



PROYEK SAINS DATA

Husni | Ika Oktavia Suzanti

PROYEK SAINS DATA

PROYEK SAINS DATA

Husni
Ika Oktavia Suzanti



PROYEK SAINS DATA

© Penerbit Perkumpulan Rumah Cemerlang Indonesia (PRCI)

Penulis:
Husni
Ika Oktavia Suzanti

Editor: Rusli

Cetakan Pertama: Juni 2023

Cover: Tim Kreatif PRCI

Tata Letak: Tim Kreatif PRCI

Hak Cipta 2022, pada Penulis. Diterbitkan pertama kali oleh:

Perkumpulan Rumah Cemerlang Indonesia
ANGGOTA IKAPI JAWA BARAT

Pondok Karisma Residence Jalan Raflesia VI D.151
Panglayungan, Cipedes Tasikmalaya – 085223186009

Website: www.rcipress.rcipublisher.org
E-mail: rumahcemerlangindonesia@gmail.com

Copyright © 2023 by Perkumpulan Rumah Cemerlang Indonesia
All Right Reserved

- Cet. I –: Perkumpulan Rumah Cemerlang Indonesia, 2023
Dimensi : 15,5 x 23 cm
ISBN: 978-623-448-547-9

Hak cipta dilindungi undang-undang
Dilarang memperbanyak buku ini dalam bentuk dan dengan
cara apapun tanpa izin tertulis dari penulis dan penerbit

Undang-undang No.19 Tahun 2002 Tentang
Hak Cipta Pasal 72

Undang-undang No.19 Tahun 2002 Tentang Hak Cipta
Pasal 72

Barang siapa dengan sengaja melanggar dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal ayat (1) atau pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling sedikit 1 (satu) bulan dan/atau denda paling sedikit Rp.1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp.5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta terkait sebagai dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp.500.000.000,00 (lima ratus juta rupiah).

KATA PENGANTAR

Puja dan puji syukur penulis haturkan kepada Allah Subhanahu Wata'ala yang telah memberikan banyak nikmat, taufik, dan hidayah. Sehingga penulis dapat menyelesaikan Buku Ajar “Proyek Sains Data” dengan baik tanpa ada halangan yang berarti.

Buku ini telah penulis selesaikan berkat kerja sama dan bantuan dari beberapa pihak. Oleh karena itu penulis sampaikan banyak terima kasih kepada segenap pihak yang telah berkontribusi secara maksimal dalam penyelesaian buku ini. Penulis mengucapkan terima kasih kepada Dekan Fakultas Teknik dan Program Studi Teknik Informatika yang telah memberi kesempatan kepada penulis untuk mendapatkan hibah buku ajar ini.

Diluar itu penulis sebagai manusia biasa menyadari sepenuhnya bahwa masih banyak kekurangan dalam penulisan laporan ini, baik dari segi tata bahasa, susunan kalimat maupun isi. Oleh sebab itu dengan segala kerendahan hati, saya selaku penyusun menerima segala kritik dan saran yang membangun dari pembaca. Semoga buku ajar ini bisa menambah refrensi dan membawa manfaat bagi para pembaca.

Penulis

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI	ii
DAFTAR GAMBAR.....	iv
BAB I PENGENALAN SAINS DATA.....	1
A. Sains Data vs Penambangan Data	1
B. Manfaat Sains Data.....	9
C. Pentingnya Sains data di Berbagai Industri.....	12
D. Aplikasi Sains data dalam Kehidupan Nyata	14
E. Sains Data di bidang Ritel.....	16
F. Kesimpulan.....	38
BAB II PROBABILITAS, STATISTIKA DAN ALJABAR LINIER.....	40
A. Apa itu Probabilitas dan Statistik dalam Sains Data?	40
B. Konsep Stastistik pada Sains Data	42
C. Konsep Probabilitas untuk Sains Data	51
D. Bagaimana Probabilitas dan Statistik Digunakan dalam Sains Data?.....	54
E. Bagaimana Cara Mempelajari Probabilitas dan Statistika pada Sains Data?	55
F. Teknik Probabilitas dan Statistik untuk Ilmuwan Data	57
G. Kesimpulan.....	59
BAB III MANIPULASI DAN VISUALISASI DATA.....	61
A. Numpy dan Pandas	61
B. Visualisasi Data.....	83
C. Kesimpulan.....	93
BAB IV SIKLUS PROYEK SAINS DATA CRISP-DM DAN TDSP.....	94

A. CRISP-DM.....	94
B. TDSP.....	111
C. Kesimpulan	120
BAB V EXPLORATORY DATA ANALYSIS (EDA).....	121
A. Apa itu Exploratory Data Analysis dalam Sains Data?	121
B. Peran EDA dalam Sains Data	122
C. Langkah-langkah dalam <i>Exploratory Data Analysis</i> (EDA)	123
D. Jenis-jenis <i>Exploratory Data Analysis</i> (EDA)	126
E. Alat-alat <i>Exploratory Data Analysis</i> (EDA).....	132
F. Keuntungan dari Penggunaan EDA.....	134
G. Contoh dari Exploratory Data Analysis (EDA)	135
H. Kesimpulan	139
BAB VI STUDI KASUS I	141
BAB VII STUDI KASUS II	191
A. Customer Churn	191
B. Decision Tree C4.5	192
C. Menyiapkan Data Training	193
D. Dataset	194
E. Script Program.....	215
REFERENSI	223

DAFTAR GAMBAR

Gambar 1. Siklus Penambangan Data	5
Gambar 2. Grup chart dengan bentuk bar.....	85
Gambar 3. Grafik bar sederhana	86
Gambar 4. Chart Bar.....	88
Gambar 5. Chart Bar Horizontal.....	90
Gambar 6. Chart Bar Kelompok.....	92
Gambar 7. Tahapan CRISP-DM	96
Gambar 8. Alur kerja dari sains data.....	114
Gambar 9. Laman Kolom Review Shopee.....	170
Gambar 10. Rata-Rata Hasil Evaluasi Full Data	188
Gambar 11. Rata-Rata Hasil Evaluasi Data Balance	188
Gambar 12. Rata-Rata Hasil Evaluasi Full Data	189
Gambar 13. Decission Tree.....	193
Gambar 14. Root Node Decision Tree.....	201
Gambar 15. Node 1	207
Gambar 16. Node 2 Decision Tree.....	213
Gambar 17. Ilustrasi Decision Tree.....	214
Gambar 18 Hasil Klasifikasi	220
Gambar 19. Rule Decision Tree C4.5	222

BAB I

Pengenalan Sains Data

A. Sains Data vs Penambangan Data

Di antara semua istilah teknis, sains data dan penambangan data mungkin yang paling banyak dibicarakan di dunia data. Bisnis dan organisasi mulai menyadari nilai besar yang tersembunyi dalam sejumlah besar data yang diambil setiap hari. Hal ini membuat pelaku bisnis maupun organisasi mulai memperkenalkan dan menggunakan berbagai teknik untuk mengolah potensi dan nilai tersebut. Tujuan utamanya untuk memperoleh informasi ataupun wawasan yang dapat ditindaklanjuti dari data yang tersedia. Dan meskipun beberapa orang menggunakan sains data dan penambangan data secara bergantian, ada perbedaan yang signifikan antara kedua istilah tersebut. Berikut akan dibahas tentang beberapa perbedaan paling menonjol antara sains data dan penambangan data.

Apa itu Sains data?

Sains data sebagai sebuah istilah, dapat ditelusuri kembali ke tahun 1974, ketika Peter Naur mengusulkannya sebagai nama alternatif untuk ilmu komputer. John Tukey pada tahun 1962, menggambarkan bidang yang menyerupai sains data modern dan menyebutnya sebagai data analisis. Pada tahun 1997, CF Jeff Wu menyarankan agar statistik diubah namanya menjadi sains data. Pada tahun berikutnya, Chikio Hayashi berpendapat bahwa sains data harus menjadi konsep interdisipliner yang sama sekali baru

yang terdiri dari tiga aspek—desain data, pengumpulan data, dan analisis data.

Sains data adalah bidang atau domain interdisipliner yang melibatkan penggunaan metode ilmiah, algoritma, proses, dan sistem untuk mengekstraksi pengetahuan dan wawasan dari sejumlah besar data terstruktur dan tidak terstruktur. Ini, pada gilirannya, digunakan untuk membangun model analitik prediktif, preskriptif, dan preskriptif.

Sains data terkait dengan big data, deep learning, dan penambangan data. Jadi secara tidak langsung, sains data merupakan irisan dari data dan komputasi dimana bahasannya dimulai dari mencari, mempersiapkan (membangun model), menganalisis (memvalidasi model), dan memanfaatkan data (menerapkan model terbaik). Sains data memadukan bisnis dengan ilmu komputer dan statistik. Enam langkah yang terlibat dalam proses sains data adalah:

- Menemukan Masalah: Sebelum menyelesaikan masalah, penting untuk mengetahui apa masalahnya; pertanyaan data terlebih dahulu harus diterjemahkan ke pertanyaan bisnis yang dapat ditindaklanjuti.
- Pengumpulan Data Mentah yang Diperlukan untuk Masalah: Data yang diperlukan harus dikumpulkan untuk mendapatkan wawasan dan kemungkinan solusi dengan memindai database internal atau membeli database dari sumber eksternal.

- **Memproses Data untuk Analisis:** Data harus diproses sebelum melangkah lebih jauh dan menganalisisnya untuk wawasan yang lebih akurat.
- **Meng-explore Data:** Langkah penting ini melibatkan pengembangan ide untuk membantu mengidentifikasi pola dan wawasan yang tersembunyi.
- **Melakukan Analisis Mendalam:** Pada tahap ini, pengetahuan matematika, statistik, dan teknologi serta alat sains data harus diterapkan untuk mengolah data dengan sukses dan menemukan serta memperoleh setiap wawasan yang mungkin serta faktor penting lainnya. Data kuantitatif dan kualitatif dapat digabungkan dan dipindahkan ke dalam tindakan.
- **Mengkomunikasikan Hasil Analisis:** Pada langkah ini, wawasan dan temuan disampaikan kepada kepala penjualan agar mereka memahami pentingnya temuan dan bagaimana temuan tersebut dapat membantu pertumbuhan bisnis.

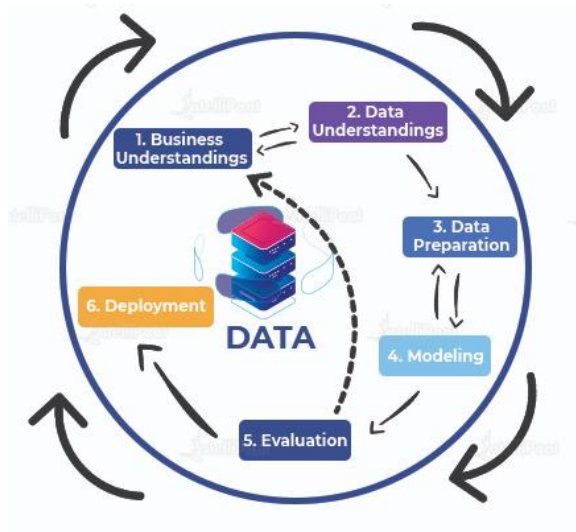
Apa itu Penambangan Data?

Istilah data mining muncul pada tahun 1990 di komunitas basis data. Penambangan data digunakan oleh perusahaan ritel dan komunitas keuangan untuk tujuan menganalisis data dan mengidentifikasi tren untuk meningkatkan basis pelanggan, dan memprediksi fluktuasi harga saham, suku bunga, dan permintaan pelanggan. Penambangan data adalah proses mengidentifikasi pola dalam kumpulan data besar. Ini melibatkan metode meliputi sistem basis data, statistik, dan pembelajaran mesin. Tujuan

keseluruhan dari subbidang ilmu komputer dan statistik interdisipliner ini adalah untuk mengekstrak informasi dari kumpulan data besar atau pustaka data dengan menggunakan algoritma matematika yang canggih dan mengubahnya menjadi struktur yang dapat dipahami untuk digunakan lebih lanjut.

Penambangan data membantu memperoleh wawasan melalui ekstraksi yang cermat, peninjauan, dan pemrosesan data mentah untuk menemukan pola dan korelasi yang dapat bermanfaat bagi bisnis. Proses penambangan data mencakup berbagai jenis layanan seperti:

- Penambangan web
- Penambangan teks
- Penambangan audio
- Penambangan video
- Penambangan data jaringan sosial
- Penambangan data bergambar



Gambar 1. Siklus Penambangan Data

Penambangan data juga disebut sebagai penemuan pengetahuan dalam data (KDD), dilakukan dengan bantuan perangkat lunak sederhana atau lanjutan. Langkah-langkah pada gambar 1 merupakan ilustrasi tahapan dalam penambangan data. Adapun penjelasannya sebagai berikut:

- **Pemahaman Bisnis:** Ini melibatkan pengenalan dan pemahaman tujuan dan pekerjaan bisnis serta memahami faktor-faktor penting yang akan membantu mencapai target bisnis.
- **Pemahaman Data:** Ini melakukan pengumpulan data dan akumulasi data. Data didaftar berdasarkan data sumber, lokasinya, cara pencapaiannya, dan jika ada masalah yang muncul. Data tersebut kemudian divisualisasikan dan diperiksa kelengkapannya.

- **Persiapan Data:** Ini melibatkan pemilihan data yang berguna, membersihkannya, membangun atribut darinya, dan integrasi data dari banyak basis data.
- **Pemodelan:** Ini melibatkan pemilihan teknik penambangan data, menghasilkan desain pengujian untuk mengevaluasi model yang dipilih, membangun model dari kumpulan data, dan mengevaluasi model dengan para ahli untuk mengetahui hasilnya.
- **Evaluasi:** Ini menentukan sejauh mana model yang dihasilkan memenuhi persyaratan bisnis dengan mengujinya berdasarkan aplikasi nyata.
- **Penyebaran:** Ini membuat rencana penerapan dan membentuk strategi untuk memeriksa kegunaan model penambangan data melalui pemeliharaan dan pemantauan.

Penambangan Data vs Sains data

- Perbedaan terbesar antara sains data dan penambangan data terletak pada istilahnya. Sementara sains data adalah bidang luas meliputi pengambilan data, analisis data, dan memperoleh wawasan yang dapat ditindaklanjuti darinya, penambangan data terutama melibatkan menemukan informasi yang berguna dalam kumpulan data dan memanfaatkannya untuk mengidentifikasi pola tersembunyi.
- Perbedaan besar lainnya antara sains data dan penambangan data adalah bahwa yang pertama adalah bidang multidisiplin yang terdiri dari statistik, visualisasi data, ilmu sosial, pemrosesan bahasa alami (NLP), dan

penambangan data, yang berarti bahwa penambangan data adalah bagian dari sains data.

- Seorang ilmuwan data dapat dianggap, sampai batas tertentu, kombinasi dari peneliti kecerdasan buatan (AI), insinyur pembelajaran mesin, insinyur pembelajaran mendalam, dan analis data. Di sisi lain, seorang profesional penambangan data tidak dapat serta merta melakukan semua peran ini, yang dapat dilakukan oleh seorang ilmuwan data.
- Perbedaan penting lainnya terletak pada jenis data yang digunakan. Sains data sebagian besar berurusan dengan semua jenis data seperti terstruktur, tidak terstruktur, dan semi-terstruktur. Namun, penambangan data sebagian besar berhubungan dengan data terstruktur.
- Saat mempertimbangkan sifat pekerjaan, ada perbedaan lain antara sains data dan penambangan data. Mengungkap pola dan menganalisisnya adalah komponen kunci dari penambangan data. Sains data melibatkan hal yang sama, tetapi juga melibatkan peramalan peristiwa masa depan dengan memanfaatkan data saat ini dan historis menggunakan berbagai alat dan teknologi.
- Sains data berfokus pada sains data, sedangkan penambangan data terutama berkaitan dengan proses mendeteksi anomali dan ketidakkonsistenan dan memprediksi hasil.

No	Sains data	Penambangan Data
1	Sains data adalah bidang studi.	Penambangan data adalah teknik yang merupakan bagian dari proses KDD.
2	Ini tentang mengumpulkan, memproses, menganalisis, dan memanfaatkan data dalam berbagai operasi.	Ini tentang mengekstraksi informasi berharga dari data.
3	Tujuannya adalah untuk membangun produk yang dominan data untuk sebuah usaha.	Tujuannya adalah untuk menyadari nilai data dan membuatnya dapat digunakan dengan mengekstraksi informasi penting.
5	Ini berkaitan dengan semua jenis data—terstruktur, tidak terstruktur, atau semi-terstruktur.	Ini terutama berkaitan dengan data terstruktur.
6	Ini melibatkan pengikisan data, pembersihan, visualisasi, statistik, dll.; oleh karena itu, ini adalah superset dari penambangan data.	Ini adalah bagian dari sains data karena aktivitas penambangan berada dalam jalur sains data.
7	Ini pada dasarnya diterapkan untuk tujuan ilmiah.	Ini terutama digunakan untuk tujuan bisnis.
8	Ini secara luas berfokus pada sains data.	Itu lebih terlibat dengan prosesnya.

B. Manfaat Sains Data

Sains data adalah studi tentang penggalian wawasan dari sejumlah besar data menggunakan berbagai pendekatan, proses, dan algoritma ilmiah. Perkembangan big data, analisis data, dan statistik kuantitatif telah memunculkan istilah "sains data". Sains data sekarang lebih penting dari sebelumnya karena digunakan untuk transformasi data. Di masa lalu, data dalam format terstruktur, ringkas, dan dapat diproses dengan alat BI (Business Intelligence) langsung. Namun, sebagian besar data saat ini bersifat semi-terstruktur atau tidak terstruktur, berbentuk multimedia seperti foto, audio, dan video. Oleh karena itu, metode analitik yang canggih yang dapat menangani volume besar dari data yang heterogen diperlukan untuk bentuk data ini.

Kebutuhan akan seorang ilmuwan data berkembang seiring dengan nilai data. Ilmuwan data semakin diperlukan di berbagai profesi, lembaga pemerintah, dan organisasi nirlaba. Ilmuwan informasi dan komputer, pemrogram basis data dan perangkat lunak, kurator, dan anotator berpengetahuan adalah contoh ilmuwan data. Semuanya penting untuk keberhasilan administrasi pengumpulan data digital. Ilmuwan data berusaha keras untuk memilah-milah segunung data untuk menemukan informasi yang relevan dan mengidentifikasi pola dan desain yang dapat digunakan untuk menentukan tujuan dan sasaran di masa depan. Ini menunjukkan mengapa sains data penting dan ilmuwan data menjadi lebih terkenal dan signifikan.

Kebutuhan akan Sains data

Tujuan sains data adalah untuk menemukan pola. Dengan demikian, sains data diperlukan karena alasan berikut:

- Pengusaha bisnis dapat mengenali pelanggan dengan lebih baik dan lebih halus dengan bantuan sains data. Pelanggan adalah pondasi dari produk apapun dan sangat penting untuk keberhasilan atau kegagalan sebuah bisnis. Sains data memungkinkan bisnis untuk berinteraksi dengan pelanggan mereka dengan cara baru, memastikan kualitas dan kekuatan superior produk.
- Pentingnya sains data dalam industri teknologi informasi, sains data memungkinkan produk untuk mengekspresikan cerita mereka dengan kuat dan menawan. Produk dan bisnis dapat terhubung dengan lebih baik dengan pelanggan mereka saat mereka menggunakan data ini untuk menceritakan kisah mereka kepada pemirsa.
- Salah satu aspek penting dari Sains data adalah bahwa hasilnya dapat digunakan di setiap industri, termasuk perjalanan, kesehatan, dan pendidikan. Industri dapat dengan cepat memeriksa masalah mereka dan berhasil mengatasinya menggunakan sains data.
- Sains data saat ini tersedia di hampir semua sektor, dan ada sejumlah besar data di dunia saat ini yang bergantung pada cara penggunaannya, dapat menentukan apakah suatu produk berhasil atau gagal. Jika data digunakan dengan benar, itu akan menjadi penting untuk upaya pencapaian tujuan produk di masa depan.

Mengapa Sains data Penting di Dunia Saat Ini?

Pada abad ke-21, sains data dianggap sebagai karir yang menguntungkan. Ini hanyalah studi matematika, statistik, dan ilmu komputer untuk mengekstraksi informasi dari data terstruktur dan tidak terstruktur. Pasar bergeser dengan cara yang luar biasa saat ini karena semakin banyak orang berbicara tentang AI dan pembelajaran mesin. Sains data, yang memecahkan masalah dengan menghubungkan data yang relevan untuk digunakan nanti, membantu teknologi baru ini. Teknologi pengenalan wajah Facebook adalah contohnya karena mengumpulkan banyak data tentang pengguna saat ini dari waktu ke waktu dan menggunakan teknik yang sama untuk mengidentifikasi pengguna baru. Seiring kemajuan alat, teknik sains data berkembang dan menjadi lebih banyak digunakan oleh masyarakat umum. Meskipun tidak semua profesional akan menjadi analitik atau ahli data, kemampuannya akan menjadi lebih mudah diakses oleh orang-orang yang tidak memiliki keterampilan teknis. Informasi yang diperoleh menghasilkan lebih banyak data untuk solusi teknologi keuangan yang baru dikembangkan seperti komputasi awan dan penyimpanan, yang mudah dibagikan oleh semua bisnis. Namun, ini bisa menjadi tantangan dan memakan waktu bagi bisnis untuk memahami sejumlah besar data tidak terstruktur untuk pengambilan keputusan yang efisien. Sains data populer di dunia modern untuk mengatasi ketidaknyamanan ini.

C. Pentingnya Sains data di Berbagai Industri

Basis data besar dari data terstruktur dan tidak terstruktur harus ditambah menggunakan teknik sains data untuk menemukan pola tersembunyi dan mendapatkan wawasan yang bermanfaat. Sains data sangat penting karena banyak aplikasi yang dapat digunakan untuk itu, mulai dari aktivitas sederhana, seperti meminta rekomendasi Siri atau Alexa, hingga yang lebih canggih, seperti mengoperasikan mobil tanpa pengemudi. Berikut penjelasan mengenai pentingnya sains data dalam masyarakat.

Pentingnya Sains data dalam Bisnis

Tujuan sains data adalah untuk membantu organisasi dalam memahami pola varians dalam data, termasuk informasi klien, tingkat pertumbuhan bisnis, volume data, atau kuantitas terukur apa pun. Dalam sains data, digunakan model statistik dan probabilistik untuk menganalisis perubahan dan peningkatan dalam data historis atau saat ini. Sains data telah mengubah operasi untuk bisnis di seluruh dunia. Berikut ini adalah beberapa manfaat utama menggunakan sains data dalam bisnis:

- Memberi manajemen dan pejabat kemampuan untuk mendorong ide-ide baru: Ilmuwan data adalah kunci untuk menemukan solusi yang lebih baik karena mereka dapat menggunakan pembelajaran mesin untuk mengidentifikasi masalah bisnis yang kompleks seperti tantangan riset operasi. Ilmuwan data bertanggung jawab untuk melaporkan tren industri, pengeluaran sumber daya internal, dan perkiraan keuntungan, mengurangi kemacetan alur kerja, dan

meningkatkan efektivitas arsitektur bisnis dengan tujuan yang terinformasi dengan baik.

- Pengalaman pengguna yang lebih baik: Meskipun ini mungkin bukan hal pertama yang terlintas dalam pikiran, hal itu pada akhirnya akan memengaruhi hal lainnya, termasuk pendapatan dan keuntungan. Peningkatan kepuasan klien mengarah pada peningkatan penjualan. Bahkan jika produk atau layanan Anda sangat baik, Anda akan kehilangan kesempatan jika Anda tidak secara konsisten meningkatkannya.

Pentingnya Sains data dalam Layanan Kesehatan

Sektor perawatan kesehatan menghasilkan kumpulan data yang sangat besar dari data berharga tentang demografi pasien, rencana perawatan, hasil pemeriksaan medis, asuransi, dll. Sejumlah besar data yang tersebar, terstruktur, dan tidak terstruktur yang dihasilkan oleh sistem perawatan kesehatan dapat diproses, dianalisis, diasimilasi, dan dikelola dengan bantuan sains data. Data ini perlu dikelola dan dianalisis secara efektif untuk mendapatkan temuan yang benar.

Analitik data besar dan sains data dapat membantu pengambilan keputusan strategis untuk sistem perawatan kesehatan dengan menawarkan wawasan yang berguna. Ini membantu dalam menciptakan gambaran lengkap tentang klien, pasien, dan profesional. Membuat keputusan berdasarkan data membuka peluang baru untuk meningkatkan kualitas layanan kesehatan. Pentingnya sains data dalam penelitian terbukti bermanfaat untuk perbaikan medis.

Pentingnya Sains Data di Masa Depan

Perusahaan saat ini memiliki akses ke database besar karena mendokumentasikan setiap aspek keterlibatan klien. Sains data memainkan peran penting dalam menganalisis dan mengembangkan model pembelajaran mesin berbasis data ini. Seiring pertumbuhan industri, lebih banyak pekerjaan harus dapat diakses karena analisis membutuhkan lebih banyak ilmuwan data. Masa depan yang cerah diharapkan bagi mereka yang tertarik berkarir di bidang sains data.

Kecerdasan buatan adalah komponen kunci di masa depan sains data. Di masa depan, AI kemungkinan akan menjadi alat paling kuat yang harus digunakan oleh ilmuwan data. Kecerdasan buatan sudah digunakan oleh bisnis untuk membuat keputusan dan menjalankan operasi mereka. Kecerdasan buatan akan digunakan dalam skenario dunia nyata untuk menggunakan solusi otomatis untuk menyaring volume data yang sangat besar untuk menemukan pola yang membantu perusahaan dalam membuat keputusan yang lebih baik.

D. Aplikasi Sains data dalam Kehidupan Nyata

Sains data baru-baru ini telah diimplementasikan hampir setiap industri di dunia. Saat ini, data sangat penting untuk setiap sektor di dunia. Berikut beberapa contoh penggunaan sains data dan penerapannya dalam kehidupan nyata.

1. Analisis Citra Medis

Prosedur seperti mendeteksi neoplasia, kekakuan arteri, dan penggambaran organ menggunakan metode dan kerangka

kerja seperti MapReduce untuk menemukan parameter terbaik pada kategorisasi tekstur paru-paru. Ini menggunakan metode pembelajaran mesin untuk klasifikasi tekstur padat, termasuk analisis wavelet, pembuatan index (indexing) obat berbasis konten, dan support vector machine (SVM).

2. Deteksi Penipuan dan Risiko

Organisasi perbankan telah menguasai seni sains data dan kontrol melalui penggunaan profil konsumen, pembelian historis, dan faktor penting lainnya untuk menghitung kemungkinan risiko dan gagal bayar.

3. Perencanaan Rute Maskapai

Bisnis penerbangan memiliki reputasi untuk mengatasi tantangan. Beberapa penyedia layanan penerbangan berusaha menjaga kondisi kerja dan tingkat hunian mereka tetap tinggi. Maskapai dapat memanfaatkan sains data untuk menetapkan strategi pemasaran seperti program loyalitas pelanggan, mengantisipasi penundaan penerbangan, memutuskan pesawat mana yang akan diperoleh, merencanakan rute dan transit, dan membuat perubahan strategis.

4. Augmented Reality

Aplikasi sains data yang tampaknya paling menjanjikan untuk masa depan adalah augmented reality ini. Sains data dan realitas virtual dapat berinteraksi karena headset VR menggunakan keahlian komputer, algoritma, dan data untuk memberikan pengalaman menonton terbaik. Kemampuan untuk menavigasi rintangan sambil melihat Pokemon di gedung, jalan, dan objek fiktif lainnya. Lokasi Pokemon dan

gym dalam game ini dipilih menggunakan data dari Ingress, aplikasi sebelumnya dari bisnis yang sama.

E. Sains Data di bidang Ritel

Industri ritel tumbuh dengan cepat. Pengecer dapat menganalisis data dan membuat citra psikologis yang unik dari pembelanja untuk mengidentifikasi kebutuhan orang tersebut. Akibatnya, konsumen seringkali terombang-ambing oleh strategi yang digunakan oleh toko. Artikel ini memberikan enam belas kasus penggunaan sains data teratas di ritel dan kebutuhan sains data di sektor ritel agar Anda tetap mengikuti perkembangan saat ini. Kami akan mengeksplorasi bagaimana penggunaan Sains data yang inovatif dalam industri ritel mengubah industri. Kita akan mulai dengan ikhtisar tingkat tinggi kasus penggunaan sains data untuk industri retail sebelum menyelami detailnya.

Mari selami lebih dalam inovasi berbasis data di industri ritel, membahas masing-masing kasus penggunaan sains data ini satu per satu.

1. Optimasi Harga

Tujuh puluh persen pembeli mengatakan biaya rendah sangat penting saat membuat keputusan akhir. Benar, menurut logika harga produsen, harga eceran juga bergantung pada jenis klien yang membeli produk dan jumlah sumber daya yang digunakan dalam manufaktur. Teknologi analisis data memungkinkan kita untuk lebih dekat dari sebelumnya untuk memecahkan masalah ini. Manfaat penting yang dibawa oleh metode pengoptimalan dapat menemukan titik harga yang memuaskan baik bagi konsumen maupun toko. Proses penentuan harga bergantung

tidak hanya pada biaya yang diperlukan untuk memproduksi suatu barang, tetapi juga pada daya beli klien biasa dan kesepakatan yang ditawarkan bisnis lain. Alat yang tersedia untuk analisis data memberikan tingkat kecanggihan baru untuk memecahkan masalah ini. Jika Anda berada di lapangan, beberapa aplikasi pengoptimalan sains data dapat membantu Anda (pengecer) dalam mengetahui kebiasaan belanja pelanggan, jika tidak, Sains data Bootcamp cocok untuk Anda. Beberapa di antaranya adalah:

Partisi Harga

Menetapkan penetapan harga yang tepat merupakan tantangan karena sulit untuk mengetahui apakah konsumen akan menganggapnya adil atau tidak. Kehilangan keuntungan potensial terjadi di kedua skenario. Dalam pandangan pembeli, produk dengan nilai yang sebanding dan beberapa manfaat tergabung dalam kelompok atau segmen. Pelanggan cukup jelas tentang harga minimum, median, dan maksimum yang bersedia mereka keluarkan untuk item di setiap sektor. Keputusan penetapan harga dapat membangun atau menghancurkan perusahaan. Ada kemungkinan bahwa meniru pesaing Anda akan menghasilkan perang harga, sementara mengandalkan tebakan dapat menghasilkan angka penjualan yang buruk.

Poin Harga 'Ajaib'

Titik harga adalah harga jual yang memungkinkan suatu produk mempertahankan tingkat permintaan yang cukup tinggi. Menurut indikator 'nilai' subjektif mereka sendiri,

konsumen menempatkan barang ke dalam segmen atau kelompok harga yang berbeda yang ditandai dengan 'ambang batas' yang jelas. Dalam zona ini, penjualan dimaksimalkan di tengah dan didorong ke titik nol di pinggiran. Karena stok penjual tidak selalu mencerminkan keteraturan ini, kinerja penjualan menjadi kurang ideal. Karena itu, sebagian besar toko akan kehilangan uang jika mereka menaikkan harga barang yang "salah" atau mendiskonnnya.

Tangga Harga

Seperti raja dalam permainan catur, pengecer harus menetapkan harga secara strategis untuk menghilangkan ancaman yang ditimbulkan oleh saingan atau memanfaatkan setiap celah yang disajikan oleh mereka. Mengelola ini menjadi lebih sulit karena lebih banyak kategori muncul dan fluktuasi harga semakin sering terjadi. Kesuksesan peritel bergantung pada kemampuannya untuk "menyelaraskan" dengan pemimpin kategori barang yang dijual. Berapa spektrum harga, jika ada, antara harga "premium" dari pesaing yang lebih mahal dan harga "wajar" dari diskon besar atau harga normal dengan diskon? Di pasar ini, konsumen memiliki sedikit preferensi di antara barang-barang yang memiliki kinerja serupa. Hal yang sama berlaku untuk kisaran harga (maksimum = harga reguler, minimum = harga jual). Penetapan harga yang ideal harus didasarkan pada harga pemimpin yang paling sering terlihat. "Spread" yang paling sesuai dengan pasar adalah kisaran harga yang ditawarkan oleh market leader. Model pengoptimalan harga dalam Sains data menggunakan beberapa algoritma untuk

melakukan analisis waktu nyata terhadap pendapat dan reaksi pelanggan terhadap promosi seperti penjualan, diskon, dan kampanye iklan.

Di perusahaan retail, metode Machine Learning dapat digunakan dengan berbagai cara untuk meningkatkan harga. Pembelajaran Mesin memungkinkan pedagang menciptakan metode yang jauh lebih efektif dan rumit untuk mencapai tujuan bisnis utama mereka. Mengoptimalkan harga memerlukan kumpulan data tertentu, yaitu survei pelanggan, perilaku, demografi, psikografis, riwayat penjualan, dll.

2. Pemasaran yang Dipersonalisasi

- Pengecer mungkin mendapat manfaat dari penerapan rencana pemasaran karena ini adalah metode paling efektif untuk menarik konsumen. Dalam hal cara kerjanya, ini mengumpulkan informasi tentang transaksi keuangan konsumen. Metode ini berpotensi meramalkan keputusan dan preferensi masa depan dalam skala besar. Skenario pemasaran lebih diuntungkan dari informasi orang dalam tentang preferensi konsumen dan pratinjau produk.
- Ini adalah mekanisme yang digunakan pengecer untuk memberikan saran yang disesuaikan kepada pelanggan berdasarkan preferensi dan tindakan mereka sebelumnya. Ini juga memungkinkan toko untuk mengembangkan strategi pemasaran hiper-spesifik yang meningkatkan ROI. Pengecer dapat melakukan semua ini

jika mereka memiliki akses ke data dan dapat mempelajari pelajaran yang bermanfaat darinya.

- Dalam banyak kasus, di sinilah sains data mungkin berguna. Prediksi perilaku konsumen di masa depan dapat dihasilkan dengan menggabungkan informasi dari beberapa sumber data pelanggan.
- Biasanya, algoritma penambahan aturan digunakan untuk menyelesaikan analisis. Itu menyaring informasi dan mengeluarkan apa yang berguna, menggunakan fungsi khusus untuk mengambil data, membaginya menurut kriteria tertentu, dan membuang semua yang tidak penting.
- Tim pemasaran Anda dapat berinteraksi lebih baik dengan ilmuwan data jika mereka memiliki pemahaman yang kuat tentang proses sains data. Regresi, klasifikasi, dan pengelompokan hanyalah beberapa pendekatan sains data yang telah mengubah pemasaran dari disiplin artistik menjadi disiplin ilmu, di mana perilaku pengguna dipelajari dan dimanipulasi secara ilmiah.
- Contoh sains data dalam ritel: salah satu strategi pemasaran yang dirintis Himalaya untuk memenangkan target demografinya adalah menyediakan solusi individual bagi pelanggan. Item yang mereka rekomendasikan didasarkan pada kebutuhan dan keinginan basis konsumen.

3. Deteksi Penipuan

- Kerugian finansial terkait penipuan adalah masalah yang tersebar luas di seluruh dunia. Risiko kredit, kerugian finansial, dan penipuan pengiriman memengaruhi pelanggan dan toko. Dalam beberapa tahun terakhir, ancaman penipuan telah tumbuh dengan digitalisasi dan transaksi online. Pelanggan kehilangan kepercayaan pada toko sebagai akibat dari tindakan tersebut. Di sinilah sektor ritel dapat mengambil manfaat dari penggunaan Sains data.
- Pendekatan berbasis aturan untuk deteksi penipuan tidak efektif karena mereka mengadu domba penjahat dengan sistem penjual. Kami dapat menggunakan sejumlah besar data yang telah diperoleh dari transaksi online untuk membuat prediksi yang akurat tentang penipuan. Pada saat yang sama, metode konvensional kaku dan tidak dapat beradaptasi dengan perubahan keadaan.
- Penerapan Sains data di ritel dapat digunakan untuk menjaga nama perusahaan tetap murni. Ada kesulitan yang berkembang dalam mengidentifikasi transaksi penipuan untuk pedagang. Sebagai akibat dari kemunduran keuangan baru-baru ini, bisnis beralih ke teknologi digital mutakhir seperti pembelajaran mendalam untuk mendapatkan bantuan. Ini memungkinkan mereka untuk memantau semuanya setiap saat dan menangkap perilaku penipuan apa pun segera setelah itu terjadi.