By Krian Bayu

**Data science methodology**

1. Business Understanding: tujuan dan konteks projeknya apa
   1. Memiliki pemahaan yang kuat terhadap projek
   2. List dan define prioritas masalahnya
   3. Tentukan tujuan bisnis
   4. Tentukan kriteria goals nya

Business metric, bisa collab denga tim digital marketing, membuat projection breng untuk mengetahui business growthnya sejauh mana.

Cth:

Metric untuk projection

1. Item dibeli
2. Active user
3. Average order value per orang
4. Analytical approach: cari tahu pendekatannya apa
   1. Descriptive: menjelaskan anomali/kejadian yangterjadi pada suatu periode ke belakang, belum dijelaskan reason terjadinya. Cth: 1 bulan kemarin rugi 10%
   2. Diagnostic: menjelasakn alasan dan faktor” terjadinya anomali tersebut. cth : rugi kemarin karena adanya error yang terjadi di perusahaan
   3. Predictive: buat apa yang akan terjadi kedepannya. Apa yang kita mau, list down data, validate result [supervised or unsupervised ]
   4. Prescriptive: menyiapkan poin” alasan dengan step” yang akan dijalankan ke depannya [reinforcement]

Cara Bagaiamana kita menggunakan data untuk mencapai target?

1. Data apa saja yang diperlukan, list down data
   1. Menyiapkan data” yang diperlukan
2. Data collection: jika belum ada datanya bisa web scrapping atau ambil data dari publik

Cth: jika menggunakan predictive analysis, jika ingin mengetahui prediksi revenue 1 bulan kedepan harus disiapkan, data” seperti orang” yang repurchase seperti apa

1. Data understanding: Mencari pattern/pola, berusaha mengerti pola secara kasat mata seprti demografi seperti apa
2. Data preparation: Setelah tau patternnya masuk data yang kita pakai,masuk ke data cleaning, feature engineering (memanipulasi data sesuai data poin yang kita mau)
   1. Data cleaning: data duplicate, null, missing data (bisa input data dengan modus mean media atau pakai data” yang dirasa tepat untuk mengisi yang missing), dan different format
   2. Formatting + feature engineering: data yang lebih ke entity user (aktivitas user)
3. Modelling: memliih algorithm yang tepat (klasifikasi or regresi)
4. Evaluation: test performance modellingnya, jika belum sesuai, bisa test ulang lgi degnan algoritma yang lain.
5. Deployment:
6. Feedback: maintain, jika terjadi anomali

mencari

Model evaluation: memprediksi dengan data yang kita punya sekarang

Accuracy:

TP + TN / Seluruh data

Precision: tingkat ketepatan dalam prediksi

TP/TP+FP (total predicted positive)

Recall: sensitifitas/kemampuan model untuk menangkap kejadian yang sebenarnya

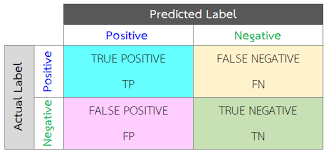
TP/TP+FN (total actual positive)

Keterangan: TP = True Positive

FP = False Positive

TN = True Negative

FN = False Negative



Cth case:

|  | selamat | Tidak selamat | Total aktual selamat |
| --- | --- | --- | --- |
| selamat |  |  |  |
| Tidak selamat |  |  |  |
|  |  |  |  |