**DATA SCIENCE**

**WHAT?**

Ilmu yang mempelajari mengolah data dengan menggunakan 3 ketrampilan sekaligus yaitu:

* Statistika
* Bisnis
* Bahasa Pemrograman

**WHY?**

* Saat ini, data berjumlah masif, variatif, dan dapat ditemukan di mana saja
* Pengambilan keputusan yang lebih akurat karena menggunakan berbagai bidang keilmuan

**DATA SCIENCE IMPLEMENTATION**

* Sektor Finansial

1. Melakukan manajemen resiko untuk mencegah kredit macet.
2. Mendeteksi potensi Fraud.
3. Menekan tingkat Customer Churn.

* Sektor Ritel & Distribusi

1. Menghasilkan prediksi untuk meningkatkan penjualan.
2. Menghasilkan Trend Forecasting.
3. Melakukan segmentasi konsumen.
4. Melakukan optimalisasi rute & armada.

* Sektor Media

1. Menghasilkan visualisasi: Data Storytelling.
2. Menghasilkan analisa sentimen positif dan negatif untuk berita.

**SAMPLE Success Story Of Data Science**

* **Alfa Group**

Tahun 2019 Alfamart meraup laba hingga 71,12% dengan kenaikan laba menjadi Rp 1,1 triliun dari sebelumnya Rp 650,14 milliar.

- mengembangkan CRM (*Customer Loyalty Management*)

- melakukan promosi yang *customized* dan *targeted*

* **Amazon**

Amazon berhasil menurunkan biaya pengiriman sebesar 40%

- Mengumpulkan data individu dan mengamati alamat pengiriman

- Melacak gudang terdekat yang dapat dipilih

**DATA SCIENCE METHODOLOGY**

Metodologi *data science* adalah langkah-langkah digunakan dalam proyek *data science* agar dapat menghasilkan hasil yang optimal yang dapat menjawab pertanyaan dari suatu masalah yang ingin diselesaikan.

Business Understanding

Analytic

Approach

Modeling

Data

Preparation

Feedback

Deployment

Evaluation

Data Requirements

Data

Collection

Data

Understanding

**Understanding Business**

* Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
* Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining.*
* Menyiapkan strategi awal untuk mencapai tujuan.

**Analytic Approach**

Pendefinisian masalah dalam konteks statistik atau Machine Learning untuk memperoleh hasil yang diinginkan

* Descriptive Analytics

Data analytics jenis ini dapat memberikan gambaran umum mengenai data yang telah dikumpulkan. Dalam analytics deskriptif ini kita hanya akan melihat informasi sederhana.

* Diagnostic Analytics

Proses analisis yang menentukan faktor dan peristiwa yang mengarah pada hasil dari peristiwa dan keadaan masa lalu.

* Predictive Analytics

Data analytics yang satu ini adalah untuk memprediksi sesuatu yang akan datang.

* Prescriptive Analytics

Analytics prescriptive merupakan proses analytics yang dapat menjawab mengapa sesuatu akan terjadi dan bagaimana cara menghadapi kondisi yang akan terjadi di masa yang akan datang.

**Data Requirements**

Tahapan seorang data science mengidentifikasi konten, format, dan sumber data yang diperlukan untuk pengumpulan data awal, dan menggunakan data ini di dalam algoritma pendekatan yang dipilih.

**Data Collection**

Dalam tahap Data Collection, data science mengidentifikasi sumber daya data yang tersedia ( *yang relevan* ) dengan domain masalah. Untuk mengambil data, dapat dilakukan dengan cara *web scrapping* pada situs web terkait, atau dapat menggunakan repositori dengan dataset premade yang siap digunakan.

\*Contoh *Data Collection* menggunakan ***pandas***.

import pandas as pd # download library to read data into dataframe

pd.set\_option('display.max\_column', None)  
dataframe = pd.read\_csv("csv\_file\_url")  
print("Data read into dataframe!")

dataframe.head() # show the first few rows  
dataframe.shape # get the dimensions of the dataframe

**Data Understanding**

Adalah tahapan seorang data science dimana lebih mencoba mengerti mengenai data yang telah dikumpulkan pada tahap “*Data Collection*”.

Data Scientist harus memeriksa type dari setiap data dan harus mempelajari atribut dan nama dari tiap data tersebut, selain itu diperlukan untuk pengecekan data yang hilang atau anomali yang muncul dari data tersebut

Output paling penting dari ***Data Understanding*** adalah memahami apakah data yang telah dikumpulkan mewakili dari masalaha yang akan disolusikan.

**Data Preparation**

Ketika data yang dimiliki telah dmengerti, tahapan selanjutnya adalah ***Data Preparation***. Pada tahapan ini meliputi manajemen data yang tidak lengkap/hilang, menghilangkan data duplikat, merubah format data ke bentuk yang seragam.

Output yang ingin di dapat pada tahapan ini adalah tidak ada kesalahan dalam data dan data yang akan di simpan sudah memiliki format yang benar untuk kemudian dipakai pada tahap *“Data Exploration”.*

**Modelling**

* Descriptive Modelling

Proses matematika yang menggambarkan peristiwa di dunia nyata dan hubungan antara faktor-faktor yang bertanggung jawab terhadap peristiwa tersebut.

Sifat :

Masa Lampau + Diagnostik

* Predictive Modelling

Proses yang menggunakan penambangan data dan probabilitas untuk memperkirakan hasil

Sifat:

Memprediksi data di masa depan

**Model Evaluation**

Data Scientist mengevaluasi kualitas model dan memverifikasi bahwa masalah bisnis telah diselesaikan secara lengkap dan memeriksa kualitas model untuk melihat apakah model tersebut memenuhi persyaratan yang ditetapkan oleh para stakeholders.

**Deployment**

Setelah mendapatkan model yang terbaik untuk pemecahan bisnis dan disetujui oleh stakeholders.

Step ini akan sering kali bekerja sama dengan tim back-end dan front-end untuk mengimplementasikan model yang dihasilkan untuk kemudahan penggunaan.

**Feedback**

Setelah implementasi model, kepentingan akan menerima feedback tentang kinerja model.

Analisa feedback memungkinkan Data Scientist untuk menyempurnakan model dan meningkatkan akurasi dan kegunaannya.