

Data Science

Pokok Bahasan : Business Understanding

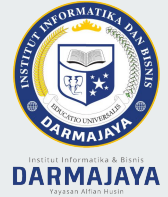
Dosen Pengampu : Hary Sabita, ST., MTI

07 Juni 2022

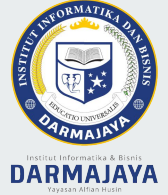


Outline :

1. Menentukan Objektif Business
2. Menentukan Tujuan Teknis Data Science
3. Membuat Rencana Proyek Data Science
4. Studi Kasus



Next - Apa tujuan mempelajari ini ??



Tujuan yang akan dicapai :

1. Mampu menentukan objektif AI didalam bisnis
2. Mampu menentukan tujuan teknis dan pondasi data science
3. Mampu membuat rencana proyek data science

Next - Apa tujuan objektif AI ??

Tujuan Objektif AI dalam Bisnis

- AI meniru keputusan dan tindakan kecerdasan manusia
- Contoh
 - Autonomous Vehicle
 - Stock Price Prediction
- **Otomatisasi** berfokus pada penyederhanaan tugas instruktif yang berulang
- Contoh
 - Menyalakan lampu menggunakan smartphone setiap jam 18:00
 - Alat otomatis pemberi makan hewan peliharaan

AI Vs Automation

Next - Bagaimana mengidentifikasinya?

Identifikasi Problem Machine Learning

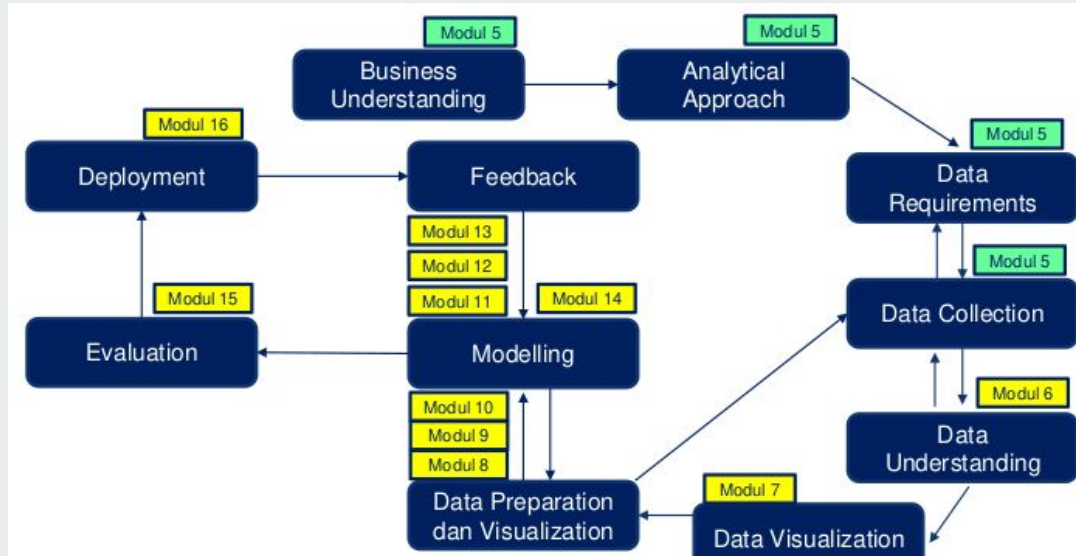
Terdapat beberapa **sistem pengolahan data**, atau **sistem pembelajaran** yang memerlukan beberapa tahapan pengolahan yang masing-masing memiliki **keterikatan** satu sama lain untuk menghasilkan **output**.

Tahapan-tahapan Business Understanding di AI



Next - Apa tujuan teknis dan pondasinya

Tujuan Teknis dan Pondasi Data Science



Next - Rencana proyek data science

Rencana Proyek Data Science

- Setiap proyek dimulai dengan pemahaman bisnis.
- Proyek Data Science merupakan proyek Bisnis, sehingga harus selalu berorientasi pada pencapaian hasil yang berfokus pada bisnis
- Proyek Data Science harus memiliki visi global yang selaras dengan strategi bisnis.
- Sponsor bisnis membutuhkan solusi analitik.
- Tahapan Proyek Data Science



Next - Bagaimana menganalisa nya

Analytical Approach

- Business Understanding example :
 - Autonomous Vehicle Problem: Kendaraan autonomous yang belum aman
- Clear question:
 - Bagaimana cara agar kendaraan autonomous dapat secara aman beroperasi?
- Measurable outcomes:
 - 40% mengurangi kecelakaan lalu lintas yang diakibatkan oleh human error

Analytical Approach

- Setelah masalah bisnis dinyatakan dengan jelas, data scientist dapat menentukan pendekatan analitik untuk memecahkan masalah
- Tahap ini mencakup pengungkapan masalah dalam konteks teknik statistik dan pembelajaran mesin.
- Tahap ini dapat membantu organisasi dalam mengidentifikasi solusi yang paling tepat

- Pendekatan analitik yang dipilih akan menentukan kebutuhan data.
- Secara khusus, metode analitik yang akan digunakan memerlukan konten, format, dan representasi data tertentu, hal ini juga dipengaruhi oleh problem domain yang dipilih.



Analytic Approach (Example)

"Memprediksi pendapatan perusahaan di kuartal berikutnya?"

- Regresi

"Apakah pasien ini menderita kanker A, kanker B, atau mereka sehat?"

- Klasifikasi

"Apakah ada kelompok pengguna yang tampaknya memiliki kesamaan satu sama lain?"

- Clustering

"Bagaimana saya bisa menargetkan diskon untuk pelanggan tertentu?"

- Rekomendasi/Personalisasi

Next - Mengumpulkan data?

Data Collection

- Pada tahap pengumpulan data awal, data scientist dapat melakukan
 - 1. mengidentifikasi dan mengumpulkan sumber data yang tersedia—
 - terstruktur
 - tidak terstruktur
 - semi-terstruktur—yang relevan dengan domain masalah.

Gathering Data

- Data dapat dikumpulkan melalui beberapa sumber, seperti:
 - 1. Data internal perusahaan (excel, database internal, dll)
 - 2. Web API, Web scraping
 - 3. Kumpulan data melalui data publik
 - 4. Kumpulan data melalui data terbuka

Studi Kasus (Stock Prediction)

- Pasar saham dikenal sebagai tempat di mana orang dapat berinvestasi dan mendapatkan margin dari investasi tersebut dengan memprediksi harga saham.
- Meskipun sangat sulit untuk memprediksi harga secara tepat.
- Bagaimana kita bisa membuat mesin memiliki kecerdasan untuk dapat memprediksi nilai saham seperti halnya broker di pasar saham?

System Flow



- Pada system ini dibutuhkan source data dalam berbagai tipe sumber, tabular, sql, csv, atau API
- Machine learning dibutuhkan untuk mempelajari nilai numerik data-data tersebut
- Hasil prediksi harga saham pada waktu yang ditentukan dapat ditampilkan oleh system berdasarkan data-data lampau

Problems

- Bagaimana cara agar data dapat di consume oleh algoritma machine learning ?
 - Solution : Data Cleansing, Data Pre-Processing, Data Understanding
- Algoritma / solusi machine learning yang perlu diterapkan agar kita bisa mendapatkan prediksi harga saham yang akurat
 - Solution : Pemilihan algoritma machine learning dengan akurasi terbaik dengan metrik yang sesuai

Komponen-Komponen Pendukung

Data Cleansing, Data Pre-processing, Data Understanding

Pilihan Machine Learning Algorithm sebagai predictor (Regression)

- Auto Regressive Moving Average
- Linear Regression
- K Nearest Neighbor
- Recurrent Neural Network (Long Short Term Memory)

Solution : Data cleansing, Data Pre-processing, Data Understanding

Modul 6

Modul 7

Modul 8

- Biasanya suatu data stock memiliki kolom sebagai berikut

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-10-08	208.00	222.25	206.85	216.00	215.15	4642146.0	10062.83
1	2018-10-05	217.00	218.60	205.90	210.25	209.20	3519515.0	7407.06

- Date, High, low, open, close, Trade Quantity, Turnover
- Date' sebagai fitur indeks. '
- High menunjukkan nilai tertinggi hari itu.
- Low menunjukkan yang terendah.
- Open adalah Harga pembukaan
- Close adalah penutupan untuk Tanggal tersebut.
- Trade Quantity adalah jumlah Saham perusahaan yang diperdagangkan pada tanggal tersebut.
- Turnover adalah omset perusahaan tertentu pada tanggal tersebut

Data Understanding (Visualization)

Modul 6

Modul 7

Modul 8



Next - Algoritma

Regression Algorithm

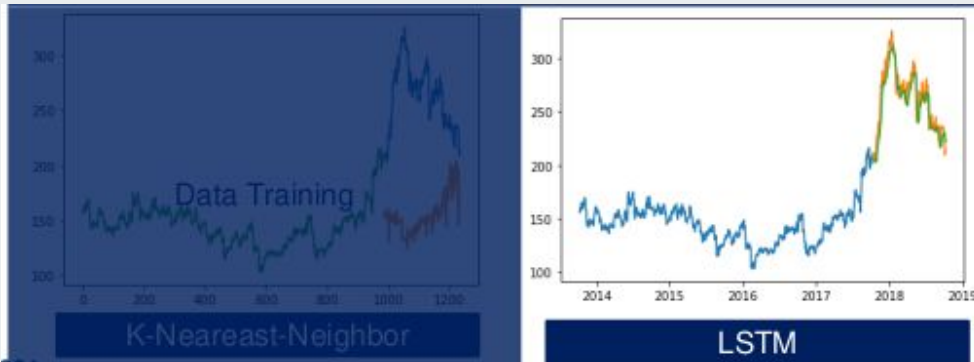
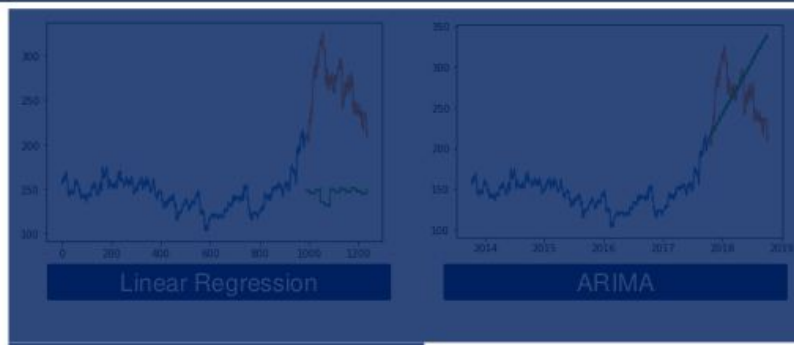
Modul 12

Modul 15

- Auto Regressive Moving Average
- Linear, Lasso, Ridge Regression
- Decision Tree Regression, Adaboost Regression,
- Recurrent Neural Network

Error Measurements

- Regression Metrics :
 - MSE (Mean Square Error)
 - RMSE (Root Mean Square Error)
 - R2 Score



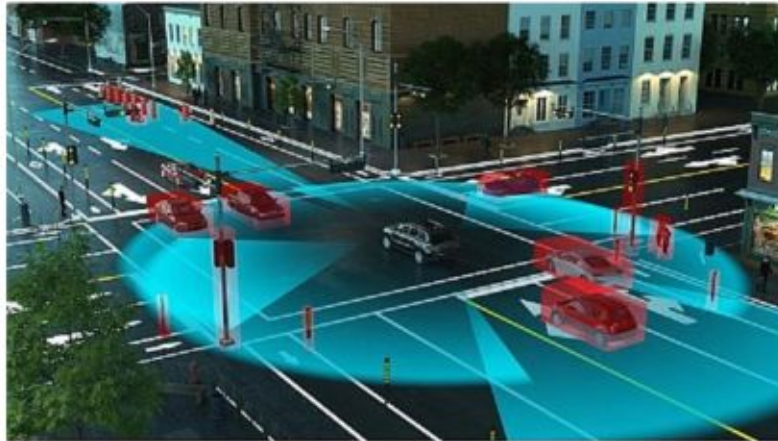
Dari ke empat algoritma berikut ini, algoritma mana yang memiliki akurasi terbaik ?

Hint : Algoritma tersebut Memiliki error paling minimal antara predicted value dan real value

Next - Studi Kasus

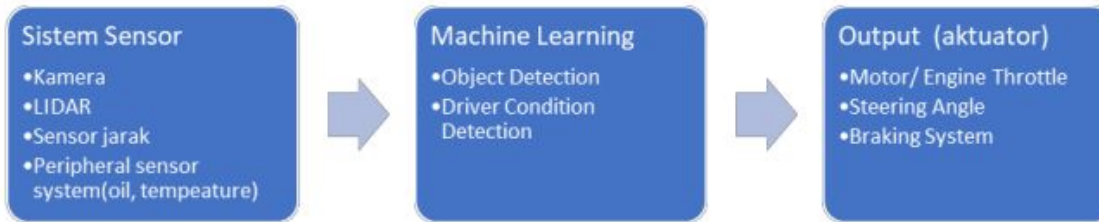
Studi Kasus (Autonomous Vehicle)

- Bagaimana menyediakan kendaraan tanpa pengemudi yang aman, ekonomis, dan praktis ?



Next - System Flow

System Flow



- Pada system ini dibutuhkan system sensor sebagai indra untuk mendeteksi lingkungan disekitar kendaraan.
- Machine learning dibutuhkan untuk menangkap hasil pengindraan sensor dan menentukan actuator untuk bekerja.

Komponen-Komponen Pendukung

Deteksi dan Klasifikasi Objek

Pengenalan Pengemudi

Sensor Fusion

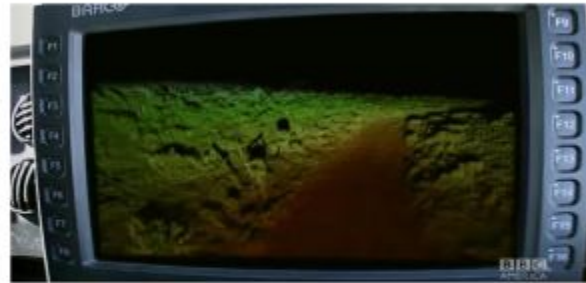
Aktuator

Privacy

Next - Deteksi Objek

Deteksi dan Klasifikasi Objek

- Pembelajaran mesin digunakan untuk membantu decision making bagi pengemudi, seperti persepsi dan pemahaman visual disekitar kendaraan.
- Hal ini melibatkan penggunaan sistem berbasis kamera untuk mendeteksi dan mengklasifikasikan objek, tetapi ada juga yang mengembangkan menggunakan LiDAR dan radar



Gambar 1 Driverless Truck (LiDAR Visualization),

Ref: <https://www.youtube.com/watch?v=1plPontZk>

Deteksi dan Klasifikasi Objek (problems)

- Salah satu masalah terbesar dalam problem mobil otonom adalah salah klasifikasi objek atau keterlambatan dalam klasifikasi objek dan memberikan reaksi.
- Data yang dikumpulkan oleh berbagai sensor kendaraan dikumpulkan dan kemudian diinterpretasikan oleh sistem kendaraan.



Deteksi dan Klasifikasi Objek (possible solution)

- Melalui pelatihan model ML yang lebih baik dan data yang lebih umum, sistem dapat meningkatkan persepsi dan mengidentifikasi objek dengan akurasi yang lebih baik.
- Melatih sistem dengan memberikan input yang lebih bervariasi pada parameter utama untuk dapat digunakan membuat keputusan.

Pengenalan Pengemudi (problems)

- Machine learning dapat mengenali pola, sehingga dapat digunakan di dalam kendaraan untuk memantau pengemudi.
- Misalnya, pengenalan wajah dapat digunakan untuk mengidentifikasi pengemudi dan memverifikasi apakah dia memiliki hak tertentu, misalnya izin untuk menyalakan mobil, yang dapat membantu mencegah penggunaan dan pencurian.
- Pengenalan kondisi pengemudi (mengantuk atau siap berkendara)



Ref : <https://marketresearch.biz/report/driver-drowsiness-detection-system-market/>

Pengenalan Pengemudi (possible solutions)

- Pengenalan wajah (face recognition) system perlu diimplementasikan untuk mengenali wajah pengendara.
- Selain itu untuk mendeteksi tingkat kelelahan kita memerlukan definisi dari face landmarks untuk membantu menentukan status kelelahan dari pengemudi.
- Dengan menggunakan fitur-fitur yang berada pada **face landmarks**, kita bisa melihat lebih detail kondisi mata, kondisi mulut, hidung, dan dahi dari pengemudi.



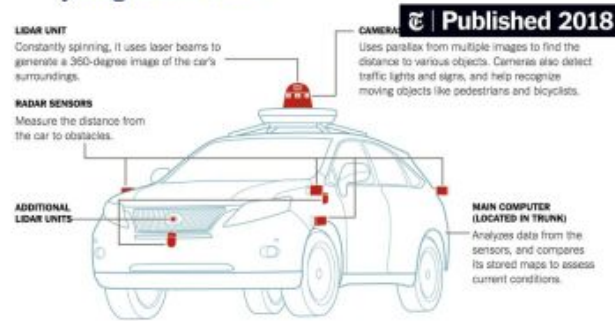
Ref :

<https://www.pyimagesearch.com/2017/05/08/drowsiness-detection-opencv/>

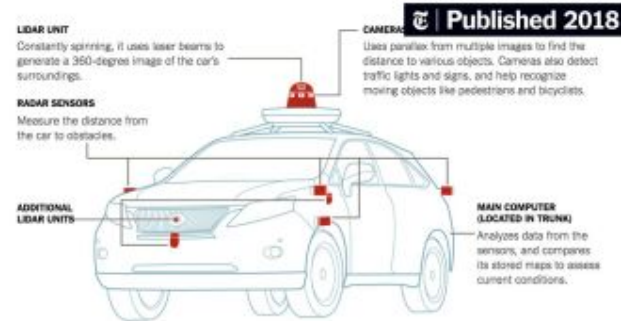
Sensors Fusion (problems & Solution)

- Kamera dapat memberikan tekstur dan pengenalan warna yang baik.
 - Tetapi kamera rentan terhadap kondisi yang dapat mengurangi jarak pandang dan ketajaman visual, seperti halnya mata manusia.
- Radar dan Lidar, sinyal dan titik digunakan untuk membuat pengelompokan yang lebih baik, untuk memberikan gambar objek 3D yang lebih akurat
 - Tidak bisa memberikan tekstur seperti kamera

- Beberapa produsen mobil otonom menggabungkan output sensor kamera dan radar untuk memperoleh keakuratan yang lebih baik.



- Pengenalan wajah (face recognition) system perlu diimplementasikan untuk mengenali wajah pengendara.
- Selain itu untuk mendeteksi tingkat kelelahan kita memerlukan definisi dari face landmarks untuk membantu menentukan status kelelahan dari pengemudi.
- Dengan menggunakan fitur-fitur yang berada pada **face landmarks**, kita bisa melihat lebih detail kondisi mata, kondisi mulut, hidung, dan dahi dari pengemudi.



Ref :

<https://www.nytimes.com/2018/03/19/technology/how-driveless-cars-work.html>

Pemeliharaan

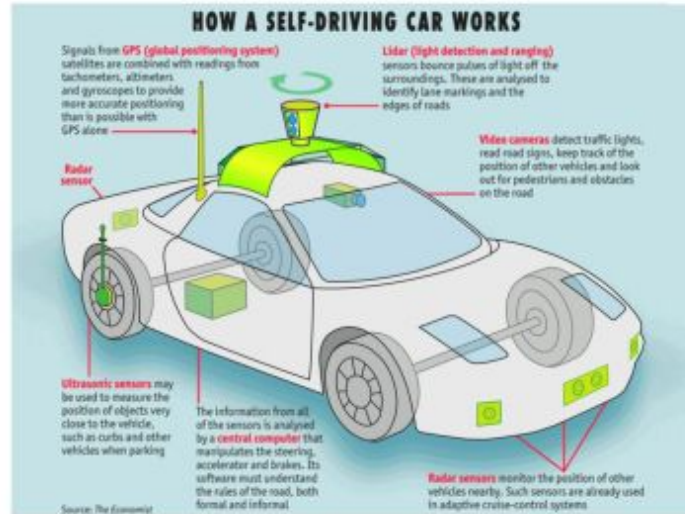
- Powertrain dan engine kendaraan biasanya menghasilkan serangkaian data disetiap waktu.
- Pembelajaran mesin dapat diterapkan pada data ini untuk meningkatkan kontrol mesin
- Dengan ML, sistem dapat beradaptasi dari waktu ke waktu terhadap penuaan kendaraan dan merespons perubahan yang terjadi.

Pemeliharaan (possible solution)

- ML dapat diterapkan ke data yang diambil oleh perangkat yang sudah terpasang. Data variabel seperti suhu, pengisian baterai, tekanan oli, dan level cairan pendingin dikirim ke system.
- Data tersebut dianalisis dan menghasilkan gambaran kinerja dan kondisi kendaraan secara keseluruhan. Indikator yang menunjukkan potensi kesalahan kemudian dapat mengingatkan pemiliknya bahwa kendaraan harus diperbaiki atau dirawat secara proaktif.

Privacy (problems)

- Dengan navigasi, informasi GPS tertentu akan memungkinkan mobil dilacak riwayat perjalanannya
- Jika kamera menghadap ke dalam dan memantau pengemudi, informasi pribadi akan dikumpulkan dari masing-masing penumpang kendaraan



Privacy (problems)

- Dengan semua data yang diambil dari sensor, muncul kekhawatiran yang dalam hal pengumpulan data . Hal ini perlu diatur agar diproses dilakukan secara legal dan benar
- Dibutuhkan aturan yang meregulasi penggunaan data khusus untuk kendaraan otonom.





Next..

Pertemuan berikutnya akan membahas data understanding..



Institut Informatika & Bisnis
DARMAJAYA
Yayasan Kita-Hidup

TERIMA KASIH

Tugas Kelompok (Tugas 6)

Jelaskan komponen-komponen yang terlibat dalam masing-masing bisnis dan coba anda buat identifikasi apakah problem-problem berikut ini membutuhkan Artificial Intelligent atau hanya butuh automation. Lalu anda perlu membuat breakdown problem, flow, dan rekomendasi solusi yang memungkinkan pada saat ini (anda dapat melihat studi kasus autonomous car sebagai referensi).

Problem 1: Sistem rekomendasi otomatis produk pada e-commerce

Problem 2: Sistem Identifikasi otomatis laporan masyarakat

Problem 3: Bank Credit approval

Problem 4 : Sistem health insurance approval

Problem 5: Early disaster warning system.

Problem 6: Post disaster management system