|  |  |
| --- | --- |
|  |  |
|  | **MODUL 1 PERKULIAHAN** |
|  |  |
|  | **Proyek Data Science** |
|  |  |
|  | **Pengertian Data Science dan contoh pemanfaatannya** |
|  |  |

# Pembahasan

**Pertemuan 1**

**Pengertian Data Science dan contoh pemanfaatannya**

Pengertian Data Science dan contoh pemanfaatannya – Ketika Kita memasuki era big data dan data science, kebutuhan untuk penyimpanan tumbuh pesat. Ini adalah tantangan utama bagi industri perusahaan hingga 2010.

Fokus utama adalah untuk membangun kerangka kerja dan solusi untuk menyimpan data. Seperti Hadoop dan kerangka kerja lainnya yang telah berhasil memecahkan masalah penyimpanan, namun sekarang fokus dunia telah bergeser ke pengolahan data . ilmu Data adalah saus rahasia di sini. Semua ide yang Anda lihat di sci-fi Hollywood, benar-benar bisa berubah menjadi kenyataan dengan Data Science. AI(Artificial Intellegent) adalah masa depan dari data science. Oleh karena itu, penting untuk memahami apa Data Science dan bagaimana hal itu dapat mengubah bisnis anda.

pengertian Data Science / ilmu data adalah sebuah bidang ilmu pengetahuan yang berfokus untuk memecahkan suatu masalah menggunakan data. baik data yang terstruktur maupun tidak terstruktur.

Sebelum kita membahas lebih jauh tentang data science, tidak ada salahnya untuk kita mengenal apa itu data. Data adalah catatan dari Kumpulan fakta. fakta fakta inilah yang nantinya akan diolah menggunakan data science dan menghasilkan informasi. setelah sebuah informasi yang di dapat, informasi tersebut bisa menjadi data lagi dan diolah lagi menjadi informasi baru.

contoh: sebuah perusahaan memiliki data penjualan selama bulan januari, dan data tersebut akan menghasilkan informasi berapa laba penjualan selama bulan januari. lalu informasi yang terkumpul itu bisa kita olah lagi untuk mendapatkan informasi laba penjualan dalam satu tahun. data sendiri bisa memiliki banyak format seperti text,angka,gambar,audio,dll.

Pengertian Data Science

Pertama, kita harus mengenal apa itu data science. Data Science merupakan perpaduan dari berbagai alat, algoritma dan prinsip-prinsip pembelajaran mesin dengan tujuan menemukan pola yang tersembunyi dari data mentah. lalu apa yang membedakannya dengan ilmu statistik pada umumnya?

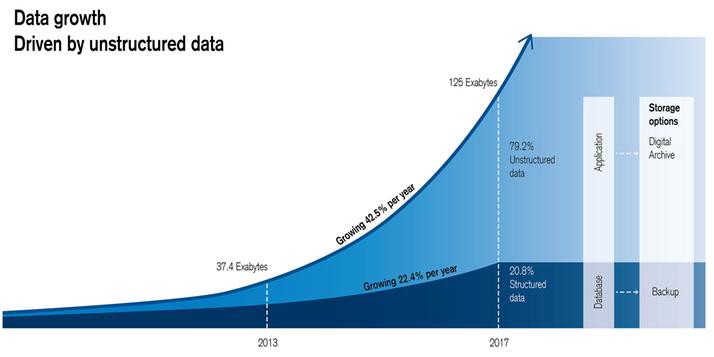
Jawabannya terletak pada perbedaan antara menjelaskan dan memprediksi.

Seperti yang Anda lihat dari gambar di atas, seorang data analyst biasanya menjelaskan apa yang terjadi dengan data . Di sisi lain, data Scientist tidak hanya melakukan analisis eksplorasi untuk menemukan wawasan dari itu, tetapi juga menggunakan berbagai algoritma pembelajaran mesin canggih untuk mengidentifikasi terjadinya peristiwa tertentu di masa depan. Seorang data Scientist akan melihat data dari berbagai sudut, kadang-kadang sudut yang sebelumnya tidak diketahui.

Ilmu Data lebih sering digunakan untuk membuat keputusan dan prediksi menggunakan analisis sebab-akibat prediktif, analytics (ilmu prediksi dan keputusan) dan machine learning.

Mengapa kita membutuhkan Data Science

Data yang kita miliki saat ini sebagian besarnya tidak terstruktur dan ukurannya terlalu kecil untuk dapat dianalisis menggunakan alat BI sederhana. Tidak seperti dahulu dimana sebagian besar data yang ada di dalam sistem adalah data yang terstruktur, sekarang sebagian besar data menjadi tidak terstruktur dan semi-terstruktur. Mari kita lihat tren data pada gambar yang diberikan di bawah ini yang menunjukkan bahwa pada tahun 2020, lebih dari 80% dari data akan tidak terstruktur.



sumber gambar: https://www.kdnuggets.com/2019/10/data-scientist-data-management.html

Data ini dihasilkan dari berbagai sumber seperti log keuangan, file teks, bentuk multimedia, sensor, dan instrumen. Alat BI sederhana tidak dapat memproses volume besar dan berbagai data. Inilah sebabnya mengapa kita perlu alat analitik dan algoritma yang lebih kompleks dan canggih untuk mengolah, menganalisis, dan menarik wawasan yang berarti dari itu.

Ini bukan satu-satunya alasan mengapa Data science menjadi begitu populer. Mari kita menggali lebih dalam dan melihat bagaimana ilmu pengetahuan digunakan di berbagai bidang data.

Sebuah perusahaan pasti ingin member produk yang sesuai untuk pelanggannya, bagaimana jika anda bisa membuat sebuah prediksi yang mem=nentukan produk yang sesuai dengan pelanggan tersebut berdasarkan data yang ada seperti sejarah masa lalu pencarian, sejarah pembelian, usia, dan gaji. tentunya hal tersebut akan meningkatkan penjualan di perusahaan tersebut.

di masa depan, seluruh mobil akan bisa mengendarai dirinya sendiri, lalu bagaimana mobil tersebut dapat mengenali lingkungan sekitarnya? Self-driving car akan mengumpulkan data langsung dari sensor, termasuk radar, kamera dan laser untuk membuat peta sekitarnya. Berdasarkan data ini, dia membuat keputusan seperti kapan harus mempercepat, saat menyalip,dan berhenti menggunakan algoritma machine learning canggih

Mari kita lihat bagaimana data sains dapat digunakan dalam analisis prediktif. Mari kita ambil sebagai contoh ramalan cuaca. Data dari kapal, pesawat, radar, satelit dapat dikumpulkan dan dianalisa untuk membangun model. Model-model ini tidak hanya akan memprediksi cuaca tetapi juga membantu dalam memprediksi terjadinya setiap bencana alam. Ini akan membantu Anda untuk mengambil tindakan yang tepat sebelum dan menyelamatkan banyak nyawa yang berharga.

data scientist

orang yang ahli dalam bilang data science biasa disebut data scientist, sedangkan pengertian Data Scientist sendiri adalah orang yang mengatur data dalam jumlah besar, mengumpulkan dan menganalisis sekumpulan besar data terstruktur dan tidak terstruktur.

Dalam suatu organisasi, data scientist bertindak sebagai orang yang menganalisis dan mempelajari pola suatu data intuk mendapatkan informasi, dan informasi tersebulah yang akan membantu perusahaan mengambil keputusan.

Data scientist tidak bisa bekerja sendiri, mereka membutuhkan bantuan tenaga lain seperti data engineer dan software engineer untuk mengaplikasikan temuan mereka ke dalam suatu aplikasi, gaji data scientist sendiri tergolong sangat besar, menurut world economic forum, data scientist menduduki peringkat dengan gaji terbesar dibanding dengan pekerjaan terkait lainnya, gaji rata-rata data scientist di us adalah $114,808 atau hampir setara 2,5M rupiah/tahun. dan di indonesia sendiri gaji seorang professional data scientist berkisar 30jt/bulan atau sekitar 360 jt/tahun dan tentunya gaji yang besar juga mendatangkan tanggung jawab yang besar pula, seorang data scientist dituntut memiliki skill yang tinggi, baik soft skill maupun hard skill.

skill yang harus dimiliki oleh seorang data scientist

Programming

Seorang data scientist harus memahami beberapa bahasa pemrograman sekaligus, tergantung kepada kebijakan dan keperluan kantor, namun bahasa pemrograman yang paling sering kita jumpai untuk data scientist adalah python dan R, karena keduanya menawarkan banyak tools dan framework yang bisa digunakan namun terkadang industri juga memerlukan bahasa lain untuk membantu pengembangan aplikasinya, ada banyak faktor yang perlu dipertimbangkan saat memilih bahasa pemrograman yang pas seperti, speed training, syntaks, framework yang di butuhkan, dan kesesuaiannya dengan aplikasi yang akan dibuat.

berikut beberapa bahasa pemrograman yang paling sering digunakan untuk data science:

* Python
* R
* Java
* SQL
* Julia
* Scala
* MatLab

Algoritma

Selain Memahami sintaks dari bahasa pemrograman, Data Scientist Juga dituntut untuk memahami Algoritma serta memilih Algoritma yang tepat untuk suatu case. inilah yang membedakan Seorang Data Scientist dengan Data analyst. Banyak sekali algoritma yang bisa diterapkan dalam memecahkan berbagai kasus, beberapa diantaranya:

Natural Language Processing

* Classification
* Clustering
* Ensemble methods
* Deep Learning
* Data Visualization / persentation tool

seorang data scientist juga harus mengerti tools untuk mempresantisakan sebuah informasi yang didapat, kenapa? karena informasi yang telah didapat tersebut akan digunakan untuk membantu pengambilan keputusan. sedangkan orang yang berhak mengambil keputusan suatu perusahaan adalah seorang manager/direktur. oleh karena itu seorang data scientist harus menjelaskan informasi yang didapatnya tsb kepada orang terkait di perusahaan tsb, tidak mungkin kan kita langsung memberikan hasil analysis kita dalam bentuk code matlab/python secara langsung.

Alat yang digunakan pun sangat beragam, untuk beberapa kasus kita bisa menggunakan tools visualisasi seperti matplotlib, dan beberapa kasus lainnya kita harus memberikan informasi yang detail berupa persentasi.

Big Data Platform

Di era big data semua data yang terkumpul jumlahnya sangatlah besar, oleh karena itu diperlukan platform yang mampu menampung data sebanyak itu dan memprosesnya. biasanya industri menggunakan data warehouse untuk urusan ini. sedangkan perseorang biasanya masih menggunakan database Relational untuk hal ini.

Kenapa Data warehouse? kamu bisa membacanya di apa itu data warehouse dan kelebihannya

ada banyak sekali big data platform, beberapa diantaranya:

* Microsoft Azure
* Hadoop
* Microsoft Azure
* Cloudera
* Sisense
* Collibra
* Tableau
* MapR
* Qualtrics
* Oracle
* MongoDB
* Datameer

Soft Skill

Seorang data scientist juga dituntut memiliki soft skill yang berguna untuk menunjang karirnya. Seperti contoh, memiliki skill komunikasi yang bagus, intuisi yang tinggi terhadap data, Mampu bekerja sama dalam team,dll.

beberapa soft skill didapat dari sifat alami orang tersebut namun kamu bisa mengasahnya dengan telus berlatih, bekerja atau mengikuti pelatihan dan bootcamp.

Analitik causal prediktif – Jika Kita menginginkan model yang dapat memprediksi kemungkinan suatu peristiwa di masa depan, Kita bisa menerapkan data analytics kausal prediktif. Katakanlah, jika Kita memberikan uang secara kredit, maka kemungkinan pelanggan melakukan pembayaran kredit tepat waktu adalah masalah yang krusial untuk kita. Di sini, kita bisa membuat model yang dapat melakukan analisis prediktif dengan menggunakan data riwayat pembayaran pelanggan untuk memprediksi apakah pembayaran di masa mendatang akan tepat waktu atau tidak.

Analitik preskriptif: Jika kita mau memiliki model yang bisa mengambil keputusan sendiri dan kemampuan untuk memodifikasi dirinya dengan parameter yang dinamis, kitabisa menggunakan analitik preskriptif. Analitik preskriptif juga akabn memberikan saran. Dalam istilah lain, tidak hanya memprediksi tetapi menyarankan serangkaian tindakan yang ditentukan dan hasil terkait.

Contoh terbaik untuk ini adalah mobil self-driving Google yang telah saya bahas sebelumnya juga. Data yang dikumpulkan oleh kendaraan dapat digunakan untuk melatih mobil yang bisa mengemudi sendiri. kita bisa menjalankan algoritma pada data ini untuk memberikan kecerdasan padanya. Ini akan memungkinkan mobil kita untuk mengambil keputusan seperti kapan harus berbelok, jalan mana yang harus diambil, kapan harus memperlambat atau mempercepat.

Machine learning untuk prediksi— Jika Anda memiliki data transaksional perusahaan keuangan dan perlu membuat model untuk menentukan tren masa depan, maka algoritma machine learning ini adalah yang terbaik. Ini berada di bawah paradigma supervised learning. Itu disebut diawasi karena Anda sudah memiliki data berdasarkan di mana Anda dapat melatih mesin Anda. Misalnya, model pendeteksian kecurangan dapat dilatih menggunakan catatan riwayat pembelian yang curang.

Machine learning untuk menemukan pola— Jika Kita tidak memiliki parameter berdasarkan data yang dikumpulkan untuk membuat prediksi, maka Kita perlu mengetahui terlebih dahulu pola tersembunyi dalam dataset untuk mendapat prediksi yang bermakna. Ini tidak lain adalah model unsupervised learning. karena Kita tidak memiliki label yang sudah ditentukan sebelumnya untuk dikelompokkan. Algoritma yang paling umum digunakan untuk penemuan pola adalah Clustering.

Katakanlah Kita bekerja di perusahaan telepon dan kita perlu membangun jaringan dengan meletakkan menara di suatu wilayah. Kemudian, Kita dapat menggunakan teknik pengelompokan untuk menemukan lokasi menara yang akan memastikan bahwa semua pengguna menerima kekuatan sinyal yang optimal.

# 6 Skill yang butuhkan untuk menjadi Data scientist

Pemanfaatkan Big Data,telah meningkatkan kebutuhan profesi data scientist di setiap perusahaan. semakin lama juga skill yang dibutuhkan data scientist juga bertambah. data science bertujuan untuk meningkatkan proses pengembangan produk, meningkatkan retensi pelanggan, atau saya melalui data untuk menemukan peluang bisnis baru – organisasi semakin mengandalkan keahlian dari para ilmuwan data untuk mempertahankan, tumbuh, dan mengalahkan pesaing mereka.

Skill Dibutuhkan untuk Menjadi Data Scientist

Artikel ini bertujuan untuk menjawab pertanyaan ini. kita akan mendalami keterampilan teknis dan non-teknis yang sangat penting untuk anda yang ingin menjadi data scientist.

Keterampilan teknis Diperlukan untuk Menjadi Scientist data

statitical analytic dan keterampilan dalam menggunakan cloud computing seperti data mining dan processing yang mengekstrak informasi dari sekumpulan data yang terstruktur maupun tidak terstruktur menjadi skill wajib yang harus dimiliki data scientist Ini berarti Anda harus terampil dalam matematika, pemrograman, dan statistik. Salah satu cara untuk memenuhi prasyarat adalah untuk memiliki latar belakang akademis yang berkaitan.

Data scientist biasanya memiliki gelar Ph.D. atau gelar Master dalam statistik, ilmu komputer atau rekayasa. Ini memberi mereka fondasi yang kuat untuk terhubung dengan poin teknis yang membentuk inti dari praktek di bidang Data Science.

keterampilan teknis lainnya yang dibutuhkan untuk menjadi seorang ilmuwan data yang meliputi:

1.Pemrograman

Anda harus memiliki pengetahuan tentang bahasa pemrograman seperti Python, Perl, C / C ++, SQL, dan java- Python menjadi bahasa yang paling umum yang digunakan dalam data science. Pemrograman membantu Anda untuk membersihkan, dan mengatur set data yang tidak terstruktur.

2.Pengetahuan tentang SAS dan Alat Analisis Lainnya

Memahami alat-alat analisis yang akan membantu Anda untuk mengekstrak informasi yang berharga dari data set yang telah dibersihkan, dan terorganisir. SAS, Hadoop, Spark, Hive, Pig, dan R adalah alat untuk data analytic yang paling populer dan paling sering digunakan. Sertifikasi dan pelatihan tools ini dapat membantu membangun portfolio berharga dan menjadikan anda ahli dalam data analytic.

3.Mahir Bekerja dengan Data yang Tidak Terstruktur

Ketika berbicara tentang keterampilan untuk dapat bekerja dengan data yang tidak terstruktur, kami sangat menekankan kemampuan data science untuk memahami dan pengolahan data yang tidak terstruktur yang berasal dari saluran yang berbeda. Dengan demikian, jika seorang data scientist yang bekerja untuk proyek-proyek pemasaran untuk membantu tim pemasaran melakukan riset, seorang data scientist juga diharuskan mencari data dari media social terkait.

Skill Non-Teknis Dibutuhkan untuk Menjadi Scientist data

Kami sekarang akan mengalihkan fokus kami pada keterampilan non-teknis yang diperlukan untuk menjadi seorang ilmuwan data. Hal ini mengacu pada keterampilan pribadi dan karena itu, bisa sulit untuk hakim hanya dengan melihat pendidikan kualifikasi, sertifikasi, dan sebagainya. Mereka:

1.Memiliki Pengetahuan Tentang Bisnis

Jika seorang data scientist tidak memiliki pengetahuan bisnis dan pengetahuan tentang unsur-unsur yang membentuk sebuah model bisnis yang sukses, maka semua ilmu yang anda miliki mejadi tidak produktif. Anda tidak akan bisa membedakan antara masalah dan potensi tantangan yang perlu diselesaikan agar bisnis dapat bertahan hidup dan berkembang. Anda tidak akan benar-benar dapat membantu organisasi Anda untuk mencari peluang bisnis baru.

2.Keterampilan Komunikasi Yang Kuat

Anda seorang data scientist dan memahami data lebih baik daripada siapa pun . Namun, agar Anda untuk berhasil dalam peran Anda, dan membantu organisasi membentuk bisnis mereka, Anda harus berhasil mengkomunikasikan pemahaman Anda dengan orang yang akan menggunakan data tsb yang tentunya mereka adalah orang non-teknis yang tidak memiliki pengetahuan tentang data. jadi Anda harus memiliki kemampuan komunikasi yang kuat agar orang lain mengerti apa yang anda maksud.

3.Memiliki intuisi yang besar terhadap data

Ini mungkin salah satu keterampilan yang paling non-teknis yang dibutuhkan paling signifikan oleh seorang data scientist dari data. Intuisi berarti memahami pola dari sebuah data yang besar di mana tidak bisa terlihat di permukaan dan mengetahui letak nilai utama pada tumpukan bit data yang belum dijelajahi. Para ilmuwan ini membuat data lebih efisien dalam pekerjaan mereka. Ini adalah keterampilan yang datang dari pengalaman dan pelatihan.boot camp adalah salah satu cara yang tepat untuk mengembangkan skill ini.

Quiz Pertemuan 1

Tuliskanlah Manfaat dari proyek data science dan berikan contoh penerapan dari studi kasus yang ada saat ini.

# 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/