CREDIT SCORING CLASSIFICATION

**Credit Score Cards merupakan metode yang biasa digunakan untuk mengontrol risiko didalam industri finance. Yaitu menggunakan informasi pribadi serta data yang telah di-inputkan oleh si pemohon kartu kredit guna memprediksi kemungkinan gagal bayar dan peminjaman kartu kredit kedepannya.**

**Pihak bank berhak menentukan apakah akan mengeluarkan credit card kepada pemohon. Credit score card ini dapat menilai secara objektif dalam mengukur risiko.**

adalah proses penilaian permohonan kredit yang dilakukan oleh lembaga kreditur/ pihak yang memberikan kredit kepada debitur selaku penerima kredit tersebut.

Tahapan ini adalah tahapan yang paling penting didalam proses kredit. Kesalahan di dalam tahapan ini akan berdampak besar pada keseluruhan tahapan pemberian kredit, dan secara global berpengaruh terhadap lembaga itu sendiri.

Bidang ilmu dari teknologi informasi, yang bisa membantu *credit scoring* adalah *data mining*. Salah satu algoritma  yang bisa digunakan di dalam *data mining* adalah Classification And Regresion Tree (CART).

1. **EXPLORATORY DATA ANALYSIS ( EDA )**

Biasa nya digunakan di awal tahap proses projek data science, sebelum model data science dibangun.

Tujuan nya adalah untuk melihat data mentah yg sudah kita dapat kemudian di eksplor dan menyatukan setiap informasi dari data tersebut yg tidak hanya membantu kita dlm membuat model yg lebih baik, tetapi juga memberikan informasi bisnis yang relevan yg berasal dari data ini.

<https://www.kaggle.com/mohammadhy/eda-credict-card-approval>

Education:

secondary special education :SMK

higher education : SMA

incomplete higher : tidak selesai SMA

lower secondary school : SMP

academic degree : Universitas

1. **Preprocessing Data**

1. Data Frame
2. Handling Duplicates values
3. Handling missing values
4. Separating categorical and numerical data
5. Encoder
6. Outliers removal

* Membuat list numeric values terpisah , karna numeric itu dpt di proses dengan proses statistika seperti quantil atas, quantil bawah.

Kalau yang non numeric values atau object type dia dihitung jumlah values nya misal berapa jumlah female dan male di kolom gender , jumlah org yang memiliki property dan yang tidak memiliki property.

* Outliers removal

Mengatasi Outlier

Mengatasi Outlier dengan menggunakan metode interquartile range (IQR). Untuk detil jelasnya, bisa dilihat di ilustrasi di bawah ini:

Gunakan hasil dari pengolahan di tahap sebelumnya df\_load untuk di olah di tahap ini. Gunakan fungsi quantile() untuk melihat quantile tertentu, dan gunakan mask() untuk mereplace nilai. Tentukan:

* Nilai Minimum dan Maximum data di tolerir
* Ubah Nilai yg di luar Range Minumum & Maximum ke dalam nilai Minimum dan Maximum

1. **Model training**

* **Split the input data** membagi dataset menjadi train data dan test data
* **Build the models** by using the training data set.
* **Determine the “best” solution** menentukan solusi terbagus berdasarkan hasil akurasi dari masing-masing algoritma

**Metode SMOTE**

Oversampling digunakan pada data latih. Metode SMOTE digunakan untuk menambah data sintetis pada dataset kelas minor (dalam kasus ini adalah kelas 1), sehingga jumlah instances pada kelas minoritas menjadi sama dengan jumlah instances pada kelas mayoritas (dalam kasus ini adalah kelas 0).

**Feature Scaling**

adalah suatu cara untuk membuat numerical data pada dataset memiliki rentang nilai (scale) yang sama. Tidak ada lagi satu variabel data yang mendominasi variabel data lainnya.

Sumber:

* <https://medium.com/analytics-vidhya/part-1-data-preparation-made-easy-with-python-e2c024402327>
* <https://medium.com/analytics-vidhya/part-2-data-preparation-made-easy-with-python-633027bf6f28>
* <https://medium.com/@nehakushwaha1308/part-3-data-preparation-made-easy-with-python-632f6647add2>
* [pandas.DataFrame.nunique — pandas 1.2.0 documentation (pydata.org)](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.nunique.html)
* <https://medium.com/sciforce/data-cleaning-and-preprocessing-for-beginners-25748ee00743>
* Data cleansing (QUANTILE) [Data Science in Telco: Data Cleansing | Hendra](https://www.hendra.my.id/2020/09/data-science-in-telco-data-cleansing.html)