

Teknik Analisis Big Data

BIG DATA – TK13025

Pendahuluan

- Analisis Big Data memadukan pendekatan analisis data statistik tradisional dengan pendekatan komputasi.
- Pengambilan sampel statistik dari suatu populasi sangat ideal bila seluruh kumpulan data tersedia, dan kondisi ini tipikal skenario pemrosesan batch tradisional.
- Namun, Big Data dapat mengalihkan pemrosesan batch ke pemrosesan real-time karena kebutuhan untuk memahami streaming data.
- Dengan data streaming, kumpulan data terakumulasi dari waktu ke waktu, dan data diatur berdasarkan waktu.
- Streaming data menekankan pada pemrosesan tepat waktu, karena hasil analitik memiliki masa simpan.
- Baik itu pengenalan peluang peningkatan penjualan yang muncul dengan sendirinya karena konteks pelanggan saat ini, atau deteksi kondisi anomali dalam lingkungan industri yang memerlukan intervensi untuk melindungi peralatan atau memastikan kualitas produk, waktu adalah esensi, dan kebaruan dari hasil analitik sangat penting.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pendahuluan

- Di bidang yang bergerak cepat seperti Big Data, selalu ada peluang untuk berinovasi.
- Contohnya adalah pertanyaan tentang bagaimana memadukan pendekatan statistik dan komputasi terbaik untuk masalah analitis tertentu.
- Teknik statistik umumnya lebih disukai untuk analisis data eksplorasi, setelah itu teknik komputasi yang memanfaatkan wawasan yang diperoleh dari studi statistik kumpulan data dapat diterapkan.
- Pergeseran dari batch ke real-time menghadirkan tantangan lain karena teknik real-time perlu memanfaatkan algoritma yang efisien secara komputasi.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pendahuluan

- Salah satu tantangan menyangkut cara terbaik untuk menyeimbangkan akurasi hasil analitik terhadap run-time dari algoritma.
- Dari perspektif penyimpanan, solusi penyimpanan multi-tingkat yang memanfaatkan RAM, solid-state drive, dan hard-disk drive akan memberikan fleksibilitas jangka pendek dan kemampuan analitik real-time dengan penyimpanan persisten jangka panjang dan hemat biaya.
- Dalam jangka panjang, sebuah organisasi akan mengoperasikan mesin analisis Big Data pada dua kecepatan: memproses data streaming saat tiba dan melakukan analisis batch data ini saat terakumulasi untuk mencari pola dan tren.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis kuantitatif

- Analisis kuantitatif adalah teknik analisis data yang berfokus pada kuantifikasi pola dan korelasi yang ditemukan dalam data.
- Berdasarkan praktik statistik, teknik ini melibatkan analisis sejumlah besar pengamatan dari kumpulan data.
- Karena ukuran sampelnya besar, hasilnya dapat diterapkan secara umum ke seluruh kumpulan data.
- Hasil analisis kuantitatif bersifat mutlak dan oleh karena itu dapat digunakan untuk perbandingan numerik.
- Misalnya, analisis kuantitatif penjualan es krim mungkin menemukan bahwa kenaikan suhu 5 derajat meningkatkan penjualan es krim sebesar 15%.

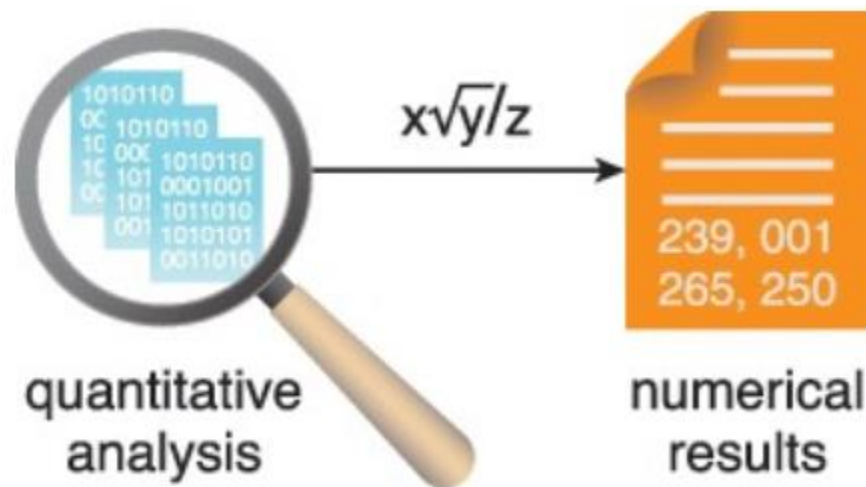


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis kuantitatif



Keluaran analisis kuantitatif bersifat numerik.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis kualitatif

- Analisis kualitatif adalah teknik analisis data yang berfokus pada penggambaran berbagai kualitas data dengan menggunakan kata-kata.
- Analisis ini melibatkan analisis sampel yang lebih kecil secara lebih mendalam dibandingkan dengan analisis data kuantitatif.
- Hasil analisis ini tidak dapat digeneralisasikan ke seluruh kumpulan data karena ukuran sampel yang kecil.
- Hasil analisis kualitatif juga tidak dapat diukur secara numerik atau digunakan untuk perbandingan numerik.
- Misalnya, analisis penjualan es krim dapat mengungkapkan bahwa angka penjualan bulan Mei tidak setinggi bulan Juni.
- Hasil analisis hanya menyatakan bahwa angka tersebut “tidak setinggi”, dan tidak memberikan perbedaan numerik. Keluaran dari analisis kualitatif adalah deskripsi hubungan dengan menggunakan kata-kata.

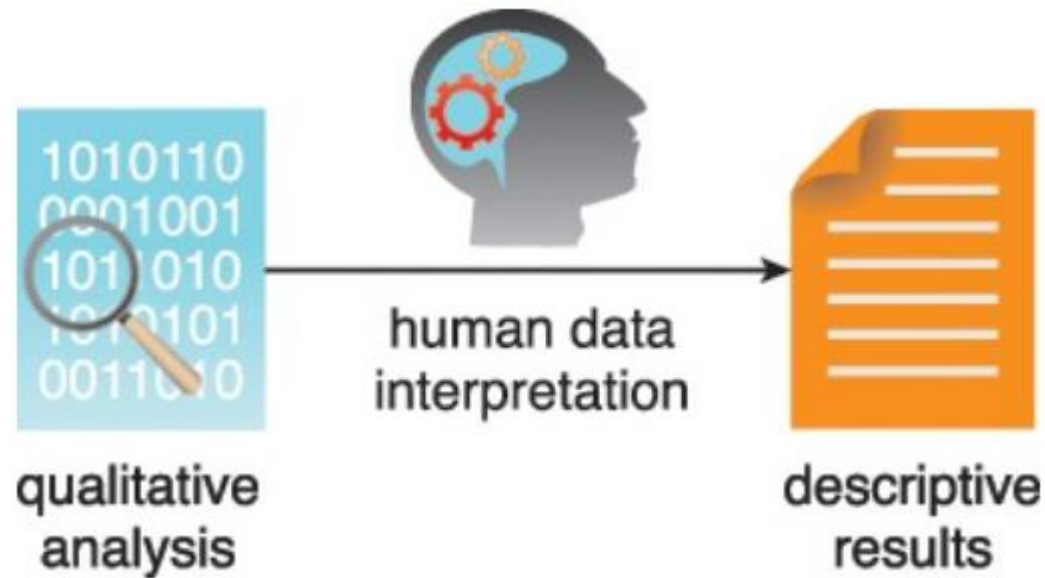


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis kualitatif



Hasil kualitatif bersifat deskriptif dan tidak dapat digeneralisasikan ke seluruh kumpulan data.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
unggul

QS STARS
RATING SYSTEM
2019

AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Data Mining

- Penambangan data, juga dikenal sebagai penemuan data, adalah bentuk khusus dari analisis data yang menargetkan kumpulan data besar.
- Sehubungan dengan analisis Big Data, penambangan data umumnya mengacu pada teknik berbasis perangkat lunak otomatis yang menyaring kumpulan data besar-besaran untuk mengidentifikasi pola dan tren.
- Secara khusus, ini melibatkan penggalian pola tersembunyi atau tidak dikenal dalam data dengan tujuan mengidentifikasi pola yang sebelumnya tidak diketahui.
- Penambangan data membentuk dasar untuk analitik prediktif dan intelijen bisnis (BI).



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis statistik

- Analisis statistik menggunakan metode statistik berdasarkan rumus matematika sebagai sarana untuk menganalisis data.
- Analisis statistik paling sering bersifat kuantitatif, tetapi bisa juga kualitatif.
- Jenis analisis ini biasanya digunakan untuk menggambarkan kumpulan data melalui ringkasan, seperti menyediakan mean, median, atau mode statistik yang terkait dengan kumpulan data.
- Ini juga dapat digunakan untuk menyimpulkan pola dan hubungan dalam kumpulan data, seperti regresi dan korelasi.
- Tipe analisis statistik:
 - ✓ A/B testing
 - ✓ Correlation
 - ✓ Regression



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

A/B Testing

- Pengujian A/B, juga dikenal sebagai pengujian split atau bucket, membandingkan dua versi elemen untuk menentukan versi mana yang lebih unggul berdasarkan metrik yang telah ditentukan sebelumnya.
- Elemennya bisa berupa berbagai hal.
- Misalnya, berupa konten, seperti halaman Web, atau penawaran untuk produk atau layanan, seperti penawaran barang elektronik.
- Versi elemen saat ini disebut versi control (control version), sedangkan versi yang dimodifikasi disebut perawatan (treatment).
- Kedua versi mengalami percobaan secara bersamaan.
- Pengamatan dicatat untuk menentukan versi mana yang lebih berhasil



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

A/B Testing

- Meskipun pengujian A/B dapat diterapkan di hampir semua domain, pengujian ini paling sering digunakan dalam pemasaran.
- Umumnya, tujuannya adalah untuk mengukur perilaku manusia dengan tujuan meningkatkan penjualan.
- Misalnya, untuk menentukan tata letak terbaik untuk iklan es krim di situs Web Perusahaan A, dua versi iklan yang berbeda digunakan.
- Versi A adalah iklan yang sudah ada (kontrol) sedangkan Versi B memiliki tata letak yang sedikit diubah (perlakuan).
- Kedua versi kemudian secara bersamaan ditampilkan ke pengguna yang berbeda:
 - ✓ Versi A ke Grup A
 - ✓ Versi B ke Grup B



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

A/B Testing

- Analisis hasil mengungkapkan bahwa Versi B dari iklan menghasilkan lebih banyak penjualan dibandingkan dengan Versi A.
- Di bidang lain seperti domain ilmiah, tujuannya mungkin hanya untuk mengamati versi mana yang bekerja lebih baik untuk meningkatkan proses atau produk .
- Contoh pertanyaan dapat mencakup:
 - ✓ Apakah obat versi baru lebih baik dari yang lama?
 - ✓ Apakah pelanggan merespon lebih baik terhadap iklan yang dikirim melalui email atau surat pos?
 - ✓ Apakah beranda situs Web yang baru dirancang menghasilkan lebih banyak lalu lintas pengguna?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

A/B Testing

Dua versi email yang berbeda dikirim secara bersamaan sebagai bagian dari kampanye pemasaran untuk melihat versi mana yang mendatangkan lebih banyak calon pelanggan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Correlation

- Korelasi adalah teknik analisis yang digunakan untuk menentukan apakah dua variabel berhubungan satu sama lain.
- Jika kedua variable tersebut ditemukan terkait, langkah selanjutnya adalah menentukan apa hubungan mereka.
- Misalnya, nilai Variabel A meningkat setiap kali nilai Variabel B meningkat.
- Mungkin peneliti lebih tertarik untuk menemukan seberapa dekat Variabel A dan B terkait, yang berarti ingin menganalisis sejauh mana Variabel B meningkat dalam kaitannya dengan peningkatan Variabel A.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Correlation

- Penggunaan korelasi membantu mengembangkan pemahaman tentang kumpulan data dan menemukan hubungan yang dapat membantu menjelaskan suatu fenomena.
- Oleh karena itu korelasi biasanya digunakan untuk penambangan data di mana identifikasi hubungan antara variabel dalam kumpulan data mengarah pada penemuan pola dan anomali.
- Korelasi dapat mengungkapkan sifat kumpulan data atau penyebab suatu fenomena.
- Ketika dua variabel dianggap berkorelasi, mereka disejajarkan berdasarkan hubungan linier.
- Artinya ketika satu variabel berubah, variabel lainnya juga berubah secara proporsional dan konstan.
- Korelasi dinyatakan sebagai angka desimal antara -1 hingga +1, yang dikenal sebagai koefisien korelasi.
- Derajat hubungan berubah dari kuat menjadi lemah ketika bergerak dari 1 ke 0 atau +1 ke 0.

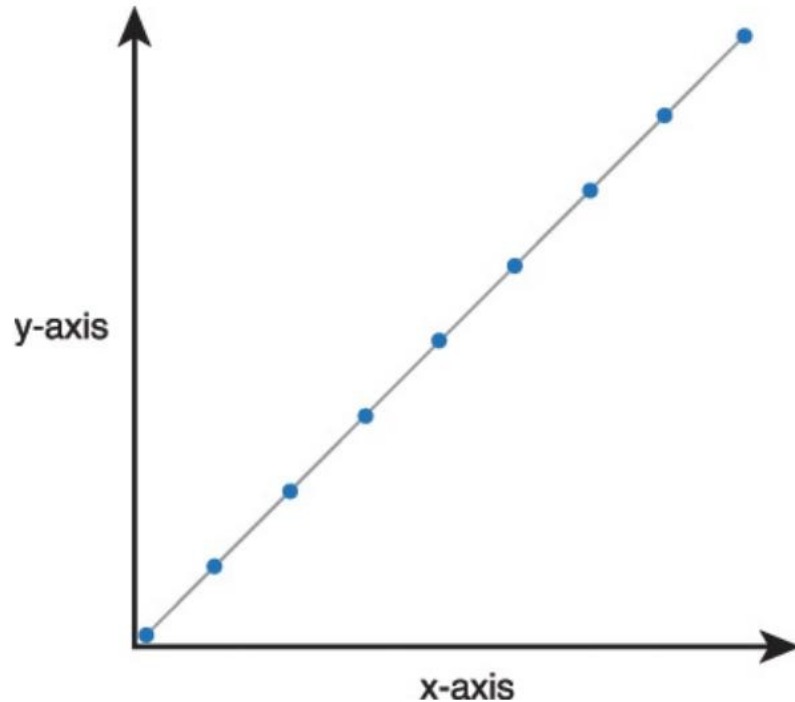


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Correlation



Gambar di samping menunjukkan korelasi +1, yang menunjukkan bahwa ada hubungan positif yang kuat antara kedua variabel.

Ketika satu variabel meningkat, yang lain juga meningkat dan sebaliknya.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
Linggi

QS STARS
RATING SYSTEM
2019

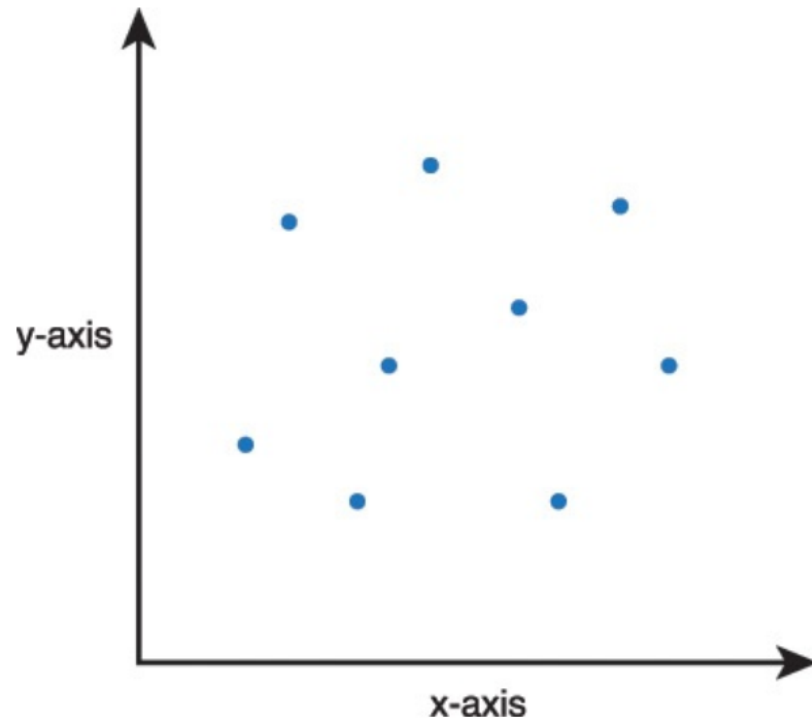
AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Correlation



Gambar di samping menunjukkan korelasi 0, yang menunjukkan bahwa tidak ada hubungan sama sekali antara kedua variabel.

Ketika satu variabel meningkat, yang lain mungkin tetap sama, atau meningkat atau menurun secara sewenang-wenang.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
lingkup

QS STARS
RATING SYSTEM
2019

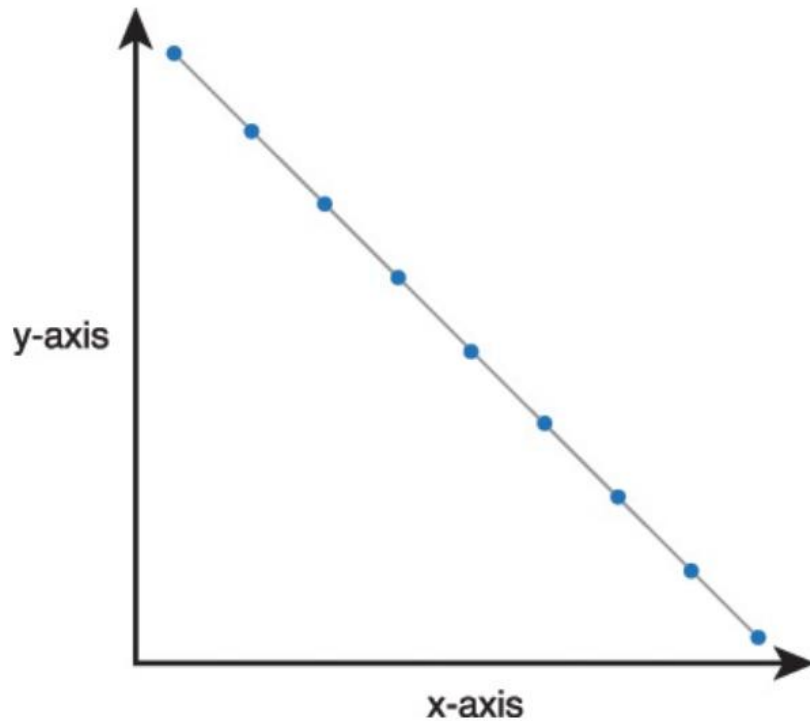
AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Correlation



Gambar di samping memiliki korelasi -1 menunjukkan bahwa ada hubungan negatif yang kuat antara kedua variabel.

Ketika satu variabel meningkat, yang lain menurun dan sebaliknya.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
Linggi

QS STARS
RATING SYSTEM
2019

AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Correlation

- Misalnya, manajer percaya bahwa toko es krim perlu menyimpan lebih banyak es krim untuk hari yang panas, tetapi tidak tahu berapa banyak yang harus disimpan.
- Untuk menentukan apakah ada hubungan antara suhu dan penjualan es krim, analis pertama-tama menerapkan korelasi dengan jumlah es krim yang terjual dan pembacaan suhu yang tercatat.
- Nilai $+0,75$ menunjukkan bahwa terdapat hubungan yang kuat antara keduanya. Hubungan ini menunjukkan bahwa dengan meningkatnya suhu, lebih banyak es krim yang dijual.
- Contoh pertanyaan lebih lanjut yang ditangani dengan korelasi dapat mencakup:
 - ✓ Apakah jarak dari laut mempengaruhi suhu kota?
 - ✓ Apakah siswa yang berprestasi baik di sekolah dasar berprestasi sama baiknya di sekolah menengah?
 - ✓ Sejauh mana obesitas terkait dengan makan berlebihan?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Regression

- Teknik analisis regresi mengeksplorasi bagaimana variabel dependen terkait dengan variabel independen dalam kumpulan data.
- Sebagai skenario sampel, regresi dapat membantu menentukan jenis hubungan yang ada antara suhu, variabel bebas, dan hasil panen, variabel terikat.
- Menerapkan teknik regresi membantu menentukan bagaimana nilai variabel dependen berubah dalam kaitannya dengan perubahan nilai variabel independen.
- Ketika variabel independen meningkat, misalnya, apakah variabel dependen juga meningkat?
- Jika ya, apakah kenaikan tersebut dalam proporsi linier atau nonlinier?



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
linggi

QS STARS
RATING SYSTEM
2019

AMBA
ACCREDITED

IAABE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Regression

- Misalnya, untuk menentukan berapa banyak stok ekstra yang harus dimiliki setiap toko es krim, analis menerapkan regresi dengan memasukkan nilai pembacaan suhu.
- Nilai-nilai ini didasarkan pada ramalan cuaca sebagai variabel independen dan jumlah es krim yang terjual sebagai variabel dependen.
- Apa yang para analis temukan adalah bahwa 15% dari stok tambahan diperlukan untuk setiap kenaikan suhu 5 derajat.
- Lebih dari satu variabel bebas dapat diuji secara bersamaan.
- Namun, dalam kasus seperti itu, hanya satu variabel independen yang dapat berubah, sementara yang lain tetap konstan.
- Regresi dapat membantu memungkinkan pemahaman yang lebih baik tentang apa itu fenomena dan mengapa itu terjadi.
- Ini juga dapat digunakan untuk membuat prediksi tentang nilai-nilai variabel dependen.



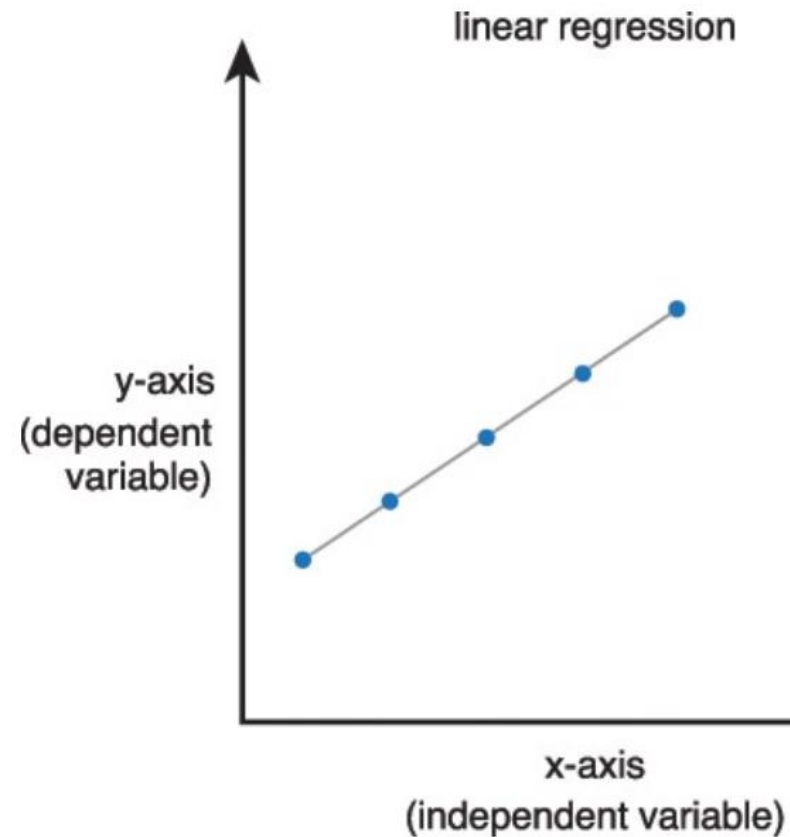
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

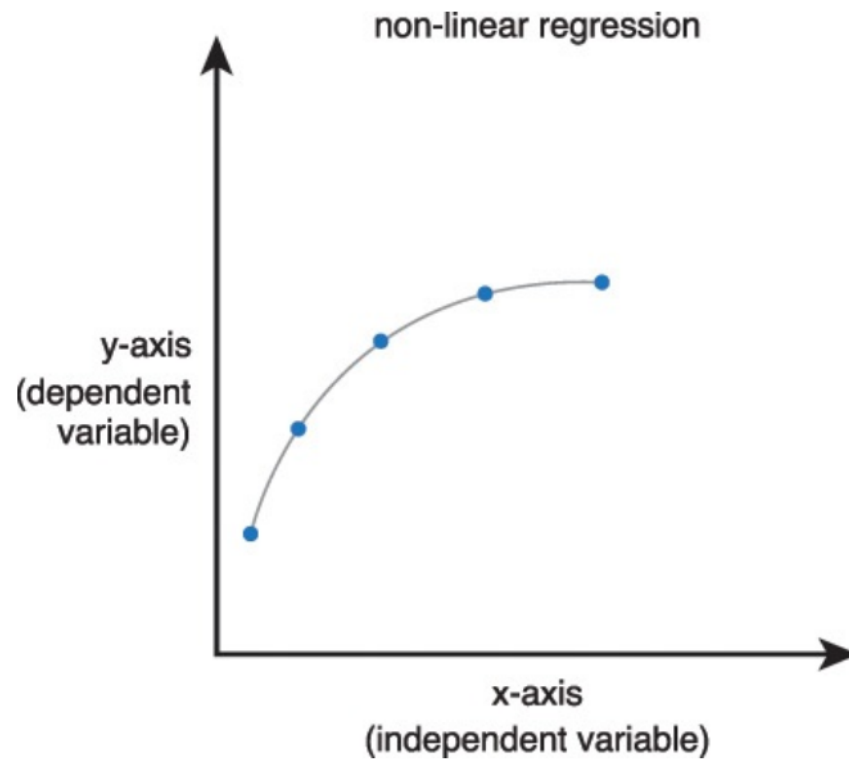
Regression

Regresi linier mewakili tingkat perubahan yang konstan



Regression

Regresi non-linier mewakili tingkat perubahan yang bervariasi



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
linggih

QS STARS
RATING SYSTEM
2019

AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Regression

- Contoh pertanyaan dapat mencakup:
 - ✓ Berapa suhu di kota yang berjarak 250 mil dari laut?
 - ✓ Berapa nilai siswa yang belajar di sekolah menengah berdasarkan nilai sekolah dasar mereka?
 - ✓ Berapa peluang seseorang menjadi gemuk berdasarkan jumlah asupan makanannya?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Regression

- Regresi dan korelasi memiliki sejumlah perbedaan penting. Korelasi tidak berarti sebab-akibat.
- Perubahan nilai salah satu variabel mungkin tidak bertanggung jawab atas perubahan nilai variabel kedua, meskipun keduanya dapat berubah pada tingkat yang sama.
- Ini dapat terjadi karena variabel ketiga yang tidak diketahui, yang dikenal sebagai faktor pengganggu.
- Korelasi mengasumsikan bahwa kedua variabel adalah independen.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Lingkar

QS STARS
RATING SYSTEM
2019

AMBA
ACCREDITED

EFMD
EQUIS

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Regression

- Regresi, di sisi lain, berlaku untuk variabel yang sebelumnya telah diidentifikasi sebagai variabel dependen dan independen dan menyiratkan bahwa ada tingkat sebab-akibat antara variabel.
- Penyebabnya bisa langsung atau tidak langsung.
- Dalam Big Data, korelasi pertama-tama dapat diterapkan untuk mengetahui apakah ada hubungan.
- Regresi kemudian dapat diterapkan untuk mengeksplorasi lebih lanjut hubungan dan memprediksi nilai-nilai variabel dependen, berdasarkan nilai-nilai variabel independen yang diketahui.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Machine Learning

- Manusia pandai melihat pola dan hubungan dalam data.
- Sayangnya, kami tidak dapat memproses data dalam jumlah besar dengan sangat cepat.
- Mesin, di sisi lain, sangat mahir dalam memproses data dalam jumlah besar dengan cepat, tetapi hanya jika mereka tahu caranya.
- Jika pengetahuan manusia dapat digabungkan dengan kecepatan pemrosesan mesin, mesin akan dapat memproses data dalam jumlah besar tanpa memerlukan banyak campur tangan manusia.
- Ini adalah konsep dasar pembelajaran mesin.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Machine Learning

- Pembelajaran mesin dan hubungannya dengan penambangan data dieksplorasi melalui cakupan jenis teknik pembelajaran mesin berikut:
 - Classification
 - Clustering
 - Outlier Detection
 - Filtering



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Classification (Supervised Machine Learning)

- Klasifikasi adalah teknik pembelajaran terawasi dimana data diklasifikasikan ke dalam kategori relevan yang dipelajari sebelumnya.
- Klasifikasi terdiri dari dua langkah:
 1. Sistem diumpankan data pelatihan yang sudah dikategorikan atau diberi label, sehingga dapat mengembangkan pemahaman tentang kategori yang berbeda.
 2. Sistem diumpankan data yang tidak diketahui tetapi serupa untuk klasifikasi dan berdasarkan pemahaman yang dikembangkan dari data pelatihan, algoritma akan mengklasifikasikan data yang tidak berlabel.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Classification

- Aplikasi umum dari teknik ini adalah untuk menyaring spam email. Perhatikan bahwa klasifikasi dapat dilakukan untuk dua atau lebih kategori.
- Dalam proses klasifikasi yang disederhanakan, mesin diberi data berlabel selama pelatihan yang membangun pemahamannya tentang klasifikasi.
- Mesin kemudian diberi masukan data tidak berlabel.



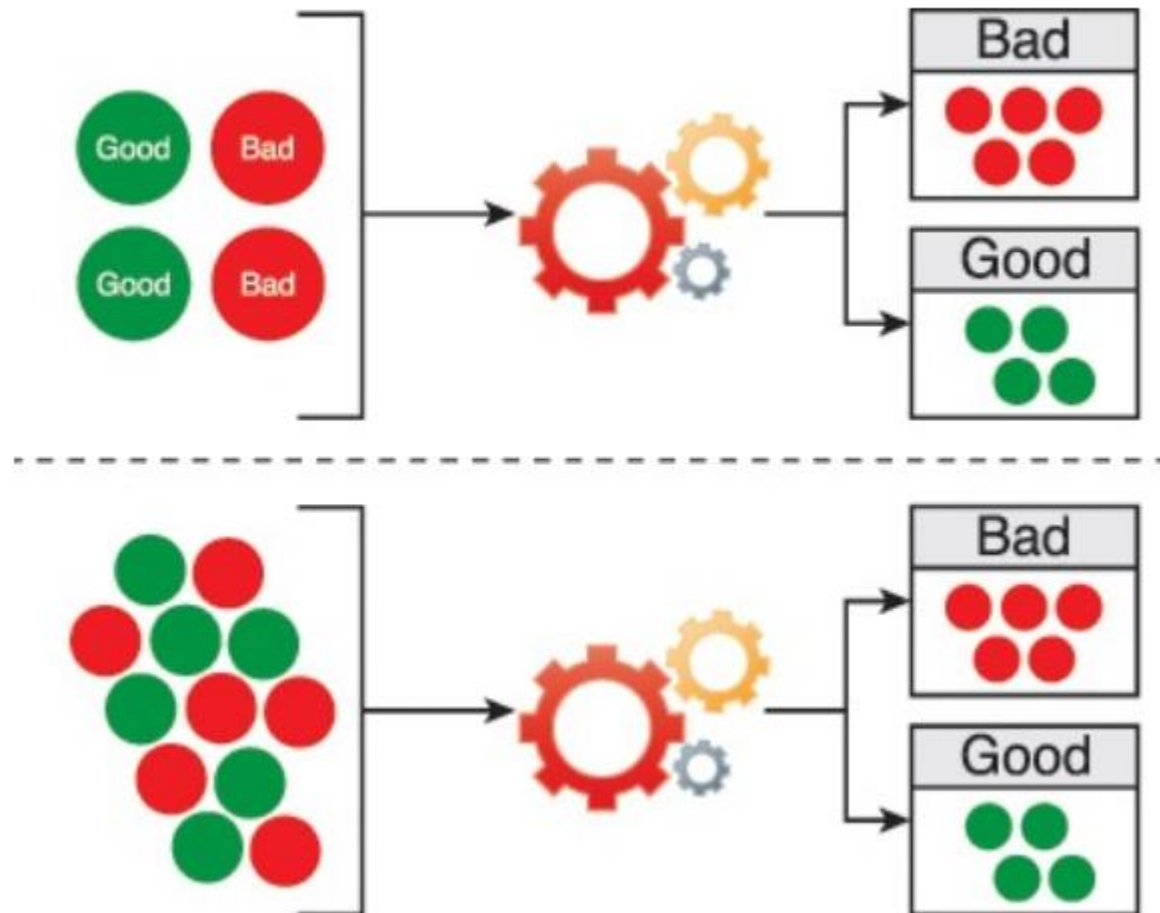
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Classification

Pembelajaran mesin dapat digunakan untuk mengklasifikasikan kumpulan data secara otomatis.



Classification

- Misalnya, bank ingin mengetahui pelanggan mana yang cenderung gagal membayar pinjaman.
- Berdasarkan data historis, kumpulan data pelatihan dikompilasi yang berisi contoh berlabel pelanggan yang pernah atau belum pernah gagal bayar.
- Data pelatihan ini diumpankan ke algoritme klasifikasi yang digunakan untuk mengembangkan pemahaman tentang pelanggan "baik" dan "buruk".
- Terakhir, data pelanggan baru yang tidak ditandai diumpankan untuk mengetahui apakah pelanggan tertentu termasuk dalam kategori default.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Classification

- Contoh pertanyaan dapat mencakup:
 - Haruskah aplikasi kartu kredit pemohon diterima atau ditolak berdasarkan aplikasi lain yang diterima atau ditolak?
 - Apakah tomat termasuk buah atau sayuran berdasarkan contoh buah dan sayuran yang diketahui?
 - Apakah hasil tes kesehatan pasien menunjukkan risiko serangan jantung?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Clustering (Unsupervised Machine Learning)

- Clustering adalah teknik pembelajaran tanpa pengawasan dimana data dibagi menjadi beberapa kelompok yang berbeda sehingga data dalam setiap kelompok memiliki sifat yang sama.
- Tidak ada pembelajaran sebelumnya dari kategori yang diperlukan. Sebaliknya, kategori secara implisit dihasilkan berdasarkan pengelompokan data.
- Bagaimana data dikelompokkan tergantung pada jenis algoritma yang digunakan.
- Setiap algoritma menggunakan teknik yang berbeda untuk mengidentifikasi cluster.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Clustering

- Clustering umumnya digunakan dalam penambangan data untuk mendapatkan pemahaman tentang properti dari kumpulan data yang diberikan.
- Setelah mengembangkan pemahaman ini, klasifikasi dapat digunakan untuk membuat prediksi yang lebih baik tentang data yang serupa tetapi baru atau tidak terlihat.
- Pengelompokan dapat diterapkan pada kategorisasi dokumen yang tidak dikenal dan kampanye pemasaran yang dipersonalisasi dengan mengelompokkan pelanggan dengan perilaku serupa.
- Grafik sebar memberikan representasi visual dari cluster

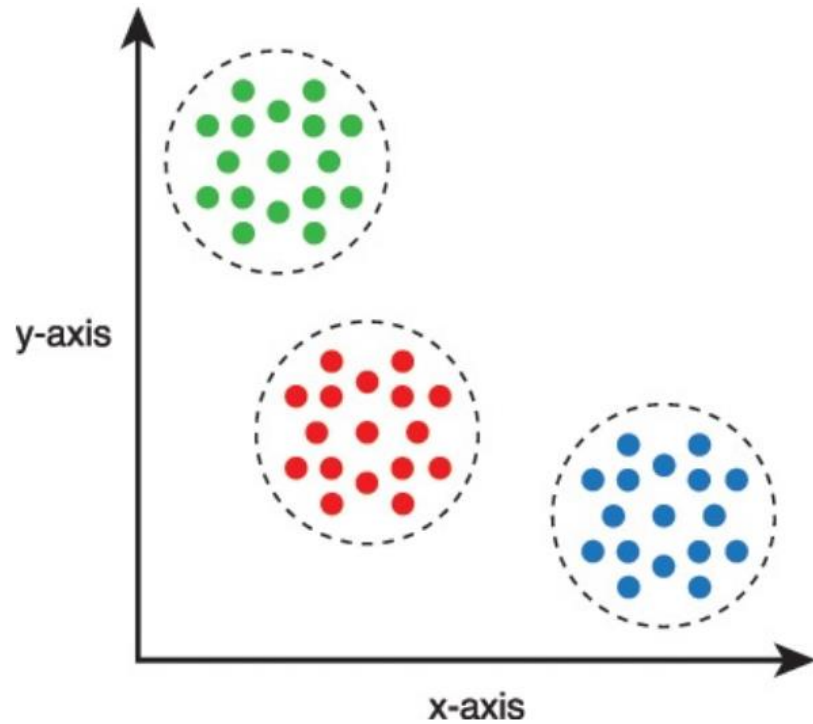


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Clustering



Grafik scatter merangkum hasil pengelompokan.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
unggul

QS STARS
RATING SYSTEM
2019

AMBA
AACSB
EQUIS

IABEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Clustering

- Misalnya, bank ingin memperkenalkan nasabah yang sudah ada ke berbagai produk keuangan baru berdasarkan profil nasabah yang tercatat.
- Analis mengkategorikan pelanggan ke dalam beberapa kelompok menggunakan pengelompokan.
- Setiap kelompok kemudian diperkenalkan dengan satu atau lebih produk keuangan yang paling sesuai dengan karakteristik profil kelompok secara keseluruhan.
- Contoh pertanyaan dapat mencakup:
 - ✓ Berapa banyak spesies pohon yang berbeda berdasarkan kesamaan antar pohon?
 - ✓ Berapa banyak kelompok pelanggan yang ada berdasarkan riwayat pembelian yang serupa?
 - ✓ Apa saja kelompok virus berdasarkan karakteristiknya?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Outlier Detection (Deteksi Pencilan)

- Deteksi outlier adalah proses menemukan data yang secara signifikan berbeda atau tidak konsisten dengan sisa data dalam kumpulan data yang diberikan.
- Teknik pembelajaran mesin ini digunakan untuk mengidentifikasi anomali, kelainan dan penyimpangan yang dapat menguntungkan, seperti peluang, atau tidak menguntungkan, seperti risiko.
- Deteksi outlier berkaitan erat dengan konsep klasifikasi dan clustering, meskipun algoritmanya berfokus pada pencarian nilai abnormal.
- Ini dapat didasarkan pada pembelajaran yang diawasi (supervised learning) atau tidak diawasi (unsupervised learning).
- Aplikasi untuk deteksi outlier termasuk deteksi penipuan, diagnosis medis, analisis data jaringan dan analisis data sensor.
- Grafik scatter secara visual menyoroti titik data yang outlier

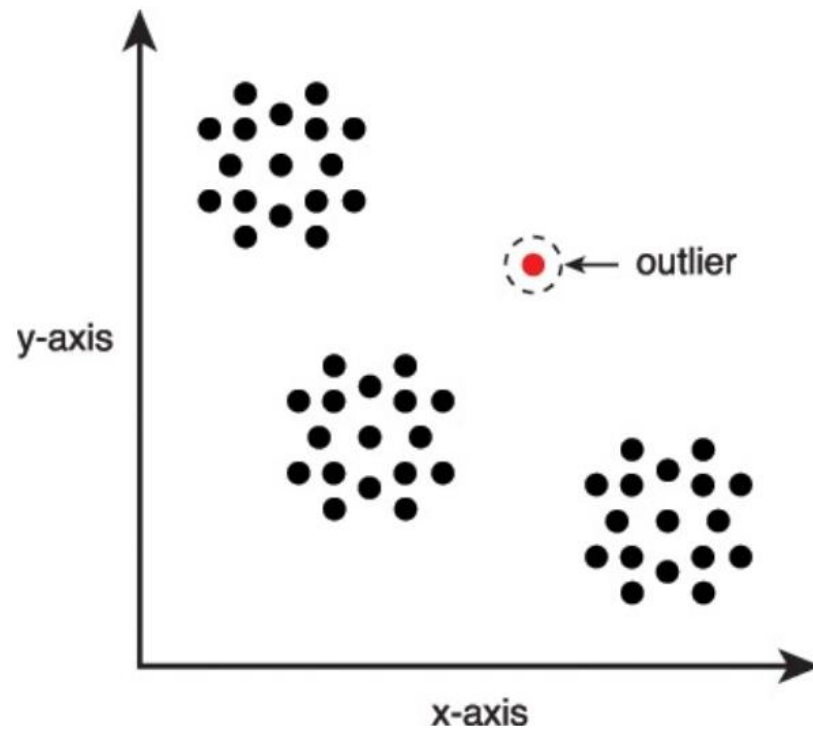


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Outlier Detection



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
Linggi

QS STARS
RATING SYSTEM
2019

AMBA
AACSB
EFMD

IAFEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Outlier Detection

- Misalnya, untuk mengetahui apakah suatu transaksi berpotensi penipuan atau tidak, tim TI bank membangun sistem dengan teknik deteksi outlier yang berbasis pembelajaran terawasi.
- Satu set transaksi penipuan yang diketahui pertama-tama dimasukkan ke dalam algoritma deteksi outlier.
- Setelah melatih sistem, transaksi yang tidak diketahui kemudian dimasukkan ke dalam algoritma deteksi outlier untuk memprediksi apakah transaksi tersebut curang atau tidak.
- Contoh pertanyaan dapat mencakup:
 - Apakah seorang atlet menggunakan obat peningkat performa?
 - Apakah ada buah dan sayuran yang salah diidentifikasi dalam dataset pelatihan yang digunakan untuk tugas klasifikasi?
 - Apakah ada jenis virus tertentu yang tidak merespon pengobatan?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Filtering

- Pemfilteran adalah proses otomatis untuk menemukan item yang relevan dari kumpulan item.
- Item dapat difilter baik berdasarkan perilaku pengguna sendiri atau dengan mencocokkan perilaku beberapa pengguna.
- Penyaringan umumnya diterapkan melalui dua pendekatan berikut:
 - ✓ collaborative filtering
 - ✓ content-based filtering



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Filtering (Penyaringan)

- Media umum di mana penyaringan diimplementasikan adalah melalui penggunaan sistem rekomendasi.
- Pemfilteran kolaboratif adalah teknik pemfilteran item berdasarkan kolaborasi, atau penggabungan, dari perilaku masa lalu pengguna dengan perilaku orang lain.
- Perilaku masa lalu pengguna target, termasuk suka, peringkat, riwayat pembelian, dan lainnya, dikolaborasikan dengan perilaku pengguna serupa. Berdasarkan kesamaan perilaku pengguna, item difilter untuk pengguna target.
- Fitering kolaboratif semata-mata didasarkan pada kesamaan antara perilaku pengguna.
- Filtering membutuhkan sejumlah besar data perilaku pengguna untuk memfilter item secara akurat.
- Filtering adalah contoh penerapan hukum bilangan besar.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Filtering

- Pemfilteran berbasis konten adalah teknik pemfilteran item yang berfokus pada kesamaan antara pengguna dan item.
- Profil pengguna dibuat berdasarkan perilaku pengguna sebelumnya, misalnya, suka, peringkat, dan riwayat pembelian mereka.
- Kesamaan yang diidentifikasi antara profil pengguna dan atribut berbagai item menyebabkan item disaring untuk pengguna.
- Berlawanan dengan pemfilteran kolaboratif, pemfilteran berbasis konten hanya didedikasikan untuk preferensi pengguna individu dan tidak memerlukan data tentang pengguna lain.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Filtering

- Sistem pemberi rekomendasi memprediksi preferensi pengguna dan menghasilkan saran untuk pengguna yang sesuai.
- Saran biasanya berkaitan dengan merekomendasikan item, seperti film, buku, halaman Web dan orang-orang.
- Sistem pemberi rekomendasi biasanya menggunakan pemfilteran kolaboratif atau pemfilteran berbasis konten untuk menghasilkan saran.
- Ini juga dapat didasarkan pada gabungan pemfilteran kolaboratif dan pemfilteran berbasis konten untuk menyempurnakan akurasi dan efektivitas saran yang dihasilkan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Filtering

- Misalnya untuk mewujudkan peluang cross-selling, bank membangun sistem rekomendasi yang menggunakan content-based filtering.
- Berdasarkan kecocokan yang ditemukan antara produk keuangan yang dibeli oleh pelanggan dan properti produk keuangan serupa, sistem pemberi rekomendasi mengotomatiskan saran untuk produk keuangan potensial yang mungkin juga diminati pelanggan.
- Contoh pertanyaan dapat mencakup:
 - ✓ Bagaimana hanya artikel berita yang diminati pengguna yang dapat ditampilkan?
 - ✓ Destinasi liburan mana yang dapat direkomendasikan berdasarkan riwayat perjalanan seorang wisatawan?
 - ✓ Pengguna baru mana yang dapat disarankan sebagai teman berdasarkan profil seseorang saat ini?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Semantic Analysis

- Sebuah fragmen teks atau data ucapan dapat membawa makna yang berbeda dalam konteks yang berbeda, sedangkan kalimat lengkap dapat mempertahankan maknanya, bahkan jika disusun dengan cara yang berbeda.
- Agar mesin dapat mengekstrak informasi berharga, data teks dan ucapan perlu dipahami oleh mesin dengan cara yang sama seperti yang dilakukan manusia.
- Analisis semantik mewakili praktik untuk mengekstraksi informasi yang bermakna dari data tekstual dan ucapan.
- Bagian ini menjelaskan jenis analisis semantik berikut:
 - ✓ Natural Language Processing
 - ✓ Text Analytics
 - ✓ Sentiment Analysis



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Natural Language Processing (NLP)

- NLP adalah kemampuan komputer untuk memahami ucapan dan teks manusia sebagaimana dipahami secara alami oleh manusia.
- Hal ini memungkinkan komputer untuk melakukan berbagai tugas yang berguna, seperti pencarian teks lengkap.
- Misalnya, untuk meningkatkan kualitas layanan pelanggan, perusahaan es krim menggunakan pemrosesan bahasa alami untuk mentranskripsikan panggilan pelanggan ke dalam data tekstual yang kemudian ditambang untuk alasan ketidakpuasan pelanggan yang sering berulang.
- Alih-alih mengkodekan aturan pembelajaran yang diperlukan, pembelajaran mesin yang diawasi atau tidak diawasi diterapkan untuk mengembangkan pemahaman komputer tentang bahasa alami.
- Secara umum, semakin banyak data pembelajaran yang dimiliki komputer, semakin tepat ia dapat menguraikan teks dan ucapan manusia.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Natural Language Processing

- NLP mencakup pengenalan teks dan ucapan.
- Untuk pengenalan ucapan, sistem mencoba memahami ucapan dan kemudian melakukan tindakan, seperti menyalin teks.
- Contoh pertanyaan dapat mencakup:
 - Bagaimana sistem pertukaran telepon otomatis yang dapat mengenali ekstensi departemen yang benar seperti yang didiktekan secara lisan oleh penelepon dapat dikembangkan?
 - Bagaimana kesalahan tata bahasa dapat diidentifikasi secara otomatis?
 - Bagaimana sebuah sistem yang dapat dengan benar memahami berbagai aksen bahasa Inggris dapat dirancang?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Text Analytics

- Teks tidak terstruktur umumnya jauh lebih sulit untuk dianalisis dan dicari dibandingkan dengan teks terstruktur.
- Analisis teks adalah analisis khusus teks melalui penerapan penambahan data, pembelajaran mesin, dan teknik pemrosesan bahasa alami untuk mengekstrak nilai dari teks yang tidak terstruktur.
- Analisis teks pada dasarnya memberikan kemampuan untuk menemukan teks daripada hanya mencarinya.
- Wawasan yang berguna dari data berbasis teks dapat diperoleh dengan membantu bisnis mengembangkan pemahaman tentang informasi yang terkandung dalam kumpulan teks yang besar.
- Sebagai kelanjutan dari contoh NLP sebelumnya, data tekstual yang ditranskripsi dianalisis lebih lanjut menggunakan analisis teks untuk mengekstrak informasi yang berarti tentang alasan umum di balik ketidakpuasan pelanggan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Text Analytics

- Prinsip dasar analitik teks adalah mengubah teks yang tidak terstruktur menjadi data yang dapat dicari dan dianalisis.
- Karena jumlah dokumen digital, email, posting media sosial, dan file log meningkat, bisnis memiliki kebutuhan yang meningkat untuk memanfaatkan nilai apa pun yang dapat diekstraksi dari bentuk data semi-terstruktur dan tidak terstruktur ini.
- Hanya menganalisis data operasional (terstruktur) dapat menyebabkan bisnis kehilangan peluang penghematan biaya atau ekspansi bisnis, terutama yang berfokus pada pelanggan.
- Aplikasi termasuk klasifikasi dan pencarian dokumen, serta membangun pandangan 360 derajat dari pelanggan dengan mengekstrak informasi dari sistem CRM.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Text Analytics

- Analisis teks umumnya melibatkan dua langkah:
 1. Parsing teks dalam dokumen untuk mengekstrak:
 - Entitas Bernama – orang, kelompok, tempat, perusahaan
 - Entitas Berbasis Pola – nomor jaminan sosial, kode pos
 - Konsep – representasi abstrak dari suatu entitas
 - Fakta – hubungan antar entitas
 2. Kategorisasi dokumen menggunakan entitas dan fakta yang diekstraksi ini.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Text Analytics

- Informasi yang diekstraksi dapat digunakan untuk melakukan pencarian konteks-spesifik pada entitas, berdasarkan jenis hubungan yang ada antara entitas.
- Entitas diekstraksi dari file teks menggunakan aturan semantik dan terstruktur sehingga dapat dicari.



documents

Name	URL	City	Country	Phone No.
≡	≡	≡	≡	≡

Text Analytics

- Contoh pertanyaan dapat mencakup:
 - Bagaimana cara mengkategorikan situs Web berdasarkan konten halaman Web mereka?
 - Bagaimana saya bisa menemukan buku yang berisi konten yang relevan dengan topik yang saya pelajari?
 - Bagaimana saya bisa mengidentifikasi kontrak yang berisi informasi rahasia perusahaan?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Sentiment Analysis

- Analisis sentimen adalah bentuk khusus dari analisis teks yang berfokus pada penentuan bias atau emosi individu.
- Bentuk analisis ini menentukan sikap pengarang teks dengan menganalisis teks dalam konteks bahasa alami.
- Analisis sentimen tidak hanya memberikan informasi tentang bagaimana perasaan individu, tetapi juga intensitas perasaan mereka. Informasi ini kemudian dapat diintegrasikan ke dalam proses pengambilan keputusan.
- Aplikasi umum untuk analisis sentimen termasuk mengidentifikasi kepuasan atau ketidakpuasan pelanggan lebih awal, mengukur keberhasilan atau kegagalan produk, dan menemukan tren baru.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Sentiment Analysis

- Misalnya, sebuah perusahaan es krim ingin mengetahui rasa es krim mana yang paling disukai oleh anak-anak.
- Data penjualan saja tidak memberikan informasi ini karena anak-anak yang mengkonsumsi es krim belum tentu pembeli es krim.
- Analisis sentimen diterapkan pada arsip umpan balik pelanggan yang tertinggal di situs web perusahaan es krim untuk mengekstrak informasi secara khusus mengenai preferensi anak-anak untuk rasa es krim tertentu daripada rasa lainnya.
- Contoh pertanyaan dapat mencakup:
 - Bagaimana reaksi pelanggan terhadap kemasan produk yang baru dapat diukur?
 - Kontestan mana yang kemungkinan besar akan menjadi pemenang kontes menyanyi?
 - Dapatkah churn pelanggan diukur dengan komentar media sosial?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis Visual

- Analisis visual adalah bentuk analisis data yang melibatkan representasi grafis dari data untuk mengaktifkan atau meningkatkan persepsi visualnya.
- Berdasarkan premis bahwa manusia dapat memahami dan menarik kesimpulan dari grafik lebih cepat daripada dari teks, analisis visual bertindak sebagai alat penemuan di bidang Big Data.
- Tujuannya adalah menggunakan representasi grafis untuk mengembangkan pemahaman yang lebih dalam tentang data yang dianalisis.
- Secara khusus, ini membantu mengidentifikasi dan menyoroti pola, korelasi, dan anomali tersembunyi.
- Analisis visual juga berhubungan langsung dengan analisis data eksplorasi karena mendorong perumusan pertanyaan dari sudut yang berbeda.
- Bagian ini menjelaskan jenis analisis visual berikut: Heat Maps, Time Series Plots, Network Graphs, dan Spatial Data Mapping



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Heat Maps

- Peta panas adalah teknik analisis visual yang efektif untuk mengekspresikan pola, komposisi data melalui hubungan sebagian-keseluruhan dan distribusi geografis data.
- Mereka juga memfasilitasi identifikasi bidang minat dan penemuan nilai ekstrim (tinggi/rendah) dalam kumpulan data.
- Misalnya, untuk mengidentifikasi wilayah penjualan es krim teratas dan terburuk, data penjualan es krim diplot menggunakan peta panas.
- Hijau digunakan untuk menyorot wilayah dengan kinerja terbaik, sedangkan merah digunakan untuk menyorot wilayah dengan kinerja terburuk.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Heat Maps

- Heat map adalah representasi visual, kode warna dari nilai data.
- Setiap nilai diberi warna sesuai dengan jenisnya atau kisarannya.
- Misalnya, Heat map dapat menetapkan nilai 0–3 untuk warna merah, 4–6 untuk kuning, dan 7–10 untuk hijau.
- Heat map bisa dalam bentuk bagan atau peta.
- Bagan (chart) mewakili matriks nilai di mana setiap sel diberi kode warna sesuai dengan nilainya.
- Itu juga dapat mewakili nilai hierarkis dengan menggunakan kotak bersarang berkode warna



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Heat Maps

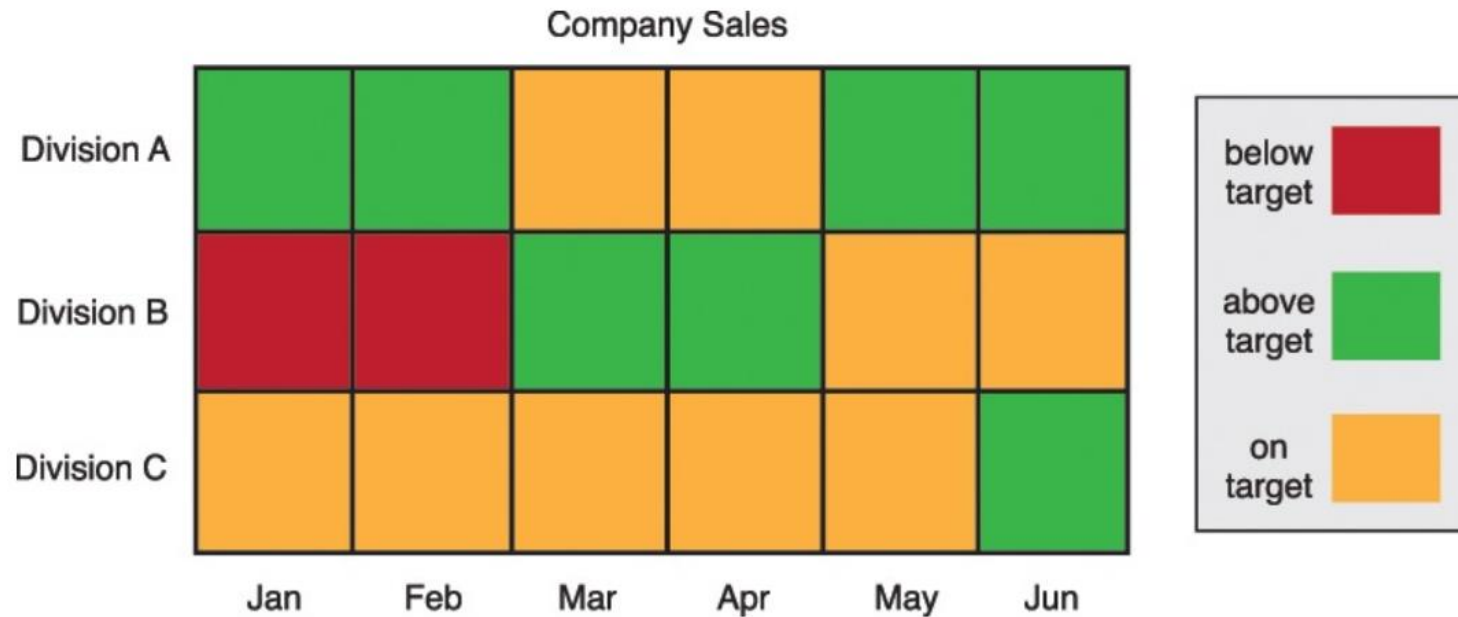


Chart heat map menggambarkan penjualan tiga divisi dalam sebuah perusahaan selama periode enam bulan.



UNTAR
Universitas Tarumanagara

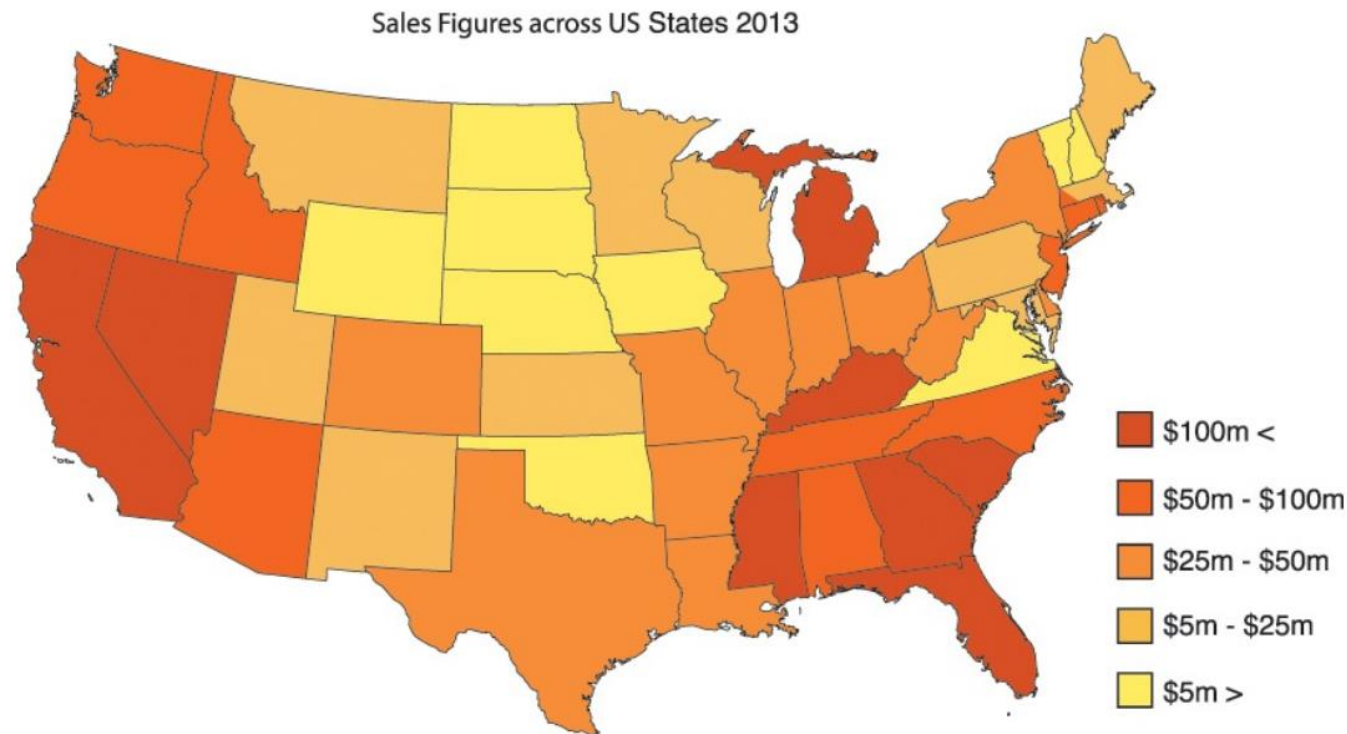


UNTAR untuk INDONESIA

Heat Maps

Peta mewakili ukuran geografis di mana berbagai wilayah diberi kode warna atau diarsir menurut tema tertentu.

Peta dapat ditumpangkan oleh lapisan yang terdiri dari kumpulan titik-titik berwarna/berarsir yang berkaitan dengan berbagai wilayah, atau bentuk berwarna/berarsir yang mewakili berbagai wilayah



Heat Maps

- Contoh pertanyaan dapat mencakup:
 - Bagaimana saya dapat mengidentifikasi secara visual pola apa pun yang terkait dengan emisi karbon di sejumlah besar kota di seluruh dunia?
 - Bagaimana saya bisa melihat jika ada pola jenis kanker yang berbeda dalam kaitannya dengan etnis yang berbeda?
 - Bagaimana cara menganalisis pemain sepak bola berdasarkan kekuatan dan kelemahannya?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Time Series Plots

- Plot deret waktu memungkinkan analisis data yang direkam selama interval waktu periodik.
- Jenis analisis ini menggunakan deret waktu, yang merupakan kumpulan nilai berurutan yang direkam dalam interval waktu yang teratur.
- Contohnya adalah time series yang memuat angka penjualan yang dicatat setiap akhir bulan.
- Analisis deret waktu membantu mengungkap pola dalam data yang bergantung pada waktu.
- Setelah diidentifikasi, pola tersebut dapat diekstrapolasi untuk prediksi masa depan.
- Misalnya, untuk mengidentifikasi pola penjualan musiman, angka penjualan es krim bulanan diplot sebagai rangkaian waktu, yang selanjutnya membantu memperkirakan angka penjualan untuk musim berikutnya.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Time Series Plots

- Analisis deret waktu biasanya digunakan untuk peramalan dengan mengidentifikasi tren jangka panjang, pola periodik musiman, dan variasi jangka pendek yang tidak teratur dalam kumpulan data.
- Tidak seperti jenis analisis lainnya, analisis deret waktu (time series) selalu menyertakan waktu sebagai variabel pembanding, dan data yang dikumpulkan selalu bergantung pada waktu.
- Plot deret waktu umumnya dinyatakan menggunakan diagram garis, dengan waktu diplot pada sumbu x dan nilai data yang direkam diplot pada sumbu y



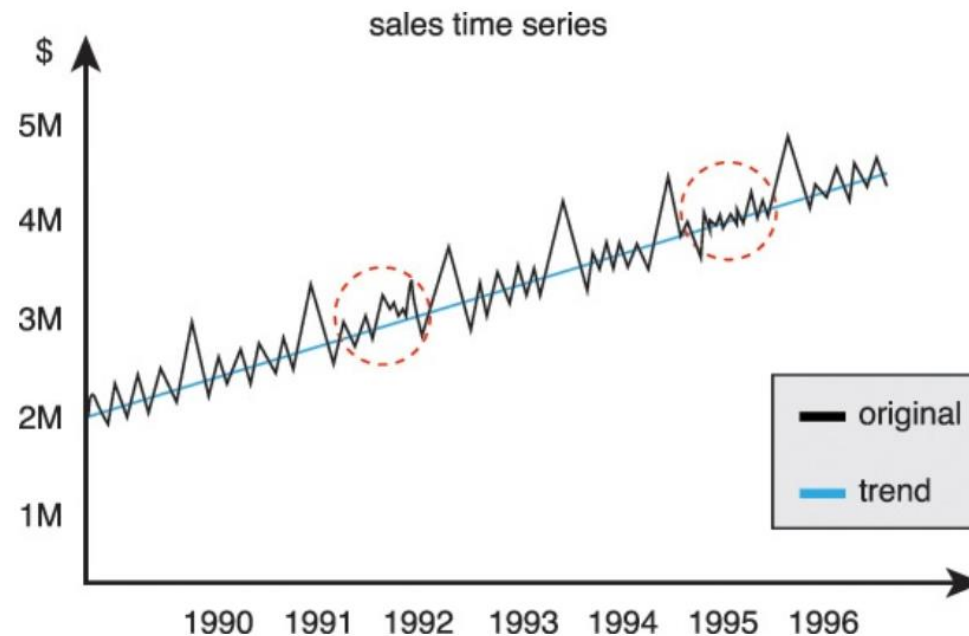
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Time Series Plots

Line chart menggambarkan deret waktu penjualan dari tahun 1990 hingga 1996.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
unggul

QS STARS
RATING SYSTEM
2010

GLAN
UNAL

IABEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Time Series Plots

- Puncak dengan jarak yang sama menjelang akhir tahun menunjukkan pola periodik musiman, misalnya penjualan Natal.
- Lingkaran merah putus-putus mewakili variasi tidak teratur jangka pendek.
- Garis biru menunjukkan tren naik, menunjukkan peningkatan penjualan.
- Contoh pertanyaan dapat mencakup:
 - Berapa banyak hasil yang diharapkan petani berdasarkan data hasil historis?
 - Berapa perkiraan pertambahan penduduk dalam 5 tahun mendatang?
 - Apakah penurunan penjualan saat ini hanya terjadi sekali atau terjadi secara teratur?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Network Graphs

- Dalam konteks analisis visual, grafik jaringan menggambarkan kumpulan entitas yang saling berhubungan.
- Entitas dapat berupa orang, grup, atau objek domain bisnis lainnya seperti produk.
- Entitas dapat terhubung satu sama lain secara langsung atau tidak langsung.
- Beberapa koneksi mungkin hanya satu arah, sehingga traversal dalam arah sebaliknya tidak dimungkinkan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Network Graphs

- Analisis jaringan adalah teknik yang berfokus pada menganalisis hubungan antara entitas dalam jaringan.
- Ini melibatkan plot entitas sebagai node dan koneksi sebagai tepi antara node.
- Ada variasi khusus dari analisis jaringan, termasuk:
 - ✓ pengoptimalan rute
 - ✓ analisis jaringan sosial
 - ✓ prediksi penyebaran, seperti penyebaran penyakit menular



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Network Graphs

- Berikut ini adalah contoh sederhana berdasarkan penjualan es krim untuk aplikasi analisis jaringan untuk optimasi rute.
- Beberapa manajer toko es krim mengeluh tentang waktu yang dibutuhkan truk pengiriman untuk berkendara antara gudang pusat dan toko di daerah terpencil.
- Pada hari-hari yang lebih panas, es krim yang dikirim dari gudang pusat ke toko-toko terpencil mencair dan tidak dapat dijual.
- Analisis jaringan digunakan untuk menemukan rute terpendek antara gudang pusat dan toko jarak jauh untuk meminimalkan durasi pengiriman.



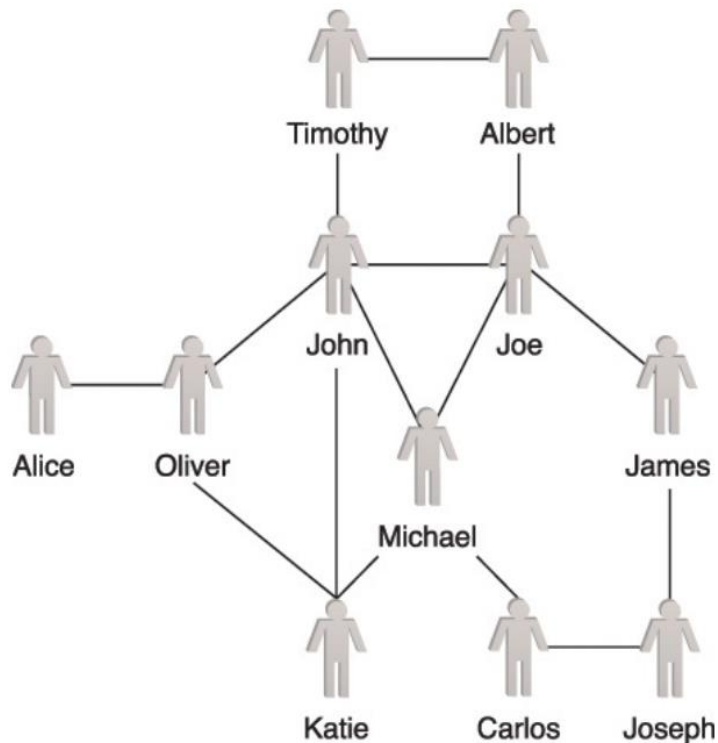
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Network Graphs

Contoh grafik jejaring sosial:



John memiliki banyak teman, sedangkan Alice hanya memiliki satu teman.

Hasil analisis jaringan sosial mengungkapkan bahwa Alice kemungkinan besar akan berteman dengan John dan Katie, karena mereka memiliki teman yang sama bernama Oliver.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Network Graphs

- Contoh pertanyaan dapat mencakup:
 - Bagaimana saya bisa mengidentifikasi pemberi pengaruh dalam kelompok besar pengguna?
 - Apakah dua individu terkait satu sama lain melalui rantai leluhur yang panjang?
 - Bagaimana saya bisa mengidentifikasi pola interaksi di antara sejumlah besar interaksi protein-keprotein?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Spatial Data Mapping

- Data spasial atau geospasial umumnya digunakan untuk mengidentifikasi lokasi geografis entitas individu yang kemudian dapat dipetakan.
- Analisis data spasial difokuskan pada analisis data berbasis lokasi untuk menemukan hubungan dan pola geografis yang berbeda antar entitas.
- Data spasial dimanipulasi melalui Sistem Informasi Geografis (SIG) yang memplot data spasial pada peta umumnya menggunakan koordinat bujur dan lintangnya.
- SIG menyediakan alat yang memungkinkan eksplorasi interaktif data spasial, misalnya mengukur jarak antara dua titik, atau mendefinisikan wilayah di sekitar titik sebagai lingkaran dengan radius berbasis jarak yang ditentukan.
- Dengan ketersediaan data berbasis lokasi yang terus meningkat, seperti data sensor dan media sosial, data spasial dapat dianalisis untuk mendapatkan wawasan lokasi.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Spatial Data Mapping

- Misalnya, sebagai bagian dari ekspansi perusahaan, lebih banyak toko es krim direncanakan untuk dibuka.
- Ada persyaratan bahwa dua toko tidak boleh berada dalam jarak 5 kilometer satu sama lain untuk mencegah toko bersaing satu sama lain.
- Data spasial digunakan untuk memplot lokasi toko yang ada dan kemudian mengidentifikasi lokasi yang optimal untuk toko baru setidaknya 5 kilometer dari toko yang ada.
- Aplikasi analisis data spasial meliputi operasi dan optimasi logistik, ilmu lingkungan dan perencanaan infrastruktur.
- Data yang digunakan sebagai input untuk analisis data spasial dapat berisi lokasi yang tepat, seperti garis bujur dan garis lintang, atau informasi yang diperlukan untuk menghitung lokasi, seperti kode pos atau alamat IP.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Spatial Data Mapping

- Selanjutnya, analisis data spasial dapat digunakan untuk menentukan jumlah entitas yang berada dalam radius tertentu dari entitas lain.
- Misalnya, supermarket menggunakan analisis spasial untuk pemasaran yang ditargetkan.
- Lokasi diambil dari pesan media sosial pengguna, dan penawaran yang dipersonalisasi dikirimkan secara real-time berdasarkan kedekatan pengguna dengan toko.
- Contoh pertanyaan dapat mencakup:
 - Berapa banyak rumah yang akan terkena dampak proyek pelebaran jalan?
 - Berapa jarak yang harus ditempuh pelanggan untuk pergi ke supermarket?
 - Dimana konsentrasi tinggi dan rendah dari mineral tertentu berdasarkan pembacaan yang diambil dari sejumlah lokasi sampel dalam suatu area?



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Spatial Data Mapping

Analisis data spasial dapat digunakan untuk pemasaran yang ditargetkan.

