilakukan melalui:

- ✚ **Data Warehousing**, dimana menyimpan data terstruktur dalam basis data yang dioptimalkan untuk *query* analitis. Contoh *data warehousing* yaitu *Amazon Redshift* dan *Google BigQuery*.
- ✚ **Data Lakes**, dimana menyimpan data dalam format aslinya di satu lokasi pusat. *Data lake* memungkinkan penyimpanan data terstruktur, semi-terstruktur, dan tidak terstruktur. Contoh *data lakes* yaitu *Hadoop HDFS* dan *Azure Data Lake*.
- ✚ **NoSQL Databases**, digunakan untuk menyimpan data yang tidak cocok untuk basis data relasional tradisional. Contoh *NoSQL Databases* adalah *MongoDB* dan *Cassandra*.

7.5.3 Pembersihan dan Praproses Data

Pembersihan data dilakukan dengan menghapus atau memperbaiki data yang tidak konsisten, duplikat, atau salah. Teknik ini penting untuk memastikan kualitas data yang tinggi. Pembersihan dan prapemrosesan data dapat dilakukan dengan beberapa teknik, diantaranya:

- ✚ **Identifikasi dan Penanganan Missing Values**, yaitu mengganti atau menghapus data yang hilang.
- ✚ **Normalisasi dan Standarisasi**, yaitu mengubah data ke dalam skala yang konsisten, misalnya skala 0-1.
- ✚ **Penghilangan Outlier**, yaitu mengidentifikasi dan menghapus data yang tidak sesuai atau ekstrem yang dapat mempengaruhi analisis.

7.5.4 Eksplorasi dan Visualisasi Data

Eksplorasi data menggunakan teknik statistik deskriptif untuk memahami karakteristik data, seperti distribusi, *mean*, *median*, dan *variance*. Sedangkan visualisasi data membantu

dalam mengidentifikasi pola, tren, dan anomali dengan menggunakan visualisasi. Contoh alat yang dapat membantu visualisasi data yaitu *Tableau*, *Power BI*, *matplotlib*, dan *Seaborn*. Teknik visualisasi data dapat juga menggunakan grafik secara umum (*Histogram*, *scatter plot*, *box plot*, *heatmap*).

7.5.5 Pemodelan Data

Tahap pemodelan data yaitu pemilihan model, pelatihan model, dan validasi model.

- ✚ **Pemilihan model**, yaitu memilih teknik pemodelan yang sesuai berdasarkan sifat data dan tujuan analisis. Contoh model data diantaranya: *regresi*, *klasifikasi*, *clustering*, dan analisis deret waktu.
- ✚ **Pelatihan model**, yaitu menggunakan data latih untuk membangun model prediktif. Pelatihan model dapat dilakukan menggunakan teknik *scikit-learn*, *TensorFlow*, dan *PyTorch*.
- ✚ **Validasi model**, yaitu menggunakan data uji untuk menilai kinerja model dan menghindari *overfitting*. Teknik validasi meliputi *cross-validation*, *confusion matrix*, dan *ROC curve*.

7.5.6 Evaluasi Model

Evaluasi model dapat dilakukan dengan menggunakan metrik evaluasi dan pemilihan model terbaik.

- ✚ **Metrik evaluasi**, yaitu menilai kinerja model menggunakan metrik yang sesuai. Contoh metrik evaluasi yaitu akurasi, presisi, *recall*, *F1-score* untuk klasifikasi; *mean squared error* (MSE), *root mean squared error* (RMSE) untuk regresi.
- ✚ **Pemilihan model terbaik**, yaitu memilih model yang memberikan kinerja terbaik berdasarkan metrik evaluasi.

7.5.7 Implementasi dan Monitoring

- ✚ **Implementasi Model**, yaitu menerapkan model yang telah dilatih ke dalam lingkungan produksi. Hal ini dapat melibatkan integrasi dengan sistem yang ada atau pengembangan aplikasi baru.
- ✚ **Monitoring dan Pemeliharaan**, yaitu memantau kinerja model secara terus-menerus untuk memastikan bahwa model tetap akurat dan relevan. Menyegarkan model dengan data baru secara berkala dan memperbarui algoritma jika diperlukan.

7.6 Masa Depan Big Data dan Analitik Lanjutan

Big Data dan analitik lanjutan telah mengalami perkembangan pesat dalam beberapa tahun terakhir dan akan terus berevolusi seiring dengan kemajuan teknologi.

7.6.1 Tren Teknologi Masa Depan

- ✚ **Artificial Intelligence (AI) dan Machine Learning (ML) yang Lebih Canggih**

AI dan ML akan semakin terintegrasi dalam analitik data, dengan algoritma yang lebih pintar dan lebih adaptif. Teknologi seperti pembelajaran mendalam (*deep learning*) akan terus berkembang, memungkinkan analisis data yang lebih kompleks dan presisi.

- ✚ **Edge Computing**

Pemrosesan data yang dilakukan dekat dengan sumber data, seperti perangkat IoT, akan menjadi lebih umum. Ini akan mengurangi latensi dan *bandwidth*, memungkinkan analisis *real-time* yang lebih efisien.

- ✚ **Quantum Computing**

Potensi komputasi kuantum dalam memproses data besar dengan kecepatan yang luar biasa akan membuka peluang

baru dalam analitik data, memungkinkan pemecahan masalah yang sebelumnya tidak mungkin diatasi dengan komputasi klasik.

7.6.2 Inovasi dalam Infrastruktur Big Data

Cloud Computing yang Lebih Andal dan Fleksibel

Penggunaan *cloud* untuk penyimpanan dan pemrosesan data akan terus meningkat, dengan peningkatan pada skalabilitas, keamanan, dan fleksibilitas. Penyedia layanan *cloud* seperti AWS, Google Cloud, dan Microsoft Azure akan memperkenalkan fitur-fitur baru yang lebih canggih.

Hybrid and Multi-Cloud Environments

Penggunaan kombinasi dari *on-premise*, *private cloud*, dan *public cloud* akan menjadi lebih umum, memberikan fleksibilitas dan optimasi biaya yang lebih baik.

5G and IoT Integration

Dengan adopsi luas dari jaringan 5G, jumlah perangkat IoT akan meningkat drastis, menghasilkan *volume* data yang lebih besar dan lebih bervariasi untuk dianalisis.

7.6.3 Evolusi Teknik Analitik

Automated Machine Learning (AutoML)

AutoML akan membuat pengembangan model ML lebih mudah dan lebih cepat, bahkan bagi mereka yang bukan ahli data. Hal ini akan mempercepat siklus pengembangan model dan meningkatkan produktivitas.





Augmented Analytics

Penggunaan AI untuk otomatisasi dan peningkatan proses analitik, memungkinkan peneliti untuk lebih cepat menemukan wawasan dari data tanpa memerlukan keahlian teknis yang mendalam.

Explainable AI (XAI)

Terdapat peningkatan fokus pada membuat algoritma AI lebih transparan dan dapat dijelaskan, memungkinkan pemahaman yang lebih baik tentang bagaimana model membuat keputusan.

7.6.4 Perubahan dalam Industri dan Aplikasi Praktis

-  **Kesehatan:** penggunaan analitik lanjutan untuk diagnosis yang lebih cepat dan akurat, personalisasi perawatan pasien, dan prediksi penyakit.
-  **Keuangan:** deteksi penipuan yang lebih baik, analisis risiko yang lebih akurat, dan personalisasi layanan keuangan.
-  **Ritel dan *E-commerce*:** personalisasi pengalaman belanja, manajemen rantai pasokan yang lebih efisien, dan prediksi tren pasar.
-  **Pemerintahan dan Kebijakan Publik:** peningkatan efisiensi layanan publik, analisis data untuk perencanaan kebijakan yang lebih baik, dan transparansi pemerintahan.

7.6.5 Tantangan dan Pertimbangan Etis

Privasi dan Keamanan Data

Dengan peningkatan volume dan sensitivitas data yang dikumpulkan, tantangan dalam melindungi privasi individu dan menjaga keamanan data akan semakin besar.

Regulasi dan Kepatuhan

Peneliti dan organisasi perlu mengikuti perkembangan regulasi yang semakin ketat terkait dengan penggunaan dan perlindungan data.

Bias dalam Algoritma

Mengurangi bias dalam algoritma AI dan ML akan menjadi tantangan utama untuk memastikan hasil analitik yang adil dan tidak diskriminatif.

7.6.6 Kolaborasi Antar Disiplin

Interdisipliner

Kolaborasi antara ahli data, ilmuwan komputer, ahli domain, dan etika akan menjadi semakin penting untuk mengatasi kompleksitas analisis *big data*.

Pendidikan dan Pelatihan

Pendidikan berkelanjutan dan pelatihan dalam teknologi baru akan menjadi kunci untuk mempersiapkan tenaga kerja yang kompeten dalam bidang analitik lanjutan.

Big data dan analitik lanjutan telah membawa revolusi dalam cara kita mengumpulkan, menyimpan, menganalisis, dan memanfaatkan data. Teknologi ini tidak hanya memberikan wawasan yang lebih mendalam dan akurat tetapi juga membuka peluang baru untuk inovasi di berbagai bidang, mulai dari kesehatan, keuangan, ritel, hingga pemerintahan. Namun, dengan segala potensi yang ditawarkannya, terdapat tanggung jawab besar yang harus diemban oleh peneliti dan praktisi dalam memastikan bahwa data digunakan secara etis dan aman.

Penting bagi peneliti untuk terus mengikuti perkembangan teknologi dan tren dalam *big data* dan analitik lanjutan, termasuk memahami teknik baru, mengadopsi alat yang lebih canggih, serta memperbarui pengetahuan tentang regulasi dan kebijakan yang mengatur penggunaan data. Dengan demikian, peneliti dapat memaksimalkan manfaat yang diperoleh dari analisis data dan meminimalkan risiko yang mungkin terjadi.

Bab 8

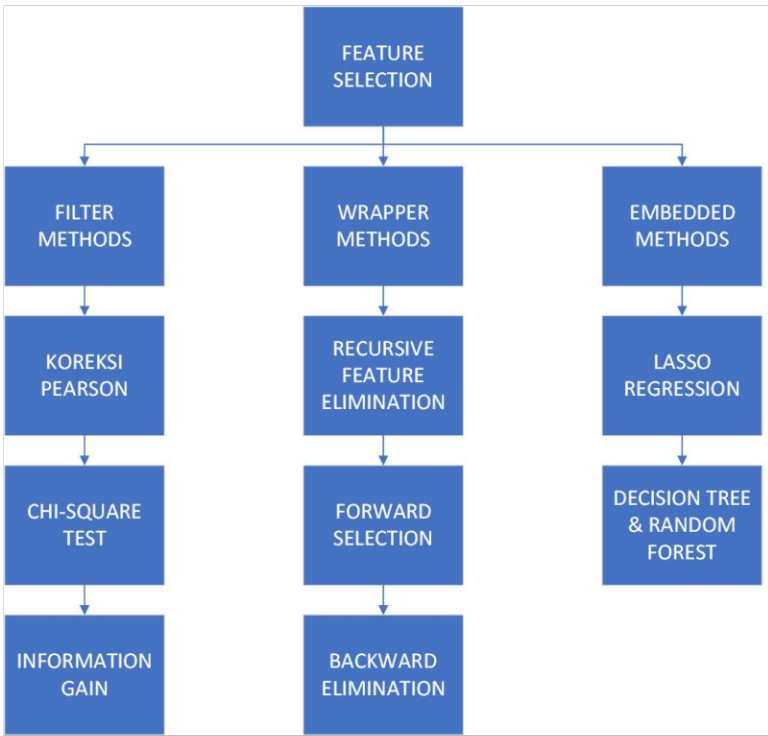
Penerapan dalam Analisis Variabel: Machine Learning dan Data Mining

8.1 Pemilihan Fitur (Feature Selection)

Pemilihan Fitur (Feature Selection) adalah proses memilih subset fitur yang paling relevan dari dataset untuk digunakan dalam membangun model machine learning. Tujuannya pemilihan fitur adalah untuk meningkatkan kinerja model dengan menghilangkan atau mengurangi jumlah fitur yang tidak relevan atau redundan sehingga mengurangi kompleksitas model dan dapat meningkatkan akurasi model, menghindari overfitting dengan menggunakan lebih sedikit fitur mengurangi risiko model mempelajari noise dari data, mengurangi waktu komputasi atau meningkatkan efisiensi komputasi dengan mengurangi jumlah fitur mempercepat proses pelatihan dan prediksi, serta menyederhanakan model dengan lebih sedikit fitur lebih mudah dipahami dan diinterpretasikan (Wiyanto and Setyaningsih, 2021). Berikut ini beberapa metode pemilihan fitur yang sering dipakai. Pertama, Filter Methods terdiri dari Koreksi Pearson dengan mengukur linearitas hubungan antara fitur dan target, Chi-square Test dengan menguji independensi antara fitur kategorikal dan target, dan Information Gain dengan mengukur penurunan entropi dengan membagi data berdasarkan fitur. Kelebihan dari metode ini yaitu cepat dan sederhana serta tidak

bergantung pada algoritma machine learning tertentu. Sedangkan kekurangan dari metode ini yaitu tidak mempertimbangkan interaksi antar fitur. Kedua, Wrapper Methods terdiri dari Recursive Feature Elimination (RFE) dengan menghapus fitur yang paling tidak penting secara iteratif, Forward Selection dengan menambahkan fitur satu per satu berdasarkan peningkatan kinerja model, Backward Elimination dengan menghapus fitur satu per satu berdasarkan penurunan kinerja model. Kelebihan dari metode ini yaitu mempertimbangkan interaksi antar fitur serta biasanya menghasilkan subset fitur yang lebih optimal untuk model tertentu. Kekurangan dari metode ini yaitu waktu komputasi yang tinggi serta bergantung pada algoritma machine learning yang digunakan. Ketiga, Embedded Methods yaitu terdiri dari Lasso Regression dengan menggunakan penalti L1 untuk mendorong koefisien fitur tidak penting menjadi nol, Decision Tree dan Random Forests secara alami melakukan pemilihan fitur selama pelatihan model. Kelebihan dari metode ini yaitu biasanya lebih efisien daripada wrapper methods serta mengoptimalkan pemilihan fitur dan pelatihan model secara simultan. Kekurangan dari metode ini yaitu bergantung pada algoritma machine learning tertentu dan mungkin tidak seefektif dalam mengidentifikasi fitur-fitur terbaik untuk semua jenis model. Contoh penerapan pemilihan fitur dalam bidang kesehatan misalkan prediksi penyakit berdasarkan fitur-fitur pasien seperti usia, berat badan, tekanan darah. Proses ini menggunakan Filter Methods dan RFE untuk mengidentifikasi dan memilih fitur yang paling relevan. Sedangkan contoh penerapan fitur dalam bidang pemasaran misalnya prediksi respons pelanggan terhadap kampanye pemasaran berdasarkan fitur seperti umur, pendapatan, riwayat pembelian. Proses ini menggunakan Chi-square Test dan Lasso Regression untuk memilih fitur yang paling berpengaruh. Pemilihan fitur adalah langkah penting dalam analisis data dan machine learning yang memastikan bahwa model hanya menggunakan fitur yang

memberikan nilai prediktif, sehingga meningkatkan efisiensi dan efektivitas model (Amzah and Bayuaji, 2024).



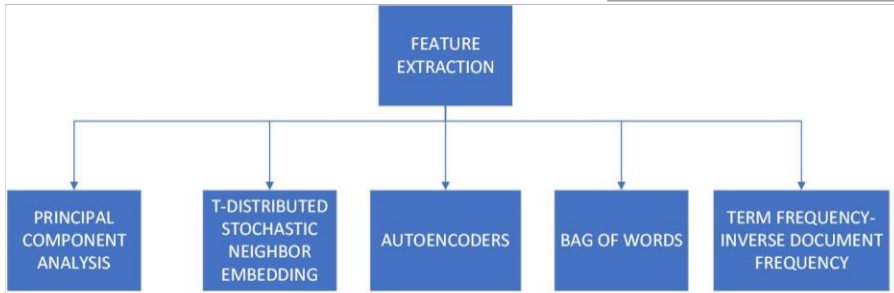
Gambar 8.1 Feature Selection Methods

8.2 Ekstraksi Fitur (Feature Extraction)

Ekstraksi fitur adalah proses transformasi data mentah menjadi fitur-fitur yang lebih bermakna dan dapat digunakan untuk pelatihan model machine learning. Ini melibatkan pembuatan fitur baru dari data yang ada atau mengubah fitur yang ada ke dalam bentuk yang lebih berguna. Tujuannya adalah untuk menyederhanakan representasi data sambil mempertahankan informasi yang paling relevan. Ekstraksi fitur dapat mengurangi dimensi data dengan mengurangi jumlah variabel yang harus dianalisis tanpa kehilangan informasi penting, meningkatkan kinerja model dengan membantu model untuk belajar lebih baik dengan menyajikan data dalam bentuk

yang lebih informatif, mengatasi masalah data sparsity dengan mengubah data yang jarang menjadi representasi yang lebih padat, mempermudah interpretasi data dengan menghasilkan fitur yang lebih mudah dipahami dan diinterpretasikan (Asrianto and Herwinanda, 2022). Berikut adalah beberapa metode ekstraksi fitur. Pertama, Principal Component Analysis (PCA) adalah teknik pengurangan dimensi yang mengubah data ke dalam komponen utama yang memaksimalkan varians. Kegunaannya mengurangi jumlah fitur dengan menjaga varians maksimum dalam data. Contoh penerapannya yaitu mengurangi dimensi data citra untuk pengenalan wajah. Kedua, t-SNE (t-distributed Stochastic Neighbor Embedding) adalah teknik visualisasi yang mengurangi dimensi data non-linear untuk memvisualisasikan data dalam dimensi rendah. Kegunaannya membantu dalam visualisasi dan eksplorasi data yang kompleks. Contoh penerapannya yaitu visualisasi kluster dalam dataset yang besar dan kompleks. Ketiga, Autoencoders adalah jenis jaringan saraf tiruan yang digunakan untuk belajar representasi rendah dimensi dari data. Kegunaannya membantu dalam mengurangi dimensi data dan menangkap struktur non-linear. Contoh penerapannya yaitu menemukan representasi rendah dimensi dari data gambar atau suara. Keempat, Bag of Words (BoW) adalah teknik ekstraksi fitur dari teks yang mengubah teks menjadi vektor kata-kata yang ada dalam dokumen. Kegunaannya memproses teks untuk model machine learning. Contoh penerapannya yaitu analisis sentimen dari ulasan produk. Kelima, Term Frequency-Inverse Document Frequency (TF-IDF) adalah teknik untuk menilai pentingnya sebuah kata dalam dokumen relatif terhadap seluruh korpus. Kegunaannya menyaring kata-kata umum yang kurang informatif dalam analisis teks. Contoh penerapannya yaitu pencarian informasi dan klasifikasi teks (Rosid, Gunawan and Pramana, 2015). Proses Ekstraksi Fitur terdiri dari identifikasi fitur dengan menentukan fitur apa yang relevan untuk masalah yang sedang dianalisis, transformasi fitur dengan mengubah fitur mentah menjadi

bentuk yang lebih bermanfaat menggunakan teknik teknik ekstraksi fitur, evaluasi fitur dengan menilai seberapa baik fitur yang diekstrak dalam membantu model machine learning melakukan tugasnya, serta penerapan model dengan menggunakan fitur yang diekstrak untuk melatih model machine learning dan melakukan prediksi. Contoh penerapan Ekstraksi Fitur yaitu dalam pengolahan citra tujuannya ekstraksi fitur dari citra untuk tugas seperti pengenalan wajah atau deteksi objek. Prosesnya menggunakan PCA atau Autoencoders untuk mengurangi dimensi citra dan menemukan fitur yang paling relevan. Selain itu, dalam analisis teks tujuannya ekstraksi fitur dari teks untuk analisis sentimen, klasifikasi dokumen, atau pengenalan topik. Prosesnya menggunakan BoW atau TF-IDF untuk mengubah teks menjadi vektor fitur yang dapat digunakan dalam model machine learning. Selanjutnya, dalam data keuangan tujuannya ekstraksi fitur dari data transaksi untuk mendeteksi penipuan. Prosesnya menggunakan teknik statistik dan machine learning untuk menemukan pola yang mencurigakan dalam data transaksi. Keuntungan Ekstraksi Fitur yaitu mengurangi kompleksitas data dengan membuat data lebih mudah diolah dan dianalisis, meningkatkan kinerja model dengan mengekstrak fitur yang lebih informatif apakah model dapat belajar lebih baik dan memberikan hasil yang lebih akurat, mengatasi masalah dimensionalitas tinggi dengan mempermudah analisis data yang memiliki jumlah fitur sangat besar, serta memperbaiki interpretabilitas dengan membantu memahami data dan model dengan lebih baik. Ekstraksi fitur adalah langkah penting dalam proses machine learning dan data mining yang memungkinkan model untuk bekerja dengan data dalam bentuk yang lebih optimal. Dengan memilih dan mentransformasi fitur yang tepat, kita dapat meningkatkan kinerja model, mengurangi waktu komputasi, dan mendapatkan wawasan yang lebih dalam dari data yang tersedia (Alizah *et al.*, 2020).



Gambar 8.2 Feature Extraction Methods

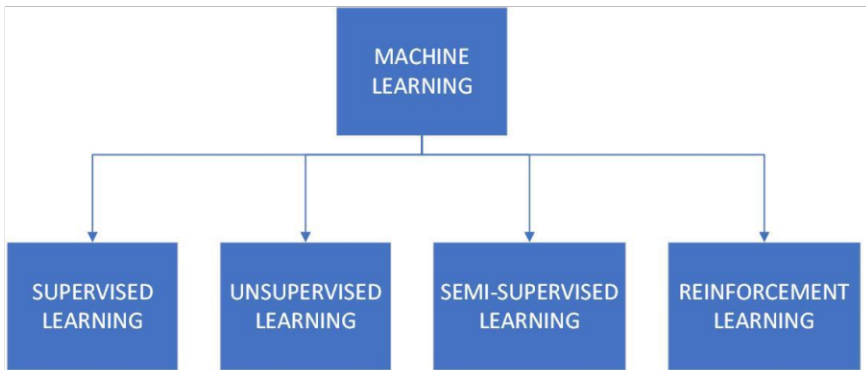
8.3 Penerapan Algoritma Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (AI) yang berfokus pada pengembangan algoritma yang memungkinkan komputer untuk belajar dari dan membuat keputusan berdasarkan data. Penerapan algoritma machine learning melibatkan berbagai langkah mulai dari pemilihan algoritma yang sesuai hingga evaluasi kinerja model yang telah dibangun. Pemilihan algoritma yang tepat, teknik evaluasi yang baik, dan pemahaman mendalam tentang dataset yang digunakan adalah kunci untuk membangun model yang efektif dan akurat (Savitri and Nursalim, 2023).

a. Jenis-Jenis Algoritma Machine Learning

Penerapan algoritma machine learning melibatkan penggunaan berbagai teknik dan model untuk mengekstrak wawasan berharga dari data dan membuat prediksi atau keputusan otomatis. Berikut adalah penjelasan lengkap mengenai penerapan algoritma-algoritma machine learning. Pertama, Supervised Learning adalah algoritma dilatih menggunakan dataset yang sudah diberi label. Contoh dari algoritma ini yaitu Regresi Linear, Regresi Logistik, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees (Handayani *et al.*, 2021). Kedua, Unsupervised Learning adalah algoritma dilatih menggunakan dataset tanpa label.

Contoh dari algoritma ini yaitu K-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), Apriori. Ketiga, Semi-supervised Learning adalah kombinasi dari dataset yang diberi label dan tidak diberi label. Contoh dari algoritma ini yaitu Semi-supervised SVM. Keempat, Reinforcement Learning adalah algoritma belajar melalui trial and error untuk memaksimalkan reward. Contoh dari algoritma ini yaitu Q-Learning, Deep Q-Networks (DQN) (Putu *et al.*, 2024).



Gambar 8.3 Machine Learning Methods

b. Studi Kasus Penerapan Algoritma Machine Learning

Penerapan algoritma machine learning adalah langkah penting dalam pengolahan data dan analisis. Dengan berbagai algoritma yang tersedia, mulai dari yang sederhana hingga yang kompleks, praktisi data dapat memilih dan menyesuaikan pendekatan terbaik untuk menyelesaikan masalah spesifik (Agustiningsih, Findawati and Alnarus Kautsar, 2023). Berikut ini contoh penerapan algoritma machine learning dalam kehidupan sehari-hari. Pertama, prediksi harga rumah menggunakan algoritma regresi linear. Prosesnya dimulai dari mengumpulkan data fitur (luas bangunan, jumlah kamar, dan lokasi), melatih model regresi linear dengan data yang ada, dan menggunakan model untuk memprediksi harga rumah baru berdasarkan fitur-fitur tersebut. Kedua, deteksi spam email menggunakan algoritma support vector machine (SVM).

Prosesnya dimulai dari mengumpulkan dataset email yang telah diberi label spam atau bukan spam, ekstraksi fitur teks menggunakan teknik seperti TF-IDF, melatih model SVM dengan data yang ada, dan menggunakan model untuk mengklasifikasikan email baru sebagai spam atau bukan spam. Ketiga, klasifikasi penyakit menggunakan algoritma decision tree dan random forests. Prosesnya dimulai dari mengumpulkan data medis pasien (gejala, riwayat kesehatan, dan hasil tes laboratorium), melatih model decision tree dengan data yang ada, dan menggunakan model untuk memprediksi kemungkinan penyakit pada pasien baru. Keempat, pengenalan wajah menggunakan algoritma neural networks dan deep learning. Prosesnya dimulai dari mengumpulkan dataset gambar wajah, melatih model neural network dengan data yang ada, dan menggunakan model untuk mengenali wajah baru dalam gambar atau video (Arrank Tonapa, D.K. Manembu and D. Kambey, 2024).

8.4 Data Mining

Data mining adalah proses ekstraksi informasi yang berguna dan pola yang menarik dari kumpulan data besar. Data mining menggunakan proses eksplorasi dan analisis sejumlah besar data untuk menemukan pola yang bermakna dan aturan yang tersembunyi. Teknik ini merupakan bagian dari ilmu data yang memanfaatkan metode dari berbagai disiplin ilmu seperti statistik, pembelajaran mesin, dan basis data untuk mengekstrak pengetahuan dari data yang besar dan kompleks (Primadona and Fauzi, 2023).

a. Tujuan Data Mining

Data mining adalah proses penting dalam ilmu data yang memungkinkan penemuan wawasan berharga dari data besar. Data mining bertujuan untuk penemuan pengetahuan dengan mengidentifikasi pola dan hubungan tersembunyi dalam data yang memberikan wawasan baru, prediksi

dengan membuat model yang dapat memprediksi kejadian atau hasil masa depan berdasarkan data historis, klasifikasi dengan mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya, dan pembuatan keputusan dengan mendukung pengambilan keputusan berbasis data di berbagai bidang seperti bisnis, kesehatan, dan keuangan (Karim, Sentinuwo and Sambul, 2017).

b. Proses Data Mining

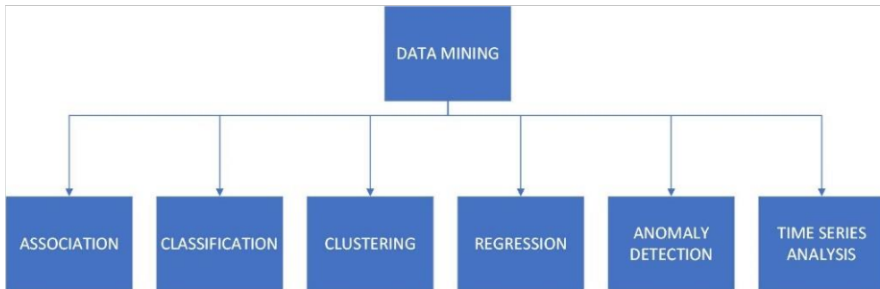
Proses data mining adalah serangkaian langkah yang sistematis untuk mengubah data mentah menjadi informasi berharga. Dengan mengikutinya, organisasi dapat mengidentifikasi pola tersembunyi, membuat prediksi yang akurat, dan mendukung pengambilan keputusan yang lebih baik. Proses data mining terdiri dari beberapa tahapan yang berurutan untuk mengolah data mentah menjadi informasi yang berguna (Arya Darmawan, Dewanta and Astuti, 2023). Berikut ini langkah-langkah dalam proses data mining. Pertama, pengumpulan data dimulai mengumpulkan data dari berbagai sumber seperti database, data warehouse, file teks, atau sensor. Contohnya mengumpulkan data penjualan dari berbagai cabang toko. Kedua, pra-pemrosesan data dimulai dari pembersihan data (mengatasi data yang hilang, menangani outliers, dan menghapus duplikasi), transformasi data (normalisasi, standardisasi, dan encoding variabel kategorikal), reduksi data (mengurangi jumlah data untuk meningkatkan efisiensi analisis, seperti dengan pemilihan fitur atau pengurangan dimensi). Contohnya membersihkan data pelanggan yang duplikat dan menangani nilai yang hilang dalam dataset penjualan. Ketiga, pemilihan data dimulai dari memilih subset data yang relevan untuk analisis lebih lanjut. Contohnya memilih data transaksi penjualan selama lima tahun terakhir. Keempat, penambangan data (data mining) menggunakan berbagai teknik untuk menemukan pola dan model dalam data. Contohnya

menggunakan algoritma klustering untuk mengelompokkan pelanggan berdasarkan pola pembelian. Kelima, mengevaluasi pola yang ditemukan untuk menentukan seberapa bermakna dan bergunanya pola tersebut. Contohnya menggunakan metrik seperti akurasi atau lift untuk mengevaluasi model klasifikasi. Keenam, representasi pengetahuan dengan menyajikan pola yang ditemukan dalam format yang mudah dipahami seperti grafik, laporan, atau visualisasi. Contohnya menampilkan hasil klustering dalam bentuk grafik untuk memudahkan interpretasi oleh manajemen.

c. Teknik-teknik Data Mining

Dengan menggunakan teknik-teknik seperti asosiasi, klasifikasi, klustering, dan regresi, serta alat dan teknologi yang tepat, data mining dapat diterapkan di berbagai bidang untuk meningkatkan pengambilan keputusan dan efisiensi operasional. Berikut ini teknik-teknik dalam data mining. Teknik asosiasi (association) yaitu mengidentifikasi aturan yang menunjukkan hubungan antar item dalam dataset. Contohnya Algoritma Apriori untuk menemukan aturan asosiasi dalam data penjualan, seperti "Jika membeli roti, maka cenderung membeli mentega". Teknik klasifikasi (classification) yaitu mengklasifikasikan item data ke dalam kategori yang telah ditentukan sebelumnya. Contohnya penggunaan algoritma decision tree, support vector machines (SVM), k-nearest neighbors (KNN), dan neural networks. Teknik klustering (clustering) yaitu mengelompokkan item data yang mirip satu sama lain ke dalam kluster. Contohnya penggunaan algoritma k-means, hierarchical clustering, dan DBSCAN untuk mengelompokkan pelanggan berdasarkan kebiasaan belanja mereka. Teknik regresi (regression) yaitu memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen untuk memprediksi nilai kontinu. Contohnya penggunaan algoritma regresi linear

untuk memprediksi penjualan berdasarkan data historis (Kusuma and Hidayat, 2024). Teknik deteksi anomali (anomaly detection) adalah mengidentifikasi data yang tidak biasa atau tidak sesuai dengan pola umum dalam dataset. Contohnya penggunaan algoritma isolation forest untuk mendeteksi transaksi keuangan yang mencurigakan. Teknik analisis deret waktu (time series analysis) yaitu menganalisis data yang dikumpulkan dalam urutan waktu untuk menemukan tren, pola musiman, dan siklus. Contohnya penggunaan algoritma ARIMA untuk memprediksi penjualan berdasarkan data bulanan (Asrirawan, Permata and Fauzan, 2022).



Gambar 8.4 Data Mining Methods

d. Aplikasi Data Mining

Data mining memiliki berbagai aplikasi yang bermanfaat di berbagai bidang. Berikut ini adalah beberapa contoh aplikasi data mining dalam berbagai bidang. Bidang pemasaran dan penjualan digunakan untuk memahami perilaku pelanggan, segmentasi pasar, dan pengembangan strategi pemasaran. Contohnya analisis keranjang belanja untuk menemukan produk yang sering dibeli bersama, personalisasi rekomendasi produk. Bidang keuangan membantu dalam mendeteksi penipuan, analisis risiko, dan keputusan investasi. Contohnya deteksi penipuan, analisis risiko kredit, prediksi harga saham. Bidang kesehatan digunakan untuk analisis data medis, prediksi penyakit, dan pengelolaan

perawatan kesehatan. Contohnya prediksi penyakit, analisis data pasien untuk perawatan kesehatan yang lebih baik. Bidang telekomunikasi membantu dalam mengelola churn pelanggan, analisis jaringan, dan peningkatan layanan. Contohnya analisis churn pelanggan, deteksi anomali dalam penggunaan jaringan. Bidang manufaktur digunakan untuk pemeliharaan prediktif, optimasi proses, dan pengelolaan rantai pasokan. Contohnya prediksi pemeliharaan mesin, optimasi rantai pasokan. Bidang pendidikan diterapkan dalam analisis data pendidikan untuk meningkatkan proses pembelajaran dan administrasi. Contohnya analisis kinerja siswa, personalisasi pembelajaran dan peningkatan administrasi (Pratama, 2021).

8.5 Evaluasi Model

Evaluasi model adalah tahap penting dalam proses data mining yang bertujuan untuk menilai kinerja dan keakuratan model yang telah dibangun, apakah model yang dibangun memiliki kinerja yang memadai dan dapat diandalkan. Proses ini melibatkan penggunaan berbagai metrik dan teknik untuk memastikan bahwa model tersebut mampu memberikan prediksi yang akurat dan dapat diandalkan. Selain itu, proses ini untuk menilai performa model berdasarkan data uji yang tidak terlihat selama pelatihan. Dengan menggunakan berbagai metrik dan teknik evaluasi, kita dapat menilai kekuatan dan kelemahan model, serta membuat keputusan yang lebih baik dalam pemilihan dan penerapan model untuk aplikasi nyata. Visualisasi hasil evaluasi juga membantu dalam memahami kinerja model secara lebih mendalam dan komunikatif (Maruli Tua Silaen, 2023).

a. Tahapan Evaluasi Model

Tahapan evaluasi model dalam data mining melibatkan beberapa langkah penting untuk memastikan bahwa model yang dibangun memiliki kinerja yang optimal dan dapat

diandalkan. Berikut ini tahapan-tahapan dari evaluasi model. Pertama, pemilihan data uji yaitu membagi dataset menjadi data latih (training set) dan data uji (testing set). Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kinerja model. Contohnya, dataset dibagi menjadi 80% data latih dan 20% data uji. Kedua, pelatihan model yaitu melatih model menggunakan data latih untuk mengidentifikasi pola dan hubungan dalam data. Contohnya, melatih model regresi linear untuk memprediksi harga rumah berdasarkan fitur-fitur seperti ukuran dan lokasi. Ketiga, pengujian model yaitu menggunakan data uji untuk menguji kinerja model yang telah dilatih. Contohnya, menggunakan data uji untuk mengukur seberapa baik model regresi linear memprediksi harga rumah. Keempat, evaluasi kinerja yaitu menggunakan metrik evaluasi untuk menilai kinerja model berdasarkan prediksi yang dihasilkan. Contohnya, menghitung metrik seperti akurasi, precision, recall, dan F1-score untuk model klasifikasi. Dengan mengikuti tahapan evaluasi model yang mencakup pemilihan data uji, pelatihan model, pengujian model, dan evaluasi kinerja menggunakan berbagai metrik dan teknik evaluasi, kita dapat memastikan bahwa model yang dibangun memiliki kinerja yang memadai dan dapat diandalkan untuk aplikasi nyata. Visualisasi hasil evaluasi juga membantu dalam memahami dan mengkomunikasikan kinerja model secara lebih efektif (Purnama, 2021).

b. Metrik Evaluasi Model

Berbagai metrik digunakan untuk mengevaluasi kinerja model tergantung pada jenis model dan tujuan analisis. Berikut adalah beberapa metrik evaluasi yang umum digunakan. Pertama, akurasi (accuracy) adalah proporsi prediksi yang benar terhadap total prediksi. Formula dari akurasi yaitu jumlah prediksi benar dibagi dengan total prediksi. Contohnya, Jika model klasifikasi memprediksi 90

dari 100 sampel dengan benar, akurasi adalah 90%.

Kedua, presisi (precision) adalah proporsi prediksi positif yang benar dari semua prediksi positif. Formula dari presisi yaitu $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$. Contohnya, jika dari 50 prediksi positif (40 benar dan 10 salah) maka presisinya adalah 80%.

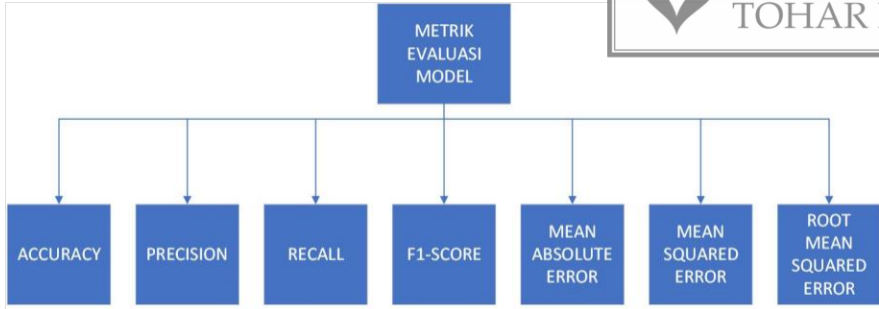
Ketiga, recall (sensitivity) adalah proporsi prediksi positif yang benar dari semua kasus positif aktual. Formula dari recall yaitu $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$. Contohnya, jika ada 50 kasus positif dan model mengidentifikasi 40 di antaranya dengan benar, recall-nya adalah 80%.

Keempat, f1-score adalah harmonic mean dari presisi dan recall, memberikan keseimbangan antara keduanya. Formula dari f1-score yaitu $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$. Contohnya, jika presisi dan recall keduanya 80%, F1-score juga 80%.

Kelima, mean absolute error (MAE) adalah rata-rata selisih absolut antara nilai aktual dan nilai prediksi. Contohnya, jika prediksi harga rumah rata-rata meleset \$10,000 dari harga sebenarnya, MAE-nya adalah \$10,000.

Keenam, mean squared error (MSE) adalah rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi, memberikan penalti yang lebih besar untuk kesalahan yang lebih besar. Contohnya, jika model memiliki MSE sebesar 100,000, ini berarti kesalahan prediksi cenderung lebih besar.

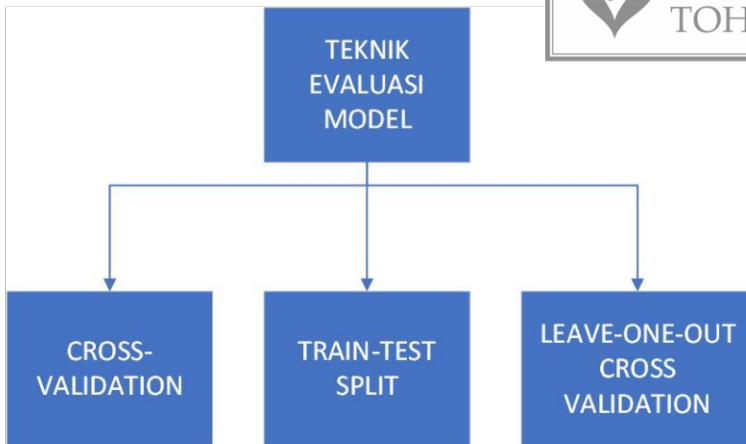
Ketujuh, root mean squared error (RMSE) adalah akar kuadrat dari MSE, memberikan interpretasi yang lebih langsung dalam unit asli dari target variabel. Contohnya, jika RMSE adalah \$10,000, ini berarti rata-rata kesalahan prediksi sekitar \$10,000 (Ihzaniah, Setiawan and Wijaya, 2023).



Gambar 8.5 Metrik Evaluasi Model

c. Teknik Evaluasi Model

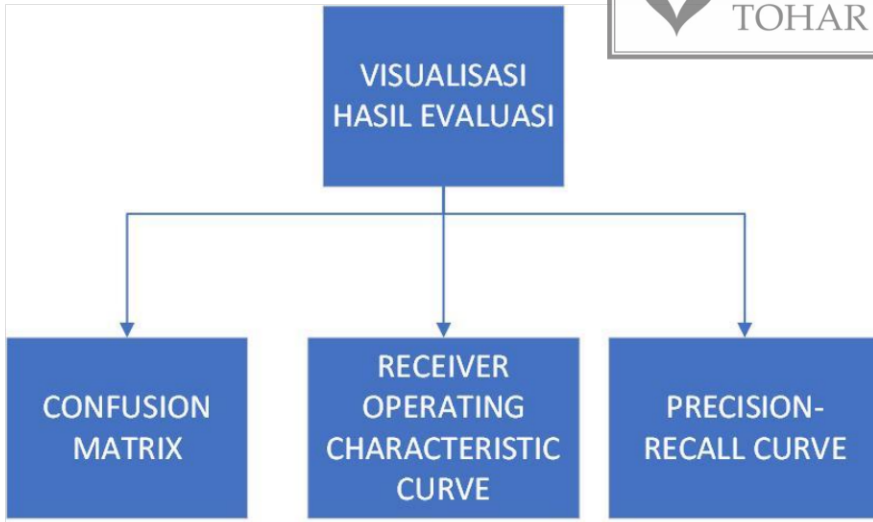
Berbagai teknik dapat digunakan untuk mengevaluasi model, terutama dalam konteks pembagian data latih dan data uji. Pertama, cross-validation yaitu membagi dataset menjadi beberapa subset (folds) dan melatih serta menguji model pada masing-masing fold secara bergantian. Contohnya, menggunakan k-fold cross-validation dengan $k = 10$ untuk mengevaluasi model klasifikasi (Ari Ahmad Dhani, Taghfirul Azhima Yoga Siswa, 2024). Kedua, train-test split yaitu membagi dataset menjadi dua bagian, satu untuk melatih model (train set) dan satu lagi untuk menguji model (test set). Contohnya, membagi dataset menjadi 70% untuk data latih dan 30% untuk data uji (Yoga Nanda Khoiril Umat, Dhiaz Rusyda Nafsyi, Diana Kusumaningsih, 2024). Ketiga, leave-one-out cross-validation (LOOCV) yaitu bentuk khusus dari cross-validation di mana satu contoh digunakan untuk menguji model, dan sisanya digunakan untuk melatih model. Ini diulang untuk setiap contoh dalam dataset. Contohnya, Jika dataset memiliki 100 contoh, model dilatih 100 kali dengan masing-masing contoh bergantian digunakan sebagai data uji (Khoirunnisa and Lestari, 2023).



Gambar 8.6 Teknik Evaluasi Model

d. Visualisasi Hasil Evaluasi

Visualisasi membantu dalam interpretasi hasil evaluasi model dan memberikan wawasan tambahan. Pertama, confusion matrix adalah matriks yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas. Contohnya, menampilkan jumlah True Positives, True Negatives, False Positives, dan False Negatives untuk model klasifikasi. Kedua, ROC Curve (Receiver Operating Characteristic Curve) adalah grafik yang menunjukkan trade-off antara true positive rate dan false positive rate untuk berbagai threshold klasifikasi. Contohnya, menggunakan ROC curve untuk mengevaluasi kinerja model klasifikasi biner. Ketiga, Precision-Recall Curve adalah grafik yang menunjukkan hubungan antara presisi dan recall untuk berbagai threshold klasifikasi. Contohnya, menggunakan precision-recall curve untuk model klasifikasi dengan data yang tidak seimbang (Sutaji, Fatichah and Navastara, 2016).



Gambar 8.7 Visualisasi Hasil Evaluasi

Penerapan analisis variabel dalam konteks machine learning dan data mining mencakup serangkaian langkah dan teknik yang bertujuan untuk memahami, memodelkan, dan memprediksi data. Proses ini melibatkan beberapa tahapan penting, mulai dari pemilihan dan ekstraksi fitur hingga evaluasi model. Penerapan analisis variabel dalam machine learning dan data mining adalah proses yang kompleks namun esensial untuk memperoleh wawasan yang mendalam dari data. Melalui pemilihan fitur, ekstraksi fitur, penerapan algoritma machine learning, dan evaluasi model, kita dapat membangun model yang akurat dan dapat diandalkan. Data mining membantu dalam mengidentifikasi pola tersembunyi dan mendukung pengambilan keputusan yang lebih baik di berbagai industri, sehingga meningkatkan efisiensi operasional dan memberikan keunggulan kompetitif. Evaluasi model dalam data mining adalah proses penting yang memastikan kinerja dan keandalan model yang telah dibangun. Evaluasi ini melibatkan pembagian data menjadi data latih dan data uji, pelatihan model, pengujian model, dan penggunaan berbagai metrik seperti akurasi, presisi, recall, F1-score, MAE, MSE, dan RMSE. Teknik evaluasi seperti

cross-validation, train-test split, dan LOOCV digunakan untuk menilai kinerja model secara menyeluruh. Visualisasi hasil evaluasi, seperti confusion matrix, ROC curve, dan precision-recall curve, membantu dalam interpretasi kinerja model. Dengan evaluasi yang tepat, kita dapat memahami kekuatan dan kelemahan model, serta membuat keputusan yang lebih baik dalam penerapan model di aplikasi nyata.

Bab 9

Pengembangan Instrumen Penelitian

9.1 Instrumen Penelitian

Instrumen penelitian merupakan alat yang digunakan untuk mengumpulkan data penelitian, baik yang bersifat kualitatif maupun kuantitatif (Jailani, 2023). Data kualitatif meliputi gambar, kata, dan benda lain yang tidak berbentuk angka, sementara data kuantitatif berupa angka. Dalam penelitian kualitatif, peneliti sendiri berperan sebagai instrumen utama (Fauzi & Setiawan, 2020). Oleh karena itu, dalam konteks ini, yang dimaksud dengan instrumen penelitian adalah instrumen untuk penelitian kuantitatif.

Data kuantitatif dapat diklasifikasikan menjadi dua jenis utama: data nominal dan data kontinu (Handayani, 2023). Data dikatakan berskala nominal jika angka-angka digunakan untuk tujuan identifikasi, yaitu untuk membedakan antara jenis subyek yang berbeda. Dalam konteks ini, angka hanya berfungsi untuk menunjukkan perbedaan antara objek atau subyek tanpa adanya informasi mengenai urutan atau ukuran. Sebaliknya, data kontinu mencakup beberapa skala seperti ordinal, interval, dan rasio. Data berskala ordinal memungkinkan untuk melakukan peringkat atau urutan, data berskala interval menyediakan informasi tentang jarak yang konsisten antara nilai-nilai, dan data berskala rasio mengandung semua fitur skala interval

ditambah dengan adanya titik nol mutlak, memungkinkan perbandingan proporsional (Umami, 2021).

Data kuantitatif umumnya diperoleh melalui proses pengukuran, yang melibatkan pemberian angka pada subyek, objek, atau atribut lainnya. Dalam hal ini, instrumen penelitian sering kali disebut sebagai alat ukur. Alat ukur ini dapat dibedakan menjadi dua jenis utama: tes dan nontes. Alat ukur dikategorikan sebagai tes jika berisi pertanyaan-pertanyaan yang memiliki jawaban benar atau salah. Sebaliknya, alat ukur yang tidak memiliki jawaban benar atau salah, seperti skala, angket, atau inventori, termasuk dalam kategori nontes. Skala digunakan untuk mengukur intensitas atau tingkat, angket untuk mengumpulkan pendapat atau persepsi, dan inventori untuk mendokumentasikan berbagai atribut atau karakteristik (Adil *et al*, 2023).

Skala umumnya digunakan untuk mengukur konstruk atau konsep psikologis, seperti sikap, minat, motivasi, pendapat, dan berbagai atribut psikologis lainnya. Skala ini memungkinkan penilaian yang lebih mendalam terhadap berbagai aspek internal dari individu. Di sisi lain, angket digunakan untuk mengumpulkan data tentang fakta atau informasi yang dianggap sebagai fakta, seperti tingkat pendidikan, jumlah anggota keluarga, penghasilan bulanan, dan sejenisnya. Angket ini berfungsi untuk mendapatkan informasi kuantitatif yang lebih jelas dan terukur mengenai aspek-aspek kehidupan yang bersifat obyektif. Sementara itu, inventori berfokus pada pengungkapan kepemilikan benda fisik atau nyata, seperti jumlah kursi, meja, atau barang-barang lainnya. Inventori ini digunakan untuk mendokumentasikan atau mengukur aset fisik yang ada pada subyek atau dalam suatu konteks tertentu.

9.2 Syarat Instrumen Yang Baik

Syarat utama dari instrumen yang efektif adalah memiliki validitas dan reliabilitas. Validitas merujuk pada sejauh mana sebuah alat ukur benar-benar dapat mengukur aspek yang dimaksudkan untuk diukur (Makbul, 2021). Validitas umumnya bersifat kontinu, yang berarti tidak ada kategori valid atau tidak valid secara mutlak, melainkan berada pada spektrum tingkat keakuratan. Validitas suatu instrumen juga bersifat spesifik terhadap tujuan tertentu; dengan kata lain, sebuah instrumen dapat dianggap valid untuk mengukur atribut A, tetapi belum tentu valid untuk mengukur atribut B. Ini berarti bahwa validitas perlu dievaluasi dalam konteks penggunaan tertentu dan tidak dapat digeneralisasi secara otomatis untuk atribut lain.

Dalam pengukuran atribut psikologis, mencapai validitas seperti yang telah dijelaskan sebelumnya sering kali merupakan tantangan besar. Hal ini dapat dipahami mengingat bahwa pengukuran variabel psikologis dan sosial cenderung mengandung lebih banyak kesalahan dibandingkan dengan pengukuran variabel fisik (Saifuddin, 2020). Kesulitan dalam menentukan validitas yang benar-benar akurat disebabkan oleh kompleksitas dan variabilitas inherent dalam aspek-aspek psikologis dan sosial. Oleh karena itu, sebagai alternatif untuk menetapkan validitas yang absolut, sering kali dilakukan estimasi validitas instrumen melalui metode perhitungan tertentu. Ini melibatkan penggunaan teknik statistik dan analisis untuk memperkirakan sejauh mana instrumen tersebut mampu mengukur atribut yang dimaksudkan secara efektif.

Ada tiga tipe validitas, yakni: (1) validitas prediktif, (2) validitas isi, dan (3) validitas konstruk (Arifin, 2017). Validitas prediktif atau ada juga yang menyebut dengan validitas kriteria terkait dicari saat instrumen akan digunakan untuk mengestimasi beberapa bentuk tingkah laku penting yang ada di luar dari hasil pengukuran instrumen itu sendiri. Atau, validitas

prediktif diestimasi manakala instrumen dimaksudkan untuk berfungsi sebagai prediktor bagi performansi di waktu yang akan datang (Kartowagiran, 2018). Dalam analisis validitas prediktif, performansi yang hendak diprediksikan disebut dengan kriteria. Besar kecilnya harga estimasi validitas prediktif suatu instrumen digambarkan dengan keofisien korelasi antara prediktor dengan kriteria tersebut.

Validitas isi dari sebuah instrumen mengacu pada sejauh mana item-item dalam instrumen tersebut mewakili komponen-komponen yang termasuk dalam keseluruhan cakupan objek yang hendak diukur, serta sejauh mana item-item tersebut mencerminkan karakteristik perilaku yang dimaksudkan untuk diukur. Validitas ini melibatkan dua aspek utama: representasi, yaitu sejauh mana item-item mencakup seluruh komponen yang relevan, dan relevansi, yaitu sejauh mana item-item mencerminkan ciri-ciri perilaku yang ingin diukur. Penilaian validitas isi dilakukan dengan cara memeriksa apakah item-item dalam instrumen sudah mewakili komponen yang akan diukur. Proses ini umumnya melibatkan penilaian subyektif dari para ahli atau penilai dan tidak melibatkan perhitungan statistik, melainkan analisis rasional.

Sementara itu, validitas konstruk mengukur sejauh mana instrumen dapat mengungkapkan suatu trait atau konstruk teoritis yang hendak diukur. Pengujian validitas konstruk adalah proses yang berkelanjutan, mengikuti perkembangan konsep trait yang diukur. Dalam psikologi, perubahan dan perkembangan konsep adalah hal yang wajar karena variabel-variabel yang diukur sering kali merupakan konsep hipotetik yang sulit dioperasionalkan secara langsung. Validitas konstruk sangat penting untuk tes yang mengukur trait tanpa kriteria eksternal yang jelas. Prosedur validasi konstruk dimulai dengan identifikasi dan batasan mengenai variabel yang diukur, kemudian dirumuskan dalam bentuk konstruk logis berdasarkan teori terkait. Dari teori ini, ditariklah konsekuensi praktis yang

diharapkan dari hasil pengukuran pada kondisi tertentu, dan konsekuensi tersebut diuji. Jika hasil pengukuran sesuai dengan harapan, maka instrumen dianggap memiliki validitas konstruk yang baik.

Selain itu, validitas konstruk juga dapat diperoleh melalui teknik analisis faktor. Analisis faktor adalah metode matematis kompleks yang digunakan untuk menganalisis hubungan antara variabel dan menjelaskan hubungan tersebut dalam bentuk kelompok variabel yang disebut faktor. Validitas yang diperoleh melalui analisis faktor dikenal sebagai validitas faktorial.

Syarat penting lainnya untuk instrumen yang baik adalah reliabilitas. Reliabilitas mengacu pada konsistensi pengukuran, yaitu sejauh mana hasil tes atau penilaian tetap stabil dari satu pengukuran ke pengukuran lainnya. Hasil penilaian memberikan ukuran performa terbatas yang diperoleh pada waktu tertentu. Kecuali jika pengukuran menunjukkan konsistensi yang baik di berbagai kesempatan, dengan penilai yang berbeda, atau sampel yang berbeda, serta domain performa yang sama, hasil penilaian yang sepenuhnya konsisten sulit dicapai.

Hasil penilaian dipengaruhi oleh berbagai faktor yang dapat menyebabkan variasi skor. Jika penilaian dilakukan pada kelompok yang sama secara berulang dalam waktu dekat, variasi skor mungkin disebabkan oleh fluktuasi memori, perhatian, usaha, kelelahan, ketegangan emosional, dan tebak-tebakan. Sebaliknya, jika penilaian dilakukan dengan jarak waktu yang lebih lama, variasi skor mungkin dipengaruhi oleh pengalaman belajar baru, perubahan kesehatan, atau faktor lupa. Variasi skor juga bisa terjadi jika tes esai atau penilaian jenis performa lainnya dilakukan oleh penilai yang berbeda.

Allen dan Yen (1979) mengidentifikasi tiga metode utama untuk mengestimasi koefisien reliabilitas, yaitu: (1) metode tes ulang, (2) metode tes paralel, dan (3) metode konsistensi internal

(Khumaedi, 2012). Masing-masing metode ini memberikan estimasi koefisien reliabilitas yang berbeda, dan hasilnya hanya merupakan estimasi karena nilai sebenarnya dari koefisien ini tidak dapat diamati secara langsung.

Metode tes ulang melibatkan pengambilan tes yang sama dua kali dengan interval tertentu dan kemudian mengkorelasikan hasilnya untuk memperoleh estimasi reliabilitas. Metode ini memberikan estimasi yang beralasan, namun memiliki beberapa kelemahan. Kelemahan pertama adalah potensi efek *carry-over*, di mana hasil tes pertama dapat mempengaruhi hasil tes kedua. Kelemahan kedua berkaitan dengan interval waktu; interval yang terlalu pendek dapat menyebabkan efek *carry-over* dalam memori, sedangkan interval yang terlalu lama dapat menyebabkan perubahan informasi.

Metode tes paralel mengestimasi reliabilitas dengan mengukur korelasi antara nilai dari dua tes paralel. Namun, dalam praktiknya, tes paralel hanya merupakan konsep teoritis, sehingga sulit untuk membuktikan bahwa dua tes benar-benar paralel. Sebagai alternatif, bentuk tes alternatif sering digunakan, yaitu dua bentuk tes yang dirancang untuk menjadi paralel, dengan rerata skor, variansi, dan korelasi dengan pengukuran lain yang serupa. Korelasi antara skor pada tes pertama dan tes alternatif memberikan estimasi reliabilitas untuk keduanya.

Reliabilitas konsistensi internal diukur dengan satu kali pelaksanaan tes, sehingga mengatasi masalah yang terkait dengan metode tes ulang. Salah satu metode yang luas digunakan dalam pendekatan ini adalah estimasi reliabilitas belah dua, di mana tes dibagi menjadi dua bagian yang dianggap sebagai tes alternatif atau paralel. Keuntungan utama dari pendekatan ini adalah hanya memerlukan satu kali pelaksanaan tes untuk menghitung estimasi reliabilitas. Namun, metode konsistensi internal tidak cocok jika tes tidak dapat dibagi menjadi bagian-bagian yang paralel atau jika tes tidak memiliki

item-item independen yang dapat dipisahkan.

Allen dan Yen dalam Retnawati *et al* (2011) mencatat tiga metode untuk membagi tes menjadi dua bagian: (1) metode gasal-genap, di mana item-item tes dikelompokkan berdasarkan nomor gasal dan genap, dan (2) metode belah dua sesuai dengan nomor urut. Teknik ini juga dapat diperluas untuk membagi tes menjadi lebih dari dua komponen, seperti dengan memodifikasi metode gasal-genap untuk membuat tiga komponen dari tes yang memiliki sembilan item.

9.3 Langkah-Langkah Pengembangan Instrumen

Langkah pertama dalam mengembangkan instrumen adalah melakukan kajian teoritik tentang substansi yang akan diukur. Peneliti harus menentukan definisi konseptual, yang kemudian dijabarkan menjadi definisi operasional. Definisi operasional ini selanjutnya dikembangkan menjadi indikator dan item-item instrumen. Enam langkah utama dalam mengembangkan instrumen alat ukur:

1. Menyusun spesifikasi alat ukur, termasuk kisi-kisi dan indikator.
2. Menulis pertanyaan.
3. Menelaah pertanyaan.
4. Melakukan ujicoba.
5. Menganalisis butir instrumen.
6. Merakit instrumen dan memberikan label.

Spesifikasi alat ukur mencakup tujuan pengukuran, kisi-kisi instrumen, skala pengukuran, dan panjang instrumen. Oleh karena itu, menentukan spesifikasi alat ukur berarti menetapkan tujuan instrumen, mengembangkan kisi-kisi instrumen, memilih skala pengukuran, dan menentukan panjang instrumen. Instrumen dapat dibagi menjadi tes dan nontes. Kisi-kisi

instrumen untuk tes dan nontes memiliki penyusunannya:

1. Kisi-Kisi Instrumen/Tes : Setelah menetapkan tujuan tes, langkah berikutnya adalah menyusun kisi-kisi tes, yang merupakan tabel matrik berisi spesifikasi soal. Kisi-kisi ini mencakup tujuan, standar kompetensi, kompetensi dasar, materi pokok, dan bentuk serta jenis penilaian. Standar kompetensi dijabarkan menjadi kompetensi dasar, yang kemudian dipecah menjadi indikator untuk membuat butir-butir instrumen. Terdapat dua bentuk tes yang umum digunakan: tes obyektif dan tes uraian. Masing-masing memiliki kelebihan dan kekurangan.
2. Kisi-Kisi Instrumen Nontes: Penyusunan instrumen nontes dimulai dengan menentukan definisi konseptual, yang kemudian dijabarkan menjadi definisi operasional. Dari definisi operasional ini, dibuatlah indikator dan item-item instrumen. Instrumen nontes meliputi skala, angket, dan inventori.

Skala digunakan untuk mengukur konstruk atau konsep psikologis seperti: sikap, minat, motivasi, pendapat, dan *trait* lainnya, sedangkan angket digunakan untuk mengukur fakta, atau yang dianggap fakta seperti: pendidikan terakhir, jumlah anggota, penghasilan setiap bulan, dll. Sementara itu, inventori digunakan untuk mengungkap kepemilikan benda nyata, seperti: jumlah kursi, jumlah meja, dll. Secara ringkas, hubungan antara tujuan, metode dan instrumen yang digunakan pada Tabel berikut.

Tabel 9.1 Hubungan Antara Tujuan, Metode, dan Instrumen

Tujuan untuk mengungkap:	Metode	Instrumen yg digunakan
- perilaku, kebiasaan, ketrampilan	observasi, wawancara mendalam	lembar observasi, lembar penilaian, catatan, peneliti sendiri
- potensi termasuk di dalamnya unjuk kerja	tes, perintah mengerjakan	soal tes, lembar perintah dilengkapi dg lembar observasi/ lembar penilaian
- afektif: motivasi, sikap, minat , kesukaan, dll	wawancara, survei	pedoman wawancara, skala
- data pribadi, data nyata	wawancara, survei	angket, inventori,
- data yang lalu, data sekunder	dokumentasi	daftar dokumen

Tabel di atas menjelaskan bahwa metode dan instrumen yang digunakan harus mengacu pada tujuan pengukuran. Hal ini penting agar tidak terjadi kesalahan pengukuran.

9.4 Cara Memvalidasi Instrumen

Di bagian sebelumnya, telah dibahas secara mendetail tentang pengertian dan berbagai jenis validitas serta reliabilitas instrumen. Secara ringkas, metode untuk memvalidasi dan mengestimasi reliabilitas instrumen dapat diterapkan dan dilihat

melalui contoh instrumen berikut. Proses ini mencakup berbagai teknik untuk memastikan bahwa instrumen yang digunakan benar-benar mengukur apa yang dimaksudkan untuk diukur (validitas) dan memberikan hasil yang konsisten (reliabilitas). Penggunaan instrumen yang telah divalidasi dan diuji reliabilitasnya akan meningkatkan keakuratan dan keandalan hasil penelitian atau penilaian yang dilakukan.

Tabel 9.2 Jenis dan Cara Memvalidasi

Jenis Validitas	Cara Memvalidasi	Keterangan
Validitas isi: validitas kurikulum, validitas tampak	<ul style="list-style-type: none"> - menggunakan kisi-kisi - konsultasi keahlinya 	-tanpa menggunakan teknik statistik
Validitas kriteria terkait atau validitas empirik: validitas prediktif, validitas konkuren	-mengkorelasikan dengan data di masa datang	Korelasi product moment
Validitas konstruk: validitas faktor	-mengkorelasikan skor butir dengan total	<ul style="list-style-type: none"> - analisis faktor - product moment - analisis butir

Tabel 9.3 Jenis Reliabilitas

Jenis Reliabilitas	Prosedur	Teknik yang dipakai
Internal Consistency: 1. data ordinal 2. data nominal	1 dan 2, tes satu kali, kemudian dianalisis ataudiestimasi reliabilitasnya 3 tes sekali, kemudian skor dibelah dua dan diestimasi	1. Koef. Alpha 2. KR 20, KR 21 3. Spearman Brown
Stabilitas	Tes dua kali dengan soal sama, kemudian hasilnya dikorelasikan.	Product moment dan korelasi intra kelas
Ekivalen	Beri tes dua kali dengan soal yang berbeda kemudian dikorelasikan	Product moment dan korelasi intra kelas

Tabel di atas menunjukkan bahwa untuk mengestimasi validitas dan reliabilitas instrumen diperlukan kerja yang sangat hati-hati, harus diupayakan agar proses dan estimasi ini dilaksanakan dengan sebaik-baiknya.

Hal-hal yang harus diperhatikan adalah: (1) butir instrumen harus sesuai indikator, (2) butir ditulis secara singkat dan jelas, (3) pilihan jawaban yang berbentuk angka atau waktu, sebaiknya diurutkan, (4) dalam satu komponen, setiap butir diberi skor sama (skor sama tidak berarti pilihan jawabannya sama), dan (5) butir ditulis dengan menggunakan bahasa baku. Selain itu, untuk menarik responden agar mau merespon dengan baik maka instrumen sebaiknya: (1) dikemas dalam bentuk yang menarik, misal dalam bentuk buku yang agak kecil, (2) diusahakan jumlah butir untuk setiap jenis responden tidak terlalu banyak

(maksimum 40 butir), dan (3) diusahakan butir pertanyaan dan jawaban pada halaman yang sama.



Bab 10

Etika dalam Penelitian dan Analisis Data

10.1 Pendahuluan

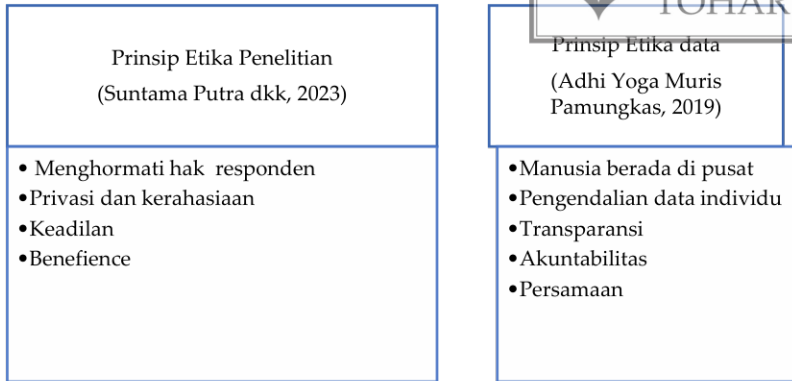
Perkembangan teknologi informasi dan komunikasi telah membawa kita ke era baru yang sering disebut dengan era big data dan kecerdasan buatan (AI). Big data merujuk pada volume data yang sangat besar dan beragam yang dihasilkan dalam berbagai aktivitas manusia sehari-hari, mulai dari aktivitas di media sosial hingga transaksi bisnis. AI, di sisi lain, adalah teknologi yang memungkinkan mesin untuk belajar dari data, membuat keputusan, dan bahkan merumuskan prediksi yang sebelumnya hanya dapat dilakukan oleh manusia. Kemajuan dalam big data dan AI telah membuka peluang luar biasa dalam berbagai bidang, seperti kesehatan, ekonomi, pendidikan, dan pemerintahan. Namun, di balik potensi yang besar ini, ada tantangan besar yang harus dihadapi, terutama yang berkaitan dengan etika dalam penelitian dan analisis data.

Etika dalam penelitian dan analisis data merupakan pedoman yang harus ditaati peneliti sejak perencanaan penelitian hingga pelaporan. Sedangkan etika dalam konteks big data dan AI mencakup berbagai isu yang kompleks dan saling terkait. Isu-isu ini mencakup privasi individu, keamanan data, keadilan dalam algoritma, transparansi, dan akuntabilitas. Dengan volume data yang begitu besar dan penggunaan AI yang

semakin meluas, penting untuk memastikan bahwa penggunaan teknologi ini dilakukan dengan cara yang etis dan bertanggung jawab. Maka, pentingnya etika digunakan untuk menentukan seberapa baik atau buruk aspek tertentu dari kegiatan penelitian.

10.2 Prinsip -Prinsip Etika Penelitian

Etika secara etimologis berasal dari bahasa Yunani "Ethos dan Ethikos", yang berarti "sifat, watak, adat, kebiasaan," dan "etika", yang berarti "susila, keadaban, atau kelakuan dan perbuatan yang baik." Dalam etika, sudut pandang normatif mengacu pada manusia dan tindakannya. (Suntama Putra dkk, 2023). Menurut Kamus Besar Bahasa Indonesia (KBBI) etika adalah ilmu yang mempelajari baik dan buruk, hak dan kewajiban moral. Selain itu Etika adalah kumpulan asas/nilai yang berkenaan dengan akhlak. Etika berkaitan dengan norma untuk berperilaku, memisahkan apa yang seharusnya dan apa yang tidak seharusnya dilakukan. Dalam konteks penelitian, etika digunakan untuk menentukan seberapa baik atau buruk aspek tertentu dari kegiatan penelitian. Fabrikasi, falsifikasi, dan plagiarisme adalah contoh kecurangan yang secara resmi dikategorikan sebagai pelanggaran etika dalam penelitian (Agung & Zarah, 2016). Dari pengertian sebelumnya dapat disimpulkan etika yaitu nilai mengenai benar dan salah yang dianut masyarakat. Sedangkan etika penelitian adalah etika mengatur berbagai hal yang harus menjadi pedoman perilaku peneliti sejak menyusun desain penelitian, mengumpulkan data di lapangan (melakukan wawancara, memberikan angket, melakukan pengamatan, meminta data pendukung), ketika menyusun laporan penelitian, sampai mempublikasikan hasil penelitian. (Martono, 2015)



Gambar 10.1 Prinsip Etika Penelitian dan Etika Data

Berikut merupakan prinsip-prinsip etika secara umum yang dapat menjadi pertimbangan peneliti dalam penelitiannya. (Suntama Putra dkk, 2023).

1. Menghormati & menghargai harkat martabat manusia sebagai subjek penelitian.

Dengan menghormati hak responden, peneliti memberikan ruang bagi responden untuk mendapatkan informasi lengkap dan akurat sebagai dasar pembuatan

keputusan keikutsertaannya dalam penelitian. Persuasi peneliti kepada responden dijaga agar tetap menghormati hak responden. Seorang peneliti wajib memperhatikan hak-hak subyek penelitian untuk mendapatkan informasi yang jelas dan terbuka berkenaan dengan jalannya penelitian serta memiliki kebebasan untuk menentukan pilihan dan tidak ada intervensi maupun paksaan untuk berpartisipasi dalam kegiatan penelitian. Oleh karena itu perlu dipersiapkan formulir persetujuan (informed consent) oleh peneliti kepada subjek penelitian.

2. Menghormati privasi dan kerahasiaan subjek penelitian.

Privasi adalah hak fundamental yang harus dilindungi dalam setiap proses penelitian dan analisis data. Dengan

meningkatnya kapasitas untuk mengumpulkan dan menganalisis data dalam jumlah besar, risiko terhadap privasi individu juga meningkat. Prinsip ini menekankan pentingnya melindungi data pribadi individu dan memastikan bahwa data tersebut hanya digunakan untuk tujuan yang sah dan disetujui. Setiap manusia memiliki hak-hak dasar individu termasuk privasi dan kebebasan individu. Oleh karenanya, seorang peneliti harus menggunakan coding atau inisial, jika yang subjek penelitian tidak berkenan untuk dipublikasikan.

3. Memegang prinsip keadilan & kesetaraan.

Prinsip keadilan berlaku pada pemilihan responden penelitian yang dilakukan dengan adil. Seluruh responden secara adil merasakan manfaat dan Semua subjek penelitian harus diperlakukan dengan baik, sehingga terdapat keseimbangan antara manfaat dan risiko yang dihadapi oleh subjek penelitian. Jadi harus diperhatikan risiko fisik, mental maupun sosial.

4. Memperhitungkan dampak positif maupun negatif dari penelitian

Peneliti melaksanakan penelitian sesuai dengan prosedur penelitian guna mendapatkan hasil yang bermanfaat semaksimal mungkin bagi subyek penelitian dan dapat digeneralisasikan di tingkat populasi (beneficence). Peneliti meminimalisasi dampak yang merugikan bagi subyek. Apabila intervensi penelitian berpotensi mengakibatkan cedera atau stres tambahan maka subyek dikeluarkan dari kegiatan penelitian untuk mencegah terjadinya cedera.

Sedangkan jika berbicara pada dalam konteks big data dan AI maka elemen penting adalah data. Data adalah bahan bakar bagi teknologi kecerdasan buatan. Menurut Luciano Floridi etika data dapat dibagi menjadi tiga area yaitu etika mengenai data itu sendiri, etika algoritma, dan etika mengenai praktik terkait. Etika mengenai data meliputi pembuatan, pencatatan, pemrosesan,

penyebaran, pembagian, dan penggunaan data tersebut. Etika algoritma meliputi artificial intelligence, artificial agents, machine learning, dan robots. Sedangkan etika mengenai praktik lain berkaitan dengan peretasan, pemrograman, dan kode profesional. Berikut prinsip etika data meliputi: (Adhi Yoga Muris Pamungkas dkk, 2019)

1. Manusia berada di pusat

Hardware proses komputer atau kepingan software yang memiliki pemikiran dan kinerja yang sesuai dengan program yang diberikan. Tidak seperti hardware, proses komputer, dan software, manusia merupakan aset yang unik. Mereka memiliki otak yang membuat mereka memiliki empati, emosi, intuisi, dan kreativitas yang dapat membuat perbedaan pemikiran dan kinerja mereka dengan pemikiran mesin yang kaku. Oleh karena itu manusia memiliki status yang lebih tinggi daripada mesin. Oleh karena itu, dapat disimpulkan bahwa manusia merupakan pusat dari penggunaan data sehingga seharusnya yang mendapatkan manfaat utama dalam penggunaan data.

2. Pengendalian data individu

Setiap manusia memiliki kendali utama dalam penggunaan data mereka. Mereka harus tahu tempat, waktu, dan bagaimana data mereka diproses. Manusia harus dapat mengendalikan data mereka sehingga mereka dapat mendapatkan manfaat dari data mereka. Oleh karena itu, nasib individu harus lebih diprioritaskan dibandingkan semua proses data, sehingga menuntut semua individu untuk berperan aktif dalam penggunaan data mereka.

3. Transparansi

Tujuan dan kepentingan penggunaan data harus harus transparan dan dapat dijelaskan sehingga dapat dipahami oleh individu. Hal ini agar individu dapat mengetahui

dampak risiko yang akan didapatkan akibat penggunaan data tersebut.

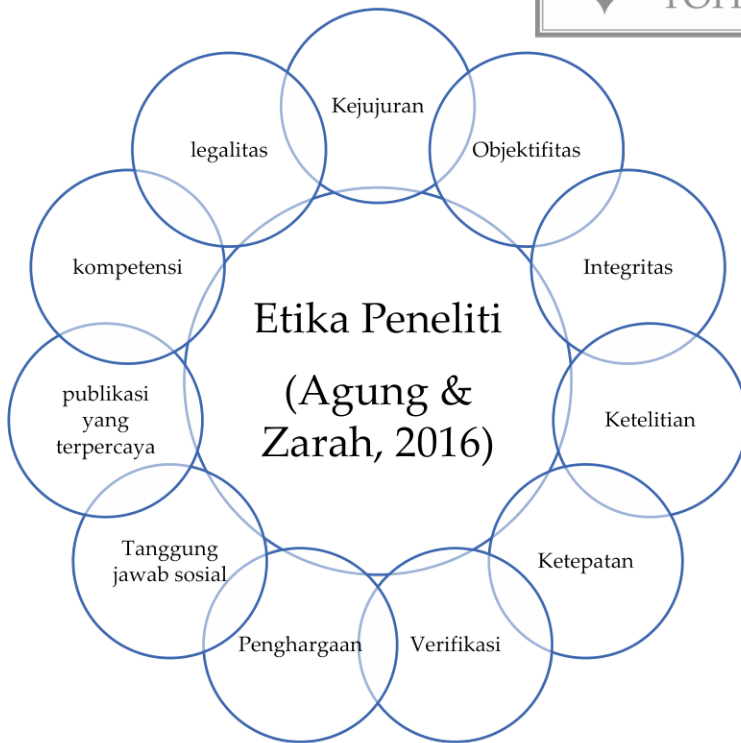
4. Akuntabilitas

Akuntabilitas merupakan hal yang penting dalam penggunaan data. Akuntabilitas menentukan penggunaan data dan perlindungan data yang bersifat pribadi. Tujuan dari akuntabilitas ini untuk memitigasi dampak risiko dalam tatanan sosial akibat dari penggunaan data yang bersifat pribadi.

5. Persamaan

Pengolahan data harus mempertimbangkan kesadaran hubungan kekuasaan masyarakat. Jangan sampai ada intimidasi, diskriminasi dan stigmatisasi pada masyarakat akibat dari pengolahan data milik mereka. Oleh karena itu, pengguna data harus memberikan perhatian khusus kepada masyarakat yang rentan terhadap pengolahan data terhadap mereka.

Selain prinsip yang harus ditaati dalam melakukan penelitian. Seorang peneliti yang baik adalah peneliti yang senantiasa menerapkan etika dalam penelitian ilmiahnya, sehingga hasil penelitiannya dapat dipertanggungjawabkan secara moril. (Agung & Zarah, 2016)



Gambar 10.2 Etika Peneliti

1. Kejujuran

Kejujuran yaitu jujur dalam pengumpulan referensi, pengumpulan data, pelaksanaan metode dan prosedur penelitian, publikasi hasil, dan sebagainya, jujur pada kekurangan atau kegagalan metode yang dilakukan (Agung & Zarah, 2016). Dalam mengkomunikasikan penelitian ilmiah hendaklah jujur melaporkan data, hasil, metode dan prosedur, Tidak boleh membuat, memalsukan/merubah data, tidak dibenarkan menipu rekan kerja, sponsor penelitian, atau masyarakat umum (Raihan, 2017).

2. Objektivitas

Objektivitas yaitu upaya untuk minimalisasi kesalahan/bias dalam melakukan analisis penelitian, penelitian harus

memiliki objektivitas baik dalam karakteristik maupun prosedurnya. objektivitas dicapai melalui keterbukaan, terhindar dari bias dan subjektivitas, dalam prosedurnya, penelitian menggunakan teknik pengumpulan dan analisis data yang memungkinkan dibuat interpretasi yang dapat dipertanggungjawabkan (Agung & Zarah, 2016). Berusahalah untuk menghindari bias dalam analisis data, interpretasi data, dan lainnya dari penelitian objektivitas diharapkan atau dibutuhkan, hindari atau kurangi bias atau penipuan diri sendiri, mengungkapkan kepentingan pribadi atau finansial yang mungkin mempengaruhi penelitian (Raihan, 2017).

3. Integritas

Integritas yaitu suatu upaya untuk selalu menjaga konsistensi pikiran dan perbuatan dalam proses peneliti (Agung & Zarah, 2016). Menepati janji dan kesepakatan, bertindak dengan tulus, berjuang untuk konsistensi pemikiran dan tindakan (Raihan, 2017).

4. Ketepatan

Ketepatan, bahwa suatu penelitian juga harus memiliki tingkat ketepatan (precision), secara teknis instrumen pengumpulan data harus memiliki validitas dan reliabilitas yang memadai, desain penelitian, pengambilan sampel dan teknik analisis datanya adalah tepat (Agung & Zarah, 2016). Hindari kesalahan dan kelalaian yang ceroboh, hati-hati dan kritis memeriksa pekerjaan anda sendiri dan pekerjaan rekan-rekan anda, buat catatan kegiatan penelitian yang bagus, seperti pengumpulan data, desain penelitian, dan korespondensi dengan agensi atau jurnal (Raihan, 2017).

5. Tanggung jawab sosial

Tanggung jawab sosial, yaitu mengupayakan agar penelitian berguna demi kemaslahatan masyarakat, meningkatkan taraf hidup, memudahkan kehidupan dan meringankan beban

hidup masyarakat, peneliti juga bertanggung jawab melakukan pendampingan bagi masyarakat yang ingin mengaplikasikan hasil penelitiannya (Agung & Zarah, 2016). Upayakan untuk mempromosikan kebaikan sosial dan mencegah atau mengurangi kerugian sosial melalui penelitian, pendidikan publik, dan advokasi (Raihan, 2017).

6. Publikasi terpercaya

Publikasi yang terpercaya, yaitu menghindari publikasi penelitian yang sama atau berulang-ulang ke berbagai media (jurnal, seminar) (Agung & Zarah, 2016). Publikasikan untuk memajukan penelitian dan dapat digunakan masyarakat (Raihan, 2017).

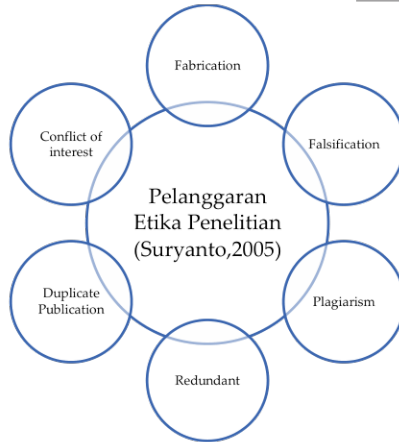
7. Kompetensi

Kompetensi, bahwa penelitian harus dilakukan oleh orang yang memiliki pengetahuan dan keterampilan sesuai bidang ilmunya (Agung & Zarah, 2016). Menjaga dan meningkatkan kompetensi dan keahlian profesional melalui pendidikan dan pembelajaran sepanjang hayat, mengambil langkah untuk mempromosikan kompetensi dalam sains secara keseluruhan (Raihan, 2017).

8. Legalitas

Legalitas, yaitu memahami dan mematuhi peraturan institusional dan kebijakan pemerintah yang terkait dengan penelitian yang dilakukan (Agung & Zarah, 2016). Mengetahui dan mematuhi hukum dan kebijakan kelembagaan dan pemerintah yang relevan (Raihan, 2017).

10.3 Pelanggaran Etika Penelitian dan Analisis Data



Gambar 10.3 Enam Kategori Pelanggaran Etika Penelitian

Menurut Suryanto, 2005 terdapat enam kategori pelanggaran etik dalam penelitian yaitu:

1. *Fabrication*

Adalah mencatat dan mempresentasikan dalam berbagai format sesuatu data yang fiktif. Fabrikasi atau pemalsuan data berupa pengaturan atau penyesuaian hasil penelitian dan mis-interpretasi data adalah suatu perbuatan yang tidak dapat dibenarkan dan tidak dapat ditolerir. Fabrikasi dan misinterpretasi data serta berbagai bentuk kecurangan lainnya dapat ditemukan dalam proses pengumpulan atau analisis data, pelaporan dan penghilangan data dengan maksud menyesatkan atau merusak catatan penelitian. Beberapa contoh pelaporan hasil penelitian yang tidak jujur seperti itu tidak dapat dibenarkan. Proposal penelitian, laporan kemajuan penelitian, presentasi, laporan penelitian internal, penulisan artikel jurnal, dan dokumen lain yang berkaitan dengan penelitian semuanya dianggap sebagai catatan data atau hasil temuan penelitian yang harus disampaikan secara benar dan jujur sesuai dengan kaidah

dan metode penelitian yang digunakan. (Suntama Putra dkk, 2023).

2. *Falsifications*

Menurut Adeleye & Adebamowo (2012), falsifikasi adalah suatu tindakan mengubah proses, bahan, ataupun peralatan penelitian, serta merekam/mencatat/melaporkan hasil penelitian palsu. Dengan demikian catatan/rekaman data hasil penelitian tersebut tidak menggambarkan hasil yang sesungguhnya. Secara lebih ringkas falsifikasi dapat di definisikan sebagai perekayasaan data serta informasi penelitian secara tidak sah dalam penelitian (mengubah untuk menipu). Tindakan falsifikasi sangat sulit dideteksi, karena praktik ini dilakukan sangat halus dan hampir tidak terlihat yang dihasilkan dari manipulasi yang cerdas, oleh karena itu tindakan ini sangat sulit dibuktikan. Hanya peneliti yang bersangkutan yang mengetahui secara pasti apakah mereka atau anggota tim peneliti yang melakukan falsifikasi. Falsifikasi mungkin dilakukan oleh peneliti itu sendiri, pengelola data (data manager), laboran, atau asisten laboran yang mencoba menyenangkan atasan dengan memberikan data diinginkan dan diharapkan oleh atasan ataupun peneliti itu sendiri.

3. *Plagiarism*

Menggunakan kata atau kata-kata gagasan, data atau hasil penelitian orang lain tanpa penghargaan kepada pemiliknya. Pencurian data atau gagasan lebih serius lagi karena hal ini tidak mungkin dilakukan tanpa disadari. Tindakan plagiarisme dalam penelitian merupakan suatu bentuk pelanggaran etika yang mendasar yang paling sering terjadi dalam proses penulisan karya ilmiah, baik itu disengaja maupun tidak. Plagiarisme yang dilakukan oleh seorang peneliti merupakan suatu perbuatan melanggar etika serius sehingga harus dihindari oleh semua peneliti. Pelanggaran

etika ini dapat dikategorikan pelanggaran tingkat sedang hingga berat, tergantung bentuk plagiarisme yang dilakukan oleh peneliti dan kriteria serta niatnya. Apabila plagiarisme ini dilakukan, pelaku/peneliti akan mendapatkan sanksi pelanggaran, mulai dari sanksi ringan hingga sanksi yang paling berat, termasuk mendapatkan sanksi penangguhan atau bahkan pembatalan kelulusan sarjana bagi mahasiswa dan konsekuensi bagi dosen seperti diberikannya Surat Peringatan hingga berpotensi dimasukkan ke dalam daftar cegah (blacklist) oleh Direktorat Jenderal Pendidikan Tinggi (Ditjen Dikti). (Suntama Putra dkk, 2023).

4. *Redundant (respetitive)*

Adalah pegulangan penerbitan sebagian atau semua yang sudah dipublikasikan sebelumnya, kecuali ada penelitian lanjutan dan ditemukan informasi baru.

5. *Duplicate publication*

Adalah publikasi satu artikel atau yang identik atau overlap secara substansial, tanpa ada ucapan terima kasih, dapat diklasifikasikan sebagai self plagiarism. Redundant dan duplication ini dapat cenderung untuk mengubah data yang ada sebagai bukti ilmiah. Apabila data dikutip dua kali misalnya, keluaran pada meta analisis yang dilakukan akan invalid.

6. *Conflict of interest*

Pejabat resmi suatu institusi kadang-kadang menghadapi hal yang tidak menyenangkan ini tetapi sukar dihindari. Konflik ini timbul bukan saja soal uang, publik, agama, tetapi dapat juga berkembang dengan posisi atau jabatan atau supervisi.

10.4 Praktik Terbaik dalam Etika Penelitian

Pengumpulan Data yang Etis

Untuk memastikan bahwa data dikumpulkan dengan cara yang etis, ada beberapa praktik terbaik yang dapat diikuti. **Pertama**, penting untuk mendapatkan persetujuan yang jelas dan diinformasikan dari individu yang datanya dikumpulkan. Ini berarti bahwa individu harus diberitahu tentang tujuan pengumpulan data, bagaimana data akan digunakan, dan apa risiko dan manfaatnya. Persetujuan harus diberikan secara sukarela dan tanpa tekanan.

Kedua, penting untuk melindungi kerahasiaan data. Ini termasuk menggunakan metode enkripsi untuk melindungi data selama penyimpanan dan transmisi, serta membatasi akses ke data hanya kepada pihak yang berwenang. Data pribadi harus disimpan dengan aman dan hanya digunakan untuk tujuan yang telah disetujui oleh individu yang bersangkutan.

Analisis Data yang Etis

Dalam analisis data, penting untuk memastikan akurasi dan kejujuran. Ini berarti bahwa data harus dianalisis dengan cara yang jujur dan tidak bias, dan hasil analisis harus dilaporkan dengan jujur dan transparan. Peneliti dan praktisi data harus bertanggung jawab atas cara mereka menganalisis data dan memastikan bahwa hasil analisis tidak dimanipulasi atau dipalsukan. Selain itu, penting untuk memastikan keadilan dalam analisis data.

Penggunaan Data yang Etis

Dalam penggunaan data, penting untuk memastikan bahwa data digunakan untuk tujuan yang sah dan bermanfaat, dan bahwa dampak negatif dari penggunaan data diminimalkan. Ini termasuk memastikan bahwa data tidak digunakan untuk tujuan yang melanggar hak-hak individu atau yang dapat merugikan kelompok-kelompok tertentu dalam masyarakat.

Selain itu, penting untuk memastikan akuntabilitas dalam penggunaan data. Peneliti dan praktisi data harus jujur tentang metode mereka, dan mereka harus bertanggung jawab atas cara mereka mengumpulkan, menganalisis, dan menggunakan data. Ini termasuk memastikan bahwa individu dan masyarakat memiliki akses ke informasi tentang bagaimana data mereka digunakan dan bagaimana keputusan yang mempengaruhi mereka dibuat.

Dalam era informasi yang semakin kompleks dan terhubung, penelitian ilmiah dan analisis data memainkan peran yang semakin krusial dalam membentuk kebijakan, inovasi teknologi, dan pemahaman kita tentang dunia. Namun, dengan meningkatnya volume dan kekayaan data yang tersedia, tanggung jawab etis para peneliti dan analis data menjadi lebih penting dari sebelumnya.

Etika penelitian adalah suatu sistem nilai dan prinsip-prinsip moral yang mengatur perilaku para peneliti dalam melakukan penelitian. Di era big data dan AI ini peran etika dalam penelitian dan analisis data menjadi semakin krusial. Perkembangan teknologi ini menawarkan potensi luar biasa untuk meningkatkan kehidupan manusia dalam berbagai aspek, seperti kesehatan, pendidikan, ekonomi, dan keamanan. Namun, bersama dengan manfaat ini, muncul pula tantangan dan risiko etis yang signifikan. Oleh karena itu, penting bagi peneliti untuk memahami dan menerapkan prinsip-prinsip etika dalam setiap tahap proses pengumpulan, analisis, dan penggunaan data.

Etika dalam penelitian dan analisis data tidak hanya berfokus pada aspek teknis, tetapi juga pada dampak sosial dan moral dari teknologi tersebut. Prinsip-prinsip seperti privasi dan kerahasiaan, akurasi dan kejujuran, keadilan dan akuntabilitas, serta beneficence harus menjadi landasan dalam setiap keputusan yang diambil. Sehingga dengan memahami dan mengikuti etika penelitian, maka dapat memastikan bahwa

penelitian yang dilakukan memberikan manfaat bagi masyarakat dan ilmu pengetahuan secara keseluruhan.



DAFTAR PUSTAKA

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.
- Abdussamad, H.Z. and Sik, M.S., 2021. Metode penelitian kualitatif. CV. Syakir Media Press.
- Adhi Yoga Muris Pamungkas dkk, 2019, Pentingnya Memahami Etika dalam Mengambil dan Mengolah Data. *Jurnal Komunikasi Profesional*. Volume 3 No 1, 31-38
- Adil, A., Liana, Y., Mayasari, R., Lamonge, A. S., Ristiyana, R., Saputri, F. R., ... & Rohman, M. M. (2023). Metode Penelitian Kuantitatif dan Kualitatif: Teori dan Praktik. Jakarta: Get Press indonesia.
- Agung, W. K., & Zarah, P. 2016. Metode Penelitian Kuantitatif. Panduan Buku.
- Agustiningasih, A., Findawati, Y. and Alnarus Kautsar, I. (2023) 'Classification of Vocational High School Graduates' Ability in Industry Using Extreme Gradient Boosting (Xgboost), Random Forest, and Logistic Regression', *Jurnal Teknik Informatika (Jutif)*, 4(4), pp. 977-985. Available at: <https://doi.org/10.52436/1.jutif.2023.4.4.945>.
- Ajija, Shochrul Rohmatul, Dyah wulan sari, Rahmat Heru Setianto et al (2011). Cara Cerdas Menguasai Eviews. Jakarta: Salemba Empat. ISBN, 9790611501, 9789790611504
- Alimy, J.I. et al. (2020) 'Pengaruh Intellectual Capital Terhadap Kinerja Keuangan: Dengan Variabel Moderasi Prospector Strategy Pada Perusahaan Manufaktur Yang Terdaftar Di Idx Periode 2016-2018', *Prosiding Seminar Nasional Pakar [Preprint]*. Available at: <https://doi.org/10.25105/pakar.v0i0.6853>.

- Alizah, M.D. et al. (2020) 'Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter', *Indonesian Journal on Software Engineering (IJSE)*, 6(2), pp. 223–229. Available at: <https://doi.org/10.31294/ijse.v6i2.8991>.
- Amzah, M.Y. and Bayuaji, L. (2024) 'Optimasi Algoritma Support Vector Machine Dengan Menggunakan Feature Selection Gain Ratio Untuk Analisis Sentimen', *JURNAL INOVTEK POLBENG*, 9(1), pp. 326–340.
- Apaydin Cirik, V., Gül, U. and Aksoy, B. (2022) 'The image of nursing among nursing and other healthcare professional university students: A mixed-method study', *Nurse Education in Practice*, 59, p. 103293. Available at: <https://doi.org/https://doi.org/10.1016/j.nepr.2022.103293>.
- Ari Ahmad Dhani, Taghfirul Azhima Yoga Siswa, W.J.P. (2024) 'Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting Improving the Accuracy of Random Forest with ANOVA and SMOTE on Stunting Data Classification', *TEKNIKA*, 13(2), pp. 264–272. Available at: <https://doi.org/10.34148/teknika.v13i2.875>.
- Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems (the original research of mathematics)*, 2(1).
- Arrank Tonapa, W., D.K. Manembu, P. and D. Kambey, F. (2024) 'Klasifikasi Ikan Cakalang dan Tongkol Menggunakan Convolutional Neural Network', *Jurnal Teknik Informatika*, 19(01), pp. 31–36. Available at: <https://doi.org/10.35793/jti.v19i01.52013>.
- Arya Darmawan, M.B., Dewanta, F. and Astuti, S. (2023) 'Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot', *TELKA - Telekomunikasi Elektronika*

Komputasi dan Kontrol, 9(1), pp. 52–61. Available at:
<https://doi.org/10.15575/telka.v9n1.52-61>.

Asrianto, R. and Herwinanda, M. (2022) 'Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine', *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), pp. 431–440. Available at:
<https://doi.org/10.37859/coscitech.v3i3.4368>.

Asrirawan, A., Permata, S.U. and Fauzan, M.I. (2022) 'Pendekatan Univariate Time Series Modelling untuk Prediksi Kuartalan Pertumbuhan Ekonomi Indonesia Pasca Vaksinasi COVID-19', *Jambura Journal of Mathematics*, 4(1), pp. 86–103. Available at:
<https://doi.org/10.34312/jjom.v4i1.11717>.

Assiddiq, D.U., Bisri, M.H., Handayani, S., Imron, M., Ambarwati, D., 2023. Komunikasi Pembangunan dan Mekanisme Bentuk Rancangan Peraturan Desa Mengenai Tata Kelola Aset Desa. *J. Nomosleca* 9, 158–178.
<https://doi.org/10.26905/nomosleca.v9i1.9910>

Baidi, B., 2022. Penggunaan metode inkuiri untuk meningkatkan hasil belajar Ips di kelas IV. *JPGI J. Penelit. Guru Indones.* 7, 11. <https://doi.org/1668757437>

BALAMURUGAN BALUSAMY, N. A. R., SEIFEDINE KADRY, AMIR H. GANDOMI 2021. *Big Data: Concepts, Technology, and Architecture*, USA, Wiley.

Barbrook-Johnson, P., & Carrick, J. (2022). Combining complexity-framed research methods for social research. *International Journal of Social Research Methodology*, 25(6). <https://doi.org/10.1080/13645579.2021.1946118>

Bauer, A.G. et al. (2022) 'Predictors of therapeutic alliance, treatment feedback, and clinical outcomes among African American women in treatment for co-occurring PTSD and

SUD', Journal of Substance Abuse Treatment, 139, p. 108766. Available at: <https://doi.org/https://doi.org/10.1016/j.jsat.2022.108766>.

BEGG, T. C. A. C. 2015. Database Systems: A Practical Approach To Design, Implementation, and Management, USA, Pearson Education.

Botter, C. H. (1982). Project management: A systems approach to planning, scheduling and controlling. European Journal of Operational Research, 10(2). [https://doi.org/10.1016/0377-2217\(82\)90164-3](https://doi.org/10.1016/0377-2217(82)90164-3)

Bustos, K. et al. (2022) 'Personality styles and adherence to treatment in adult patients with Multiple Sclerosis', Multiple Sclerosis and Related Disorders, 60, p. 103699. Available at: <https://doi.org/https://doi.org/10.1016/j.msard.2022.103699>.

Chatfield, C. (2018). Introduction to multivariate analysis. Routledge, New York.

Creswell, W. John & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. In Sage (Vol. 53, Issue 9).

Cronk, B.C., 2020. How To Use SPSS. A-Step-by-Step Guide to Analysis and Interpretation, Eleventh Edition. ed. Taylor & Prancis, London.

Dietterich, T. G. (1990). Machine learning. Annual review of computer science, 4(1), 255-306.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? (pp. 3-11). Springer International Publishing.

Elise, E. (2019) 'Analisis Pengaruh Keterlibatan Keluarga dengan Pengendalian Internal sebagai Variabel Moderasi

terhadap Biaya Agensi', null [Preprint]. Available at: <https://doi.org/null>.

- EXCELSIOR. 2022. Big Data, Explained: The 5V s of Data [Online]. Available: https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1 [Accessed 17 July 2024].
- Fauzi, A., & Setiawan, H. (2020). Etnomatematika: Konsep geometri pada kerajinan tradisional sasak dalam pembelajaran matematika di sekolah dasar. *Didaktis: Jurnal Pendidikan Dan Ilmu Pengetahuan*, 20(2).
- Field, A., 2018. *Discovering Statistics Using IBM SPSS Statistics, Fifth Edition*. ed. SAGE Publications, Inc., Universitas of Sussex.
- Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100.
- Grotenhuis, M. te, Matthijssen, A., 2016. *Basic SPSS Tutorial*. SAGE Publications, Inc, London.
- Handayani, F. et al. (2021) 'Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung', *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 7(3), p. Vol. 7 No. 3.
- Handayani, L. T. (2023). *Buku Ajar Implementasi Teknik Analisis Data Kuantitatif (Penelitian Kesehatan)*. PT. Scifintech Andrew Wijaya.
- Hastono Sutanto Priyo. (2007), *Analisis Data Kesehatan*. Depok: Fakultas Kesehatan Masyarakat, Universitas Indonesia.
- Heriyanto, B. (2022). *Metode Penelitian Kuantitatif (Teori dan Aplikasi)*. In Book.

- Ho, R., 2006. Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Taylor & Prancis, London.
- Ihzaniah, L.S., Setiawan, A. and Wijaya, R.W.N. (2023) 'Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing', *Jambura Journal of Probability and Statistics*, 4(1), pp. 17–29. Available at: <https://doi.org/10.34312/jjps.v4i1.18948>.
- Indriani, S. et al. (2023) 'Pengaruh Penyuluhan Melalui Media Audio Visual Terhadap Tingkat Pengetahuan Bahaya Seks Bebas Pada Remaja Di SMAN 1 Sukomoro Tahun 2023', *Jurnal Mahasiswa Kesehatan [Preprint]*. Available at: <https://doi.org/10.30737/jumakes.v5i1.5187>.
- Ismayani, A. (2019). Metodologi penelitian. Syiah Kuala University Press.
- Jailani, M. S. (2023). Teknik pengumpulan data dan instrumen penelitian ilmiah pendidikan pada pendekatan kualitatif dan kuantitatif. *IHSAN: Jurnal Pendidikan Islam*, 1(2), 1-9.
- Kambuaya, A.E., Masinambow, V.A.J. and Sumual, J.I. (2019) 'Analisis Variabel-Variabel (Faktor-Faktor) Yang Mempengaruhi Permintaan Jasa Angkutan Kota Di Kecamatan Malalayang Kota Manado', null [Preprint]. Available at: <https://doi.org/null>.
- Karim, B., Sentinuwo, S. and Sambul, A. (2017) 'Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan Data Mining', *Jurnal Teknik Informatika*, 11(1). Available at: <https://doi.org/10.35793/jti.11.1.2017.16555>.

- Kartowagiran, B. (2018). Optimalisasi uji tingkat kompetensi di SMK untuk meningkatkan soft skill lulusan. *Jurnal Dinamika Vokasional Teknik Mesin*, 3(2), 101-109.
- Kelley, K., & Bolin, J. H. (2013). Multiple regression. In *Handbook of quantitative methods for educational research* (pp. 69-101). Brill.
- Khoirunnisa, V. and Lestari, S. (2023) 'Implementasi Klasifikasi Kehamilan Beresiko Dengan Metode Naive Bayes Pada Puskesmas Kelurahan Malaka Jaya', *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4(3), pp. 1680–1693. Available at: <https://doi.org/10.35870/jimik.v4i3.396>.
- Khumaedi, M. (2012). Reliabilitas instrumen penelitian pendidikan. *Jurnal Pendidikan Teknik Mesin*, 12(1).
- Kline, P. (2014). *An easy guide to factor analysis*. Routledge, New York.
- Knight, R. et al. (2022) 'Dietary inflammation score is associated with perceived stress, depression, and cardiometabolic health risk factors among a young adult cohort of women', *Clinical Nutrition ESPEN*, 51, pp. 470–477. Available at: <https://doi.org/https://doi.org/10.1016/j.clnesp.2022.06.01>.
- Kriyantono, R. (2020). *Teknik praktis Riset Komunikasi kuantitatif dan kualitatif*. Jakarta: Prenadamedia Group.
- KUNAL, G. 2023. The Power of Big Data Analytics in Industrial IoT [Online]. Softobotics. Available: <https://www.softobotics.com/blogs/the-power-of-big-data-analytics-in-industrial-iot/> [Accessed 17 July 2024].
- Kuntoro, H., 2021. *Teori dan Aplikasi Analisis Multivariat Lanjut*. Zifatama Jawaara.
- Kusuma, M.D.H. and Hidayat, S. (2024) 'Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI', *Jurnal*

Indonesia : Manajemen Informatika dan Komunikasi, 5(2),
 pp. 1097–1110. Available at:
<https://doi.org/10.35870/jimik.v5i2.629>.

Kusumastuti, A. and Khoiron, A.M., 2019. Metode penelitian kualitatif. Lembaga Pendidikan Sukarno Pressindo (LPSP).

Lasmita, Y., Misnaniarti, M. and Idris, H., 2021. Analisis penerimaan vaksinasi Covid-19 di Kalangan masyarakat. Jurnal Kesmas (Kesehatan Masyarakat) Khatulistiwa, 8(4), p.195.

Lestari, A.D., Suberi, Moch., Anom, L., 2022. Pemanfaatan Limbah Bawang Merah Untuk Pembuatan Media Tanam Di Desa Jamberejo Kecamatan Kedungadem Kabupaten Bojonegoro. J. Manaj. Dan Penelit. Akunt. 15, 79–84. <https://doi.org/10.58431/jumpa.v15i2.203>

Lusiana, E.D. and Mahmudi, M., 2020. Teori dan Praktik Analisis Data Univariat dengan PAST. Universitas Brawijaya Press.

Makbul, M. (2021). Metode pengumpulan data dan instrumen penelitian.

Martono. 2015. Metode Penelitian Sosial. Jakarta. PT Raja Grafindo Persada.

Maruli Tua Silaen (2023) 'Klasifikasi Karakteristik Kepribadian Siswa Berdasarkan the Big Five Personality Dengan Menggunakan Metode K- Nearest Neighbor (Knn)', Jurnal Informatika dan Rekayasa Elektronik, 6(1), pp. 121–129. Available at: <https://doi.org/10.36595/jire.v6i1.860>.

MAYER-SCHÖNBERGER, V. & CUKIER, K. 2013. Big Data: A Revolution that Will Transform how We Live, Work, and Think, Houghton Mifflin Harcourt.

- MINELLI, M., CHAMBERS, M. & DHIRAJ, A. 2013. *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, Wiley.
- Mueller, R. O., & Hancock, G. R. (2018). Structural equation modeling. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 445-456). Routledge, New York.
- Mulyani, S. and Mulyani, S.H. (2021) 'Multimetode Sebagai Pendorong Motivasi Dan Hasil Belajar Matematika Kompetensi Dasar Persamaan Linier Dua Variabel', *Jurnal Lingkar Mutu Pendidikan* [Preprint]. Available at: <https://doi.org/10.54124/jlmp.v18i2.29>.
- Neil M. Richards, Jonathan H. King. 2014. *Big Data Ethics*. 393-431
- NG, A. 2018. *Machine Learning Yearning*, USA, MIT.
- Norfai, S.K.M., 2022. *Analisis data penelitian (Analisis Univariat, Bivariat dan Multivariat)*. Penerbit Qiara Media.
- Nurviana, N., 2020. A Survey on Smart Analytics: Method, Tools, and Open Research Issues. *J. Sist. Cerdas* 3, 54–64. <https://doi.org/10.37396/jsc.v3i1.54>
- Nuryono, M. et al. (2019) 'Pengaruh Kepemilikan Manajerial, Kepemilikan Institusional, Komisaris Independen, Komite Audit, Serta Kulit Audit Pada Nilai Perusahaan', *JURNAL ILMIAH EDUNOMIKA* [Preprint]. Available at: <https://doi.org/10.29040/jie.v3i01.457>.
- Pallant, J., 2020. *SPSS Survival Manual. A Step by Step Guide to Data Analysis Using IBM SPSS*, Seventh Edition. ed. Routledge, New York.
- Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice - Michael Quinn Patton - Google Books*. In Sage Publication.

- Pratama, E.E. (2021) 'Klasifikasi Kategori Permasalahan Akademik Mahasiswa Comdev Universitas Tanjungpura Menggunakan Algoritma Naive Bayes', *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 7(2), p. 106. Available at: <https://doi.org/10.26418/jp.v7i2.47199>.
- Pratyay Chattopadhyay dkk, 2023, ETHICS IN ARTIFICIAL INTELLIGENCE: PRINCIPLE AND CHALLENGES. *International Research Journal of Modernization in Engineering Technology and Science*. DOI:10.56726/IRJMETS44968
- Primadona, P. and Fauzi, R. (2023) 'Penerapan Data Mining Pada Penjualan Produk Elektronik', *Computer and Science Industrial Engineering (COMASIE)*, 9(4). Available at: <https://doi.org/10.33884/comasiejournal.v9i4.7712>.
- PROVOST, F. & FAWCETT, T. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly Media.
- Purnama, D.I. (2021) 'Model Regresi Hurdle Negative Binomial (HNB) untuk Pemodelan Konsumsi Rokok di Provinsi Sulawesi Tengah', *Jurnal Ilmiah Matematika Dan Terapan*, 18(1), pp. 21–31. Available at: <https://doi.org/10.22487/2540766x.2021.v18.i1.15506>.
- Putra, A.E. et al. (2020) 'Pengaruh Dewan Komisaris Independen, Komite Audit, Kepemilikan Manajerial, Kepemilikan Institusional dan Leverage terhadap Manajemen Laba Pada Perusahaan Industri Barang Konsumsi Yang Terdaftar di Bursa Efek Indonesia', null [Preprint]. Available at: <https://doi.org/10.29264/jiam.v4i4.5769>.
- Putu, I. et al. (2024) 'Suricata Accuracy Optimization Based on Live Analysis Using One-Class Support Vector Machine Method and Streamlit Framework', *Jurnal Teknik*

Informatika (JUTIF), 5(2), pp. 415–427. Available at:
<https://doi.org/10.52436/1.jutif.2024.5.2.1822>.

Rahmawati, I., Umkabu, T. and Lestari, N.P. (2021) 'Analisis Kesalahan Peserta Didik Dalam Penyelesaian Soal Sistem Persamaan Linear Dua Variabel Berdasarkan Kriteria Watson Di SMP Negeri 3 Kaureh Kabupaten Jayapura', *Jurnal Inovasi Pembelajaran Matematika (JIPM)* [Preprint]. Available at:
<https://doi.org/10.36379/jipm.v2i2.161>.

Ramadhan, M., 2021. Metode penelitian. Cipta Media Nusantara.

Ramírez, A.S., Wilson, M.D. and Soederberg Miller, L.M. (2022) 'Segmented assimilation as a mechanism to explain the dietary acculturation paradox', *Appetite*, 169, p. 105820. Available at:
<https://doi.org/https://doi.org/10.1016/j.appet.2021.105820>

Retnawati, H., Kartowagiran, B., Hadi, S., & Hidayati, K. (2011). Identifikasi kesulitan peserta didik dalam belajar matematika dan sains di sekolah dasar. *Jurnal Kependidikan Penelitian Inovasi Pembelajaran*, 41(2).

Roflin, E., & Riana, F. (2022). Analisis Korelasi dan Regresi. Penerbit NEM.

Romaní-Romaní, F., Gutiérrez, C. and Azurin-Salazar, J. (2023) 'Tendencia en la retención de conocimientos de ciencias básicas en una prueba de progreso entre estudiantes de Medicina', *Educación Médica*, 24(4), p. 100830. Available at:
<https://doi.org/https://doi.org/10.1016/j.edumed.2023.100830>.

Rosid, M.A., Gunawan, G. and Pramana, E. (2015) 'Centroid Based Classifier With TF – IDF – ICF for Classification of Student's Complaint at Appliation E-Complaint in

Muhammadiyah University of Sidoarjo', *JEEE-U (Journal of Electrical and Electronic Engineering-UMSIDA)*, 1(1), pp. 17–24. Available at: <https://doi.org/10.21070/jeee-u.v1i1.23>.

Rosmawati, - (2022) 'Manajemen laba terhadap kinerja keuangan dengan pengungkapan CSR sebagai variabel moderasi', *Accounting and Management Journal* [Preprint]. Available at: <https://doi.org/10.33086/amj.v6i2.3199>.

Sahir, S. H. (2021). *Metodologi penelitian*. Penerbit KBM Indonesia.

Saidin , M.Syahrani Jailani. 2023. Memahami Etika Dalam Penelitian Ilmiah. *Qosim Jurnal Pendidikan, Sosial dan Humaniora*, Volume 1 No 1, 24-29

Saifuddin, A. (2020). *Penyusunan skala psikologi*. Prenada Media.

Saragih, D.W.S.B. (2019) 'Pengaruh Arus Kas Bebas, Leverage Dan Peluang Pertumbuhan Terhadap Kebijakan Dividendengan Likuiditasdan Ukuran Perusahaan Sebagai Variabel Kontrol Pada Perusahaan Manufaktur Di BEI', null [Preprint]. Available at: <https://doi.org/null>.

Sari, I.N., Lestari, L.P., Kusuma, D.W., Mafulah, S., Brata, D.P.N., Iffah, J.D.N., Widiatsih, A., Utomo, E.S., Maghfur, I., Sofiyana, M.S. and Sulistiana, D., 2022. *Metode penelitian kualitatif*. Unisma Press.

Savitri, L. and Nursalim, R. (2023) 'Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning', *Diophantine Journal of Mathematics and Its Applications*, 2(01), pp. 30–36. Available at: <https://doi.org/10.33369/diophantine.v2i01.28260>.

Semiawan, C.R., 2010. *Metode penelitian kualitatif*. Grasindo.

- Seng Hansen, dkk.2023. ETIKA PENELITIAN, TEORI DAN PRAKTIK. Podomoro University Press (PU PRESS). Jakarta Barat
- Sihombing, S.O., 2022. Pengantar Metode Analisis Multivariat. Penerbit NEM.
- Silverman, D. (2017). David Silverman discusses qualitative research. In SAGE Publications (Vol. 9781473992).
- Slamet Widodo dkk. 2023., Buku Ajar Metode Penelitian. CV Science Techno Direct
- Sugiyono, D. (2013). Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D. Universitas Gresik.
- Suntama Putra, Risnita , M. Syahrani Jailani, Faisal Hakim Nasution, 2023. Penerapan Prinsip Dasar Etika Penelitian Ilmiah. Jurnal Pendidikan Tambusai, Volume 7 Nomor 3, 27876-27881.
- Suryanto, D. 2005. Etika Penelitian, 25(1), pp. 17–22. doi:10.30883/jba.v25i1.906.
- Sutaji, D., Faticah, C. and Navastara, D.A. (2016) ‘Segmentasi pembuluh darah retina pada citra fundus menggunakan gradient based adaptive thresholding dan region growing’, Register: Jurnal Ilmiah Teknologi Sistem Informasi, 2(2), pp. 105–116. Available at: <https://doi.org/10.26594/register.v2i2.553>.
- Tersiana, A., 2018. Metode penelitian. Anak Hebat Indonesia.
- Timm, N. H. (Ed.). (2002). Applied multivariate analysis. New York, NY: Springer New York.
- Umami, A. (2021). Konsep Dasar Biotatistik. Pelita Medika.
- Unaradjan, D.D., 2019. Metode penelitian kuantitatif. Penerbit Unika Atma Jaya Jakarta.

- Weigand, A.J. et al. (2022) 'Cerebral blood flow, tau imaging, and memory associations in cognitively unimpaired older adults', *Cerebral Circulation - Cognition and Behavior*, 3, p. 100153. Available at: <https://doi.org/https://doi.org/10.1016/j.cccb.2022.100153>.
- Wiyanto, W. and Setyaningsih, Z. (2021) 'Sentiment Analysis Pemutusan Hubungan Kerja Akibat Pandemi Covid-19 Menggunakan Algoritma NaïveBayes Dan PSO', *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 10(3), pp. 426–431. Available at: <https://doi.org/10.32736/sisfokom.v10i3.1299>.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.
- Yoga Nanda Khoiril Umat, Dhiaz Rusyda Nafsyi, Diana Kusumaningsih, L.H. (2024) 'Analisis faktor yang mempengaruhi pemilihan gubernur daerah khusus jakarta menggunakan algoritma naive bayes dan regresi logistik 1)', *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, 9(2), pp. 211–224.

BIODATA PENULIS

Aisyah Mutia Dawis, S.Kom., M.Kom. lahir di Kota Surakarta pada tanggal 3 Juni 1991. Ia Lulus S1 pada tahun 2013 hingga mendapat gelar Sarjana Komputer di Universitas Muhammadiyah Surakarta serta mendapatkan penghargaan sebagai lulusan terbaik Se Fakultas Komunikasi dan Informatika, selanjutnya ia lulus S2 pada tahun 2019 hingga mendapat gelar Magister Komputer di Universitas Amikom Yogyakarta. Saat ini ia tercatat

sebagai dosen tetap di Program Studi Sarjana Sistem dan Teknologi Informasi di Universitas 'Aisyiyah Surakarta. Selain mengajar ia aktif dalam kegiatan tridharma lainnya diantaranya ialah penelitian dan pengabdian kepada masyarakat. Kegiatan penelitian internal dan eksternal pernah dilakukannya. Beberapa penelitiannya berjudul : "*Utilization of Virtual Reality Technology in Knowing the Symptoms of Acrophobia and Nyctophobia*". Ia sering menjadi *invited speaker* diantaranya menjadi tenaga pengajar BPPTIK Kementerian Kominfo. Serta sebagai bentuk pengabdian kepada masyarakat, ia pun pernah terlibat aktif sebagai trainer bimbingan teknis bantuan TIK SMP, SMA, Mahasiswa, Tenaga Pendidik, Guru dan Dosen dari tahun 2020 hingga saat ini. Adapun karya buku yang telah ditulisnya sejak tahun 2023, diantaranya berjudul :

1. *Artificial Intelligence* : Konsep Dasar Dan Kajian Praktis
2. *Data Lake Insights*
3. *Data Cloud*
4. Pengantar *Data Science*
5. Rekamaya Perangkat Lunak : Panduan Praktis Untuk Pengembangan Aplikasi Berkualitas



Dr. Hj. Rulyanti Susi Wardhani,
S.E., M.Si., CSRS., CSRA, Lahir di Medan 4 Juni 1967. Saat ini Penulis. Saat ini Penulis adalah staf Pengajar di Fakultas Ekonomi Jurusan Akuntansi Universitas Bangka Belitung. Penulis Menempuh Pendidikan S1 Bidang Ilmu Akuntansi di Universitas IBA Palembang (1988-1993). Penulis S2 bidang Ilmu Akuntansi di UNPAD

Bandung (1999-2003). Kemudian S3 penulis tempuh di UNSRI Palembang dengan bidang Kajian Umum Akuntansi. Buku yang pernah ditulis antara lain:

1. Pengantar Akuntansi
2. Akuntansi Biaya
3. Tata Kelola Perguruan Tinggi
4. Aplikasi Keuangan Untuk Petani
5. Kampung KB dan Kesejahteraan Keluarga Miskin
6. Strategi dan pengembangan Wisata Budaya Tempilang
7. Komoditas Unggulan Perkebunan Karet. Kopi dan Kelapa Sawit: Analisis Lq, SWOT dan RCA
8. Akuntansi Manajemen
9. Mengenal Saham
10. Jalan Menuju Organisasi Berkelanjutan



Hartono Nurlette, S.Pi., M.Si lahir

di Buano Pada Tanggal 19 Maret 1987. Ia mendapatkan gelar Magister dari Program Studi Ilmu Kelautan Universitas Pattimura Ambon pada Tahun 2017. Saat ini Ia tercatat sebagai dosen tetap pada Program Studi Ilmu Kelautan Universitas Muhammadiyah Maluku dan ditugaskan menjadi sekretaris Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM). Pada kegiatan tridarma Ia

ditugaskan mengampuh mata kuliah analisis kuantitatif. Selain mengajar ia aktif dalam melakukan kegiatan penelitian dan pengabdian masyarakat. Adapun karya buku yang telah ditulis adalah “Metodologi Penelitian Kuantitatif”.

Ritha Widyaprawati



Lisa Astria Milasari, S.T., M.T. lahir

di Kota Samarinda. Penulis Lulus pada tahun 2014 mendapat gelar Magister Teknik Sipil Minat Kekhususan Perencanaan Wilayah dan Kota di Universitas Brawijaya. Saat ini Dosen tetap pada Program Studi Perencanaan Wilayah & Kota, Fakultas Teknik, Universitas 17 Agustus 1945 Samarinda. Menyelesaikan pendidikan S1 pada

Jurusan Teknik Planologi, ITN Malang dan melanjutkan S2 pada Magister Teknik Sipil Minat Kekhususan Perencanaan Wilayah dan Kota, Universitas Brawijaya. Penulis saat ini menekuni

bidang penjaminan dan audit mutu. Aktif melakukan penelitian pada bidang perencanaan wilayah dan kota, yang salah satunya berhasil didanai oleh Kemdikbudristekdikti pada tahun 2022. Adapun karya buku yang telah ditulisnya sejak tahun 2020, diantaranya berjudul Ilmu Kealamiah Dasar, Statistik Ekonomi, Kebijakan Sistem Penjaminan Mutu Internal Universitas 17 Agustus 1945 Samarinda Tahun 2022-2026, Buku Saku Pedoman Auditor SPMP-AMI, Monograf Efektivitas Fungsi Taman Kota dan buku referensi lainnya.



Mohamad Zaenudin, S.Pd., M.Sc.Eng.

telah menjadi salah satu Dosen di Universitas Global Jakarta dengan status *Senior Lecturer*. Beliau menyelesaikan pendidikan S1 di Universitas Negeri Jakarta pada program studi S1 Pendidikan Teknik Elektronika dan melanjutkan studi S2 di Management and Science University dengan fokus pada bidang ilmu material. Selama perjalanan akademisnya, Mohamad Zaenudin telah meneliti dan menyelesaikan tesisnya mengenai simulasi dinamika molekuler pada fenomena welding difusi material Aluminium dan Nikel. Pengalaman riset ini memberikan kontribusi besar dalam pemahaman tentang material dan teknik pengolahannya, yang menjadi fokus utama dalam pengajaran beliau di Universitas Global Jakarta. Sebagai seorang dosen, Mohamad Zaenudin telah mengajar berbagai mata kuliah, termasuk Matematika Teknik, Fisika, dan sekarang fokus pada mata kuliah keteknikan mesin, terutama pada bidang material. Selain aktif dalam pengajaran, beliau juga aktif dalam riset ilmiah. Beliau telah menerbitkan berbagai artikel ilmiah di bidang teknik mesin dan material pada jurnal internasional terindeks bereputasi dan jurnal nasional terindeks serta menjadi *reviewer* pada jurnal-jurnal tersebut. Prestasinya dalam riset telah

diakui dengan berbagai hibah dari DIKTI, DRTPM dan BELMAWA. Selain itu, Mohamad Zaenudin juga turut berkontribusi dalam pengembangan pendidikan tinggi di Indonesia sebagai seorang *reviewer* pada Program *Matching Fund* tahun 2023 dan Program Dana Padanan tahun 2024, Dewan Pendidikan Tinggi, Dirjen Pendidikan Tinggi, Kemendikbud-Ristek.

Kontak: mzaenudin@jgu.ac.id

Profil Google Scholar: <https://bit.ly/GS-MZaenudin>

Profil *Researchgate*: <https://bit.ly/RG-MZaenudin>

Profil Scopus: <https://bit.ly/SC-MZaenudin>



Inayatul Inayah, S.Si., M.Si lahir di Kota Demak pada tanggal 21 Desember 1994. Ia Lulus pada tahun 2020 dan mendapat gelar Magister Fisika dengan konsentrasi Elektronika dan Komputasi di Institut Teknologi Bandung melalui beasiswa LPDP. Sejak tahun 2021, ia telah tercatat sebagai dosen tetap program studi Fisika Institut Teknologi dan Sains Nahdlatul Ulama (ITSNU) Pekalongan. Selain mengajar, ia juga

aktif dalam kegiatan tridharma lainnya, yaitu penelitian dan pengabdian. Ia pun pernah diamanahi sebagai ketua Bidang Audit Mutu Internal ITSNU Pekalongan, *Reviewer* Proposal Penelitian dan Pengabdian ITSNU Pekalongan, Auditor Mutu Internal ITSNU Pekalongan, dan Pembina UKM riset. Saat ini, ia sedang menempuh Pendidikan Doktor di Jurusan Fisika Institut Teknologi Bandung dan menjadi penerima beasiswa BPI-LPDP tahun 2023. Beberapa penelitian dan publikasi yang telah

dilakukan berfokus pada bidang elektronika dan instrumentasi, *internet of things*, dan *machine learning*. Pada tahun 2022, ia berhasil mendapatkan dana hibah penelitian DIKTI (PDP) dengan judul Rancang Bangun Sistem Pemantauan Panel Surya Secara *Realtime* Berbasis *Internet of Things*, dan pada tahun 2023-2024 ia juga berhasil mendapatkan dana hibah Penelitian Kompetitif Nasional (PT-JH) dengan judul Pengembangan Sistem *Smart Agriculture* untuk Peningkatan Produktivitas Hasil Pertanian Berbasis *Internet of Things* dan *Machine Learning*. Selain aktif melakukan kegiatan penelitian dan pengabdian kepada masyarakat, ia juga aktif sebagai penulis buku dengan judul “*Artificial intelligence : konsep dasar dan kajian praktis*” dan “*Algoritma Pemrograman*” pada tahun 2022 dan 2024. Ia juga aktif menjadi narasumber dalam kegiatan seminar maupun *workshop*.



Deny Haryadi, S.Kom., M.Kom

Penulis lahir di Bekasi pada tanggal 14 Juli 1991. Ia Lulus pendidikan Sarjana pada tahun 2017 di STT Pelita Bangsa pada Program Studi Teknik Informatika. Setelah itu pada tahun 2020 Lulus pendidikan Magister di Universitas Presiden pada Program Studi Teknik Informatika. Ia telah menulis dua buku kolaborasi berjudul “*data sebagai fondasi kecerdasan buatan*” dan “*kecerdasan buatan dan IOT*”, sedangkan ini merupakan buku ketiga. Saat ini ia tercatat sebagai dosen tetap pada Program Studi S1 Teknologi Informasi (Kampus Jakarta) di Universitas Telkom. Selain mengajar ia aktif dalam kegiatan tridarma lainnya diantaranya ialah penelitian dan pengabdian kepada masyarakat. Saat ini ia diamanahi sebagai Ketua Program Studi S1 Teknologi Informasi (Kampus

Jakarta) Universitas Telkom, managing Informatics and Communication Technology (JICT) dari Universitas Telkom, editorial board Journal of Practical Computer Science (JPCS) dari Universitas Pelita Bangsa, reviewer mitra bestari Sinkron: Jurnal dan Penelitian Teknik Informatika dari Politeknik Ganesha Medan, reviewer mitra bestari Journal of Informatics, Information System, Software Engineering and Applications (INISTA) dari Institut Teknologi Telkom Purwokerto, dan reviewer Indonesia Journal on Computing (Indo-JC) dari Universitas Telkom. Penelitian eksternal yang berhasil didanai oleh Ristekdikti dari tahun 2022: SIKUMELAWIT (Sistem Informasi Kualitas Minyak Kelapa Sawit) Aplikasi Deteksi Kualitas Kejernihan Minyak Kelapa Sawit menggunakan Pengolahan Citra Berbasis Android. Selain itu Penelitian Internal yang berhasil didanai: Implementasi Algoritma K-Medoids Clustering untuk Pengelompokan Ekspor Minyak Kelapa Sawit berdasarkan Negara Tujuan (2021), Prediksi Harga Bahan Bakar Nabati Jenis Biodiesel dengan menggunakan Algoritma Regresi Linear Berganda (2022), SIKUMELAWIT – IoT “Sistem Informasi Kualitas Minyak Kelapa Sawit Berbasis Internet of Things” (2023), Analisa Deteksi Logam Berat Timbal (PB) Berbasis Sensor Fiber Optik pada Air Minum (2023), Deteksi Kunci Gitar Secara Real-Time Menggunakan Metode Berbasis Deep Convolutional Neural Network (2023), dan Deteksi Logo Kendaraan Menggunakan Metode Detection Transformer (2023). Deny memiliki ketertarikan lebih pada bidang ilmu komputer, kecerdasan buatan, data mining, machine learning, deep learning, computer vision, dan text mining. Anda dapat menghubunginya melalui email denyharyadi@telkomuniversity.ac.id atau denyharyadi14@gmail.com



Yoana Nurul Asri, S.Si., M.Pd lahir di

Kota Bandung pada tanggal 17 November. Ia Lulus pada tahun 2016 hingga mendapat gelar Magister Pendidikan Fisika di Universitas Pendidikan Indonesia. Saat ini ia tercatat sebagai dosen tetap untuk mata kuliah Fisika di Universitas Nurtanio Bandung. Selain mengajar ia aktif dalam kegiatan tridarma lainnya diantaranya ialah

penelitian dan pengabdian. Saat ini ia pun diamanahi sebagai Sekertaris Jenderal Forum Komunikasi Dosen, Ketua Biro Research, Publikasi Karya Ilmiah dan Jurnal di Perkumpulan Dosen Peneliti Indonesia Kabupaten Bandung Barat, Instruktur Visitasi AKMI Tahun 2022, Trainer Bimbingan Teknis Bantuan TIK (Chromebook) dari Direktort SMP tahun 2021, menjadi editor beberapa judul buku, koordinator buku ajar dan buku referensi editorial board jurnal Abdi Wiralodra dari Universitas Wiralodra Indramayu, dan menjadi reviewer Jurnal Pendidikan Fisika dan Sains (JPFS) dari Universitas Nahdlatul Ulama Cirebon. Beberapa penelitian yang berhasil didanai oleh Ristekdikti dan hibah internal dari tahun 2019 hingga sekarang berjudul : Pembuatan Indikator Kelayakan Oli Mesin Kendaraan, Pembuatan Set Eksperiment Untuk Mengukur Percepatan Gravitasi Berbasis Arduino (2019), Pengembangan Kemampuan STEM Dengan Penggunaan Jaringan *Online Technology* Pada Pembelajaran Konsep dan *Hands On* (2021), Strategi Meningkatkan *Financial Literacy* Berorientasi Pada Sistem *Financial Technology* di Era *Cashless Society* (2022), dan Monitoring Kualitas Air Dan Efisiensi Energi Listrik Pada Sistem *Vertical Aquaponic* Berbasis Mikrokontroler (2022), Studi Komparatif Penggunaan Platform Zoom Dan Google Classroom Sebagai Media Perkuliahan Daring Di Universitas Nurtanio Bandung (2022), Pendampingan Umkm Dalam Pembuatan Desain Packaging Inovatif Kentang Mustofa Dan Optimalisasi Penjualan

Untuk Meningkatkan Daya Jual Produk (2023), Deteksi Jumlah Konsumsi Air Minum Harian Dengan Smart Dispenser Berbasis Iot (2023), Formulasi Sediaan Obat Kumur (Mouthwash) Dengan Ekstrak Cengkeh Dan Peppermint (2023), Penerapan Green Human Resource Management Terhadap Green Employee Behaviour yang dimediasi Top Management Commitment Pada Perguruan Tinggi (2024)



Nurul Kholisatul 'Ulya, M.Eng lahir dan menetap di Kota Solo. Menyelesaikan pendidikan Magister Teknologi Informasi di Universitas Gadjah Mada. Kini Bekerja di salah satu Perguruan Tinggi di kota Solo. Selain Mengajar ia juga aktif di kegiatan penelitian dan pengabdian. Saat ini ia mendapat amanah menjadi ketua program studi Informatika Program Sarjana di Institut Teknologi Sains dan Kesehatan PKU

Muhammadiyah Surakarta. Adapun karya buku yang telah ditulisnya sejak tahun 2020, diantaranya berjudul Modul Algoritma dan Pemrograman (2020), Modul Struktur Data (2021), Modul Sistem Operasi (2022) dan Modul Pemrograman Jaringan (2022), Sistem Basis Data (2024). Pengembangan Sistem Berbasis Data (2024)



E-BOOK
TOHAR MEDIA

PANDUAN PRAKTIS ANALISIS VARIABEL UNTUK PENELITI

Buku ini menawarkan panduan lengkap untuk peneliti dalam memahami dan menerapkan teknik analisis variabel. Dimulai dengan penjelasan mendasar tentang variabel dan metode pengumpulan data, buku ini melanjutkan dengan pembahasan tentang penggunaan software statistika untuk analisis data, serta teknik analisis univariant, bivariant, dan multivariant. Pembaca akan mendapatkan pemahaman praktis mengenai cara mengolah data menggunakan teknik seperti regresi berganda dan analisis faktor, serta penerapan big data dan analitik lanjutan.

Selain itu, buku ini juga mencakup penerapan machine learning dan data mining untuk analisis variabel, dan memberikan panduan dalam pengembangan instrumen penelitian yang efektif. Aspek etika dalam penelitian juga dibahas, menekankan pentingnya integritas dan perlindungan peserta penelitian. Dengan pendekatan praktis dan terstruktur, buku ini menjadi sumber berharga untuk peneliti yang ingin meningkatkan keterampilan mereka dalam analisis data secara efektif dan etis.



TOHAR MEDIA

No Anggota IKAPI : 022/SSL/2019

Workshop : JL. Adiyaksa Baru, Ruko Zamrud Blok I no 9

Redaksi : JL. Muhktar dg Tompo Kabupaten Gowa
Perumahan Nayla Regency Blok D No 25
Telp. (0411) 8987659 Hp. 085299993635
<https://toharmedia.co.id>

ISBN 978-623-8705-31-3



9 786238 705313