|  |  |
| --- | --- |
|  |  |
|  | **MODUL 3 PERKULIAHAN** |
|  |  |
|  | **Proyek Data Science** |
|  |  |
|  | **Pengertian Data Science dan contoh pemanfaatannya** |
|  |  |

# Pembahasan

**Pertemuan 3**

**Bahasa pemrograman untuk data science dan Tools pendukung data science**

1. Python

Bahasa syntaxnya yang mudah merupakan salah satu keunggulan Python dibandingkan yang dengan bahasa pemrograman lainnya. Uniknya, Python juga tepat untuk kamu yang belum pernah mempelajari coding.

1. Sederhana

Sederhananya bahasa pemrograman Python sudah tidak diragukan lagi. Untuk kamu yang baru mengenal coding pun mampu mengoperasikan Python dengan mudah. Jika kamu ingin mendalami ilmu coding, maka Python merupakan bahasa pemrograman yang tepat karena bisa dilakukan dengan efektif dan cepat.

2. Library : Banyak dan Lengkap

Banyaknya library yang tersedia pada Python memberikan kemudahan untuk kamu menganalisis, memvisualisasikan dan memformulasikan data-data mentah menjadi informasi yang berharga. Oleh karena itu, library yang ada pada Python merupakan sumber dari berbagai fungsionalitas.

3. Mudah Diakses Statisican dan Programmer

Dengan Python seorang yang ingin membuat project atau product, tepat untuk menggunakan bahasa pemrograman ini. Seseorang dengan latar belakang statistik dan programming sering kali memilih bahasa pemrograman Python untuk menjadi salah satu tools dalam pengolahan data mereka.

4. Ramah IoT

Internet of things merupakan salah satu nilai plus dari Python. Bahasa pemrograman ini menarik perhatian dalam ruang coding IoT selama beberapa tahun terakhir dalam proyek pemrograman. Python pada IoT dibuat menggunakan Raspberry Pi dan platform Bluemix IBM untuk home Automation System.

1. R

Dalam dunia data terdapat banyak jenis bahasa pemrograman salah satunya yang cukup populer adalah R. Bahasa pemrograman R sendiri memiliki fungsi untuk mengembangkan bahasa baik dari data analisis, model grafik dan statistik.

Mengapa bahasa pemrograman ini memiliki nama R? Karena, bahasa pemrograman ini pertama kali dibuat oleh Ross Ihaka dan Robert Gentleman. Mereka merupakan statistikawan asal Selandia Baru. Namun jangan salah, R kini dikembangkan oleh kolaborasi dari pakar-pakar statistikawan dan komputer di seluruh dunia.

1. Bersifat Gratis

Bahasa R bersifat gratis dan open source yang artinya, tidak memerlukan biaya lisensi yang pada umumnya cukup mahal untuk perangkat lunak pengolahan data. Kebebasan untuk mengakses R memberikan peluang untuk kamu yang ingin memulai karir di bidang data dengan memulai belajar dasar pengolahan data.

Walaupun R merupakan bahasa pemrograman yang bersifat gratis, kemampuannya tidak kalah dengan bahasa pemrograman lain yang komersial, bahkan dari segi tertentu R merupakan bahasa pemrograman yang lebih baik.

2. Mampu Mengolah Data dalam Jumlah Besar

R menjadi salah satu solusi untuk kamu yang ingin mengolah data dengan skala besar yang tidak mampu diolah menggunakan excel. Untuk perhitungan yang lebih banyak dan kompleks, perintah-perintah R ditulis lebih dahulu dalam sebuah text editor, kemudian dijalankan untuk diolah.

3. Bahasa Pemrograman yang Powerfull

Selain R merupakan bahasa pemrograman yang bersifat gratis, R pun sangat bersifat powerful dimana memiliki banyak tools yang dapat digunakan untuk pengolahan data. Banyaknya jenis packages, test, dan tools yang mudah dipahami, R menjadi salah satu bahasa pemrograman yang tepat untuk kamu yang ingin mengolah data dengan hasil akhir berupa visual.

**3 Bahasa Pemrograman Untuk Menjadi Data Science**

Kemajuan teknologi dan sejumlah besar data online yang tersedia mempengaruhi setiap sektor dan memiliki dampak yang luar biasa pada ekonomi. Hal ini disebut “data avalanche” bukan hanya tentang volume data belaka, tapi juga kecepatan perubahan dan pertumbuhannya, dan beragam jenis data yang tersedia.

Hal ini membuat developer dengan kemampuan data science banyak di cara oleh perusahaan besar yang menawarkan gaji yang menggiurkan. Dan juga tingkat pertumbuhan untuk lowongan pekerjaan data science sangat tinggi selama empat tahun terakhir ini.

Beberapa institusi pendidikan di luar sana bahkan sudah membuat program sarjana untuk jurusan data science, diantaranya Northeastern University, Boston University, CUNY dan Merrimack College. Beberapa dari kampus ini juga membuka kursus online, dan juga ada program biaya rendah dan seminar yang tersedia melalui IEEE Computer Society.

Jika kamu berencana untuk mengejar karir pada bidang data science, maka kamu harus mempertimbangkan untuk mempelajari tiga bahasa pemrograman ini.

1. R

R merupakan bahasa pemrograman dan juga framework yang digunakan untuk data miners untuk mengembangkan software statistik atau analisis data.

Bahasa pemrograman ini terlihat memiliki lonjakan besar karena data analysis dan data science menjadi lebih umum dalam beberapa tahun terakhir. R memiliki tool yang dibangun bagi data scientist, dengan ekstensi dan plugin yang khusus dibuat untuk tujuan tersebut.

Ketika belajar bahasa R, individu juga harus mengerti dasar-dasar dari ilmu matematika. Akan menjadi bencana jika kita hanya percaya kepada output dari software tanpa mengetahui apa yang diukur dan tanpa mengerti data yang disediakan sebagai input.

2. Python

Python merupakan bahasa pemrograman untuk tujuan umum, yang memiliki tool yang sesuai dengan bermacam-macam lingkungan seperti website atau mobile. Bahasa ini juga lebih mudah dibaca ketimbang R.

Jika ditanya bahasa pemrograman apa yang harus dipelajari untuk menjadi data scientist atau ilmuwan data, maka bahasa R dan Python harus kamu pelajari.

3. Java

Java merupakan salah satu bahasa pemrograman terfavorit, menurut survei yang dilakukan oleh WP Engine. Sama seperti Python, Java merupakan bahasa pemrograman untuk tujuan umum yang mendukung berbagai platform.

Java juga merupakan bahasa interpreted, namun tidak seperti C dan C++, Java tidak membutuhkan pengetahuan yang banyak mengenai hardware, menjadikannya lebih mudah untuk dipelajari. Java juga menjadi bahasa pemrograman yang paling banyak dicari di iklan lowongan pekerjaan.

9 Data analytics tools terbaik sesuai kebutuhanmu

1. Tableau

Sebagai tombak terdepan dalam analisis data, Tableau merupakan software dengan kemampuan Big Data yang tak perlu diragukan.

Seperti yang diketahui fungsi big data adalah untuk mengumpulkan dan mengolah informasi pelanggan sehingga pengusaha dapat mengetahui kebutuhan konsumen dan menjadi lebih dekat dengan mereka.

Sebagai tools untuk big data, Tableau dapat menghubungkan data dari beragam sumber. Mulai dari data berbasis Microsoft Excel, jaringan internet, maupun Data

Hasil pengolahan data Tableau juga bermacam-macam, bisa berupa data visualisasi dalam bentuk bagan serta dashboard. Terlebih lagi, hasil tersebut dapat diperbarui (update) secara real time.

Bila dibutuhkan, Tableau pun menyediakan fitur unduh dengan beragam format berkas sesuai keperluan. Kamu juga bisa membagikan solusi analisis data di sosial media maupun pada klien apabila diperlukan.

2. Domo

Domo merupakan ruang manajemen bisnis yang berbasis cloud. Software ini mampu menggabungkan sumber data seperti database, spreadsheet, dan tentu saja sofware berbasis cloud.

Kecepatan dan pemahaman data merupakan fokus dari Domo. Kemampuan business intelligence dari software ini memungkinkan kolaborasi melalui visualisasi dan fitur pemberian komentarnya.

Perangkat lunak ini juga mendukung penggunaan melalui gawai mobile. Berkat itu, pengguna bisa mengaksesnya dengan mudah dan di mana saja.

3. Looker

Looker termasuk tool yang lebih cocok bagi pengguna yang telah berpengalaman dalam membaca atau memakai data. Dengan proses extract, load, transform (ELT) yang tersedia, pemakai dapat membentuk dan mengubah data sesuai kebutuhan.

Salah satu data analytic tools ini bisa menyampaikan data langsung ke tool dan aplikasi yang kamu gunakan. Looker juga memungkinkan pengguna untuk memberikan pertanyaan rumit, selama memakai bahasa bisnis yang umum.

4. Sisense

Sisense mampu menyederhanakan analisis data. Oleh karena itu, hasil olahan datanya tidak hanya dapat dipahami oleh tim analisis, tapi juga pengguna yang lebih awam.

Tool ini menawarkan beragam fitur. Mulai dari pengaturan tampilan dashboard sesuai kenyamanan dan kebutuhan, visualisasi yang interaktif, hingga kemampuan analisis yang mantap.

5. Knime

Knime atau Konstanz Information Miner dapat menggabungkan berbagai komponen untuk menggali informasi (data mining) dan machine learning. Tool ini menerapkan open source akan membantumu dalam menganalisis dan menampilkan data melalui visual programming.

data analytics tools - EKRUT

Skytree adalah contoh tools analisa big data lainnya yang bisa kamu gunakan - EKRUT

6. Skytree

Skytree merupakan salah satu tools untuk big data lainnya yang biasa digunakan untuk membangun model yang lebih akurat dan cepat.

Ada beberapa fitur dalam tools ini seperti algoritma yang dapat diukur, Artificial Intelligence, model interpretability, serta akses GUI dan programatik.

7. Pentaho

Pentaho merupakan data analytics tools yang menekankan pada pengumpulan data IoT dan pencampurannya dengan sumber data lain seperti sistem ERP dan CRM serta tools untuk big data seperti Hadoop dan NoSQL.

Integrasi internal dengan IoT end points dan fungsi injeksi metadata yang unik mempercepat pengumpulan data dari berbagai sumber.

8. Xplenty

Xplenty adalah solusi ETL berbasis cloud yang menyediakan jalur pipa data sederhana yang kemudian divisualisasikan untuk aliran data di berbagai sumber dan tujuan secara otomatis.

Beberapa keunggulan yang ada pada tool ini seperti fokus pada keamanan, serta fleksibilitas tujuan dalam hal mengirim data ke database, data warehouse dan salesforce.

9. RapidMiner

RapidMiner adalah open-source tool dengan banyak kelebihan yang dapat digunakan untuk ananlisa data secara mendalam pada berbagai jenis bisnis.

Tool ini mengotomatiskan sebagian besar proses analitik dan menyederhanakan prosedur operasional sehingga kamu hanya perlu fokus pada interpretasi data saja.

Beberapa kelebihan lain dari RapidMiner adalah memiliki 1,5 ribu algoritma asli, data tracking dan analytics yang real-time, menunjang transformasi dan data cleansing, mendukung integrasi pihak ketiga, kemampuan pelaporan yang komprehensif, skalabilitas untuk digunakan oleh tim dengan ukuran apa pun, dan adanya fitur keamanan yang superior.

Menentukan pilihan data analytics tools yang tepat terkadang tidak mudah dan butuh waktu yang tidak sedikit. Kamu juga harus menyesuaikannya dengan kebutuhan perusahaan dan sistem teknologi informasi yang selama ini digunakan.

3 Aplikasi Untuk Mulai Belajar Data Science

Inzightlite

Inzightlite merupakan aplikasi yang dibuat oleh peneliti dari Universitas Auckland, New Zealand. Aplikasi ini didesain agar memudahkan pelajar sekolah menengah dan para pemula untuk belajar data science. Kegunaan utamanya adalah menganalisis dan memvisualisasikan data sekaligus.

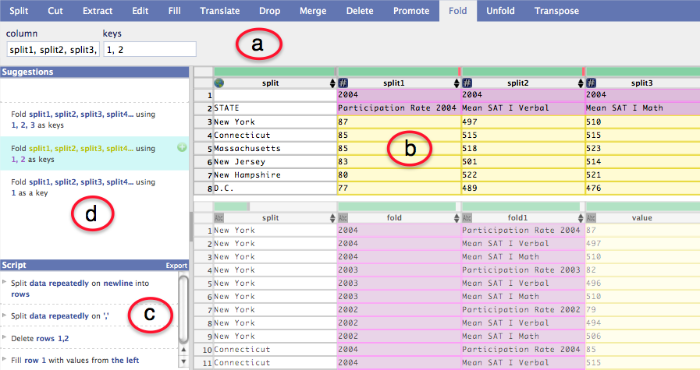
Cara menggunakannya cukup sederhana. User tinggal masuk ke website dan upload dataset ke website tersebut. Setelah itu, kita bisa membuat visualisasi box plot, scatter plot, hingga peta. Inzight juga memberi summary dan gambaran data seperti mean, median, dan modus dari data yang kita miliki.



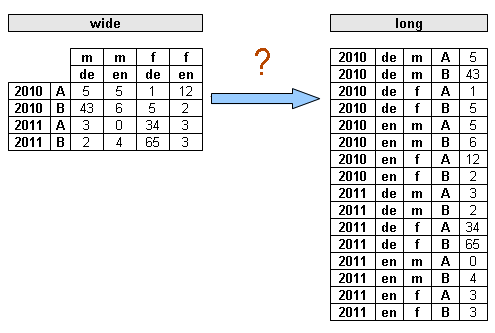
Gambar 1 Inzight lite

Data Wrangler

Data wrangler adalah salah satu aplikasi yang dibuat oleh Stanford University. Aplikasi ini memudahkan dalam manipulasi data yang akan kita analisis. Aplikasi ini juga tersedia secara gratis.

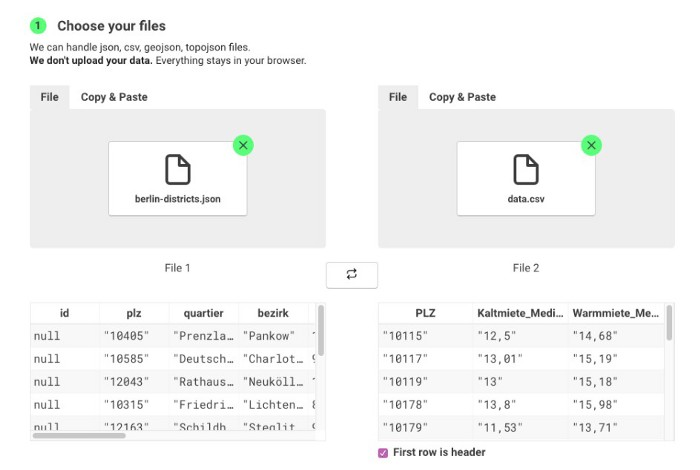


Kemudian, apa yang dimaksud dengan manipulasi data? Seringkali kita menemukan data dalam format lebar. Dalam format data lebar, kita tidak bisa melakukan analisis dengan bantuan excel ataupun tool analisis data lainnya. Sementara software analisis data membutuhkan data dalam format panjang.



Geo Datamerger

Aplikasi lainnya adalah Geo Datamerger. Apliksai ini didesain oleh Funke Interaktiv https://medium.com/@interaktiv. Kegunaan dari aplikasi ini adalah menggabungkan dua dataset berbeda menjadi satu. Caranya mudah, hanya dengan copy paste dataset yang dimiliki ke website.



Quiz Pertemuan 3

Buatlah tutorial atau panduan cara menginstall dan menggunakan tools PYTHON dan R pada word.

# Daftar Pustaka

[1] “Data Scientist: The Sexiest Job of The 21st Century” Diakses 01 April 2019 dari https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[2] “50 Best Jobs in America 2019” Diakses 01 April 2019 dari https://www.glassdoor.com/List/Best-Jobs-in-America-LST\_KQ0,20.htm

[3] “Tech in Asia Jobs” Diakses 02 Desember 2018 dari https://www.techinasia.com/jobs

[4] “Kalibrr: Where Jobs Find You” Diakses 02 Desember 2018 dari https://www.kalibrr.com/

[5] “Data Scientist Earning More Than CAs, Engineers” Diakses 01 April 2019 dari https://timesofindia.indiatimes.com/india/Data-scientists-earning-more-than-CAs-engineers/articleshow/52171064.cms

[6] “5 Facts About Software Engineers, Like Which One Gets Paid The Most” Diakses 01 Apil 2019 dari https://learning.linkedin.com/blog/tech-tips/the-american-city-that-pays-software-engineers-the-most–and-oth

[7] “Top 10 Reasons Why You Should Learn Data Analytics” Diakses 01 April 2019 dari https://bigdata-madesimple.com/10-reasons-why-you-should-learn-data-analytics/

[8] “5 Reasons Why Everybody Should Learn Data Analytics” Diakses 01 April 2019 dari https://www.sas.com/en\_au/insights/articles/analytics/5-reasons-why-everybody-should-learn-data-analytics.html

[9] “Permintaan Tenaga Data Scientist Melonjak, Jadikan Profesi Ini Kian Menjanjikan” Diakses 10 Desember 2018 dari https://id.techinasia.com/talk/profesi-data-scientist-menjanjikan

[10] “Go-Jek Buka Kantor Data Science di Singapura, Apa Alasannya?” Diakses 01 April 2019 dari https://www.liputan6.com/tekno/read/2998722/go-jek-buka-kantor-data-science-di-singapura-apa-alasannya

[11] Introduction to Data Science. A Python Approach to Concepts,Techniques and Applications. Laura Igual, Santi Segui. Tahun

2017.

[12] Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly. 2014.

[13] Jure Leskovek, Anand Rajaraman and Jerey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014. (free online)

[14] Mohammed J. Zaki andWagner Miera Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press. 2014.

[15] https://anaktik.com/data-science/

[16] https://anaktik.com/skill-data-scientist/