FINAL TASK DATA SCIENTIST ID/X PARTNER – VIX RAKAMIN – CREDIT RISK PREDICTION

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

Created by:
Candraditya Dwaya Putra
Email : candradityaputra1@gmail.com
linkedIn:
https://www.linkedin.com/in/candraditya-dwaya-putra-77719a66

"Have more than 11 years professional career in GIS and Management geodatabase. Good knowledge WebGIS Development, with strong experience Geospatial data in Forestry, environmental and regional planning. Skilled at data collection and analysis that elicits accurate and valuable information utilizing technical principles and theories. Technical proficiencies include SQL, Python, Machine Learning, MS Office, ESRI GIS, QGIS, HTML, CSS/Bootstrap, Javascript, geomorphology and Geospatial Software."

Overview

"Sebuah perusahaan di Indonesia ingin memprediksi *credit risk* untuk mengurangi risiko kredit macet, memaksimalkan dalam pengambilan keputusan pemberian kredit, mengetahui target pemberian kredit dan meminimalkan kerugian perusahaan. Oleh karena itu dibutuhkan analisis mendalam dari data yang telah tersedia yang kemudian dapat dikomunikasikan kepada pemangku kepentingan, managemen perusahaan bahkan bagian keuangan perusahaan. Sehingga perusahaan dapat menyalurkan kredit tepat kepada individu yang layak.

Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 466285 entries, 0 to 466284
Data columns (total 75 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   Unnamed: 0                  466285 non-null  int64
 1   id                          466285 non-null  int64
 2   member_id                   466285 non-null  int64
 3   loan_amnt                   466285 non-null  int64
 4   funded_amnt                 466285 non-null  int64
 5   funded_amnt_inv             466285 non-null  float64
 6   term                        466285 non-null  object
 7   int_rate                    466285 non-null  float64
 8   installment                 466285 non-null  float64
 9   grade                       466285 non-null  object
 10  sub_grade                   466285 non-null  object
 11  emp_title                   438697 non-null  object
 12  emp_length                  445277 non-null  object
 13  home_ownership              466285 non-null  object
 14  annual_inc                  466281 non-null  float64
 15  verification_status         466285 non-null  object
 16  issue_d                     466285 non-null  object
 17  loan_status                 466285 non-null  object
 18  pymnt_plan                  466285 non-null  object
 19  url                         466285 non-null  object
 20  desc                        125983 non-null  object
 21  purpose                     466285 non-null  object
 22  title                       466265 non-null  object
 23  zip_code                    466285 non-null  object
 24  addr_state                  466285 non-null  object
 25  dti                         466285 non-null  float64
 26  delinq_2yrs                 466256 non-null  float64
 27  earliest_cr_line            466256 non-null  object
 28  inq_last_6mths              466256 non-null  float64
 29  mths_since_last_delinq      215934 non-null  float64
 30  mths_since_last_record      62638 non-null   float64
 31  open_acc                    466256 non-null  float64
 32  pub_rec                     466256 non-null  float64
 33  revol_bal                   466285 non-null  int64
 34  revol_util                  465945 non-null  float64
 35  total_acc                   466256 non-null  float64
 36  initial_list_status         466285 non-null  object
 37  out_prncp                   466285 non-null  float64
 38  out_prncp_inv               466285 non-null  float64
 39  total_pymnt                 466285 non-null  float64
 40  total_pymnt_inv             466285 non-null  float64
 41  total_rec_prncp             466285 non-null  float64
 42  total_rec_int               466285 non-null  float64
 43  total_rec_late_fee          466285 non-null  float64
 44  recoveries                  466285 non-null  float64
 45  collection_recovery_fee     466285 non-null  float64
 46  last_pymnt_d                465909 non-null  object
 47  last_pymnt_amnt             466285 non-null  float64
 48  next_pymnt_d                239071 non-null  object
 49  last_credit_pull_d          466243 non-null  object
 50  collections_12_mths_ex_med  466140 non-null  float64
 51  mths_since_last_major_derog 98974 non-null   float64
 52  policy_code                 466285 non-null  int64
 53  application_type            466285 non-null  object
 54  annual_inc_joint            0 non-null       float64
 55  dti_joint                   0 non-null       float64
 56  verification_status_joint   0 non-null       float64
 57  acc_now_delinq              466256 non-null  float64
 58  tot_coll_amt                396009 non-null  float64
 59  tot_cur_bal                 396009 non-null  float64
 60  open_acc_6m                 0 non-null       float64
 61  open_il_6m                  0 non-null       float64
 62  open_il_12m                 0 non-null       float64
 63  open_il_24m                 0 non-null       float64
 64  mths_since_rcnt_il          0 non-null       float64
 65  total_bal_il                0 non-null       float64
 66  il_util                     0 non-null       float64
 67  open_rv_12m                 0 non-null       float64
 68  open_rv_24m                 0 non-null       float64
 69  max_bal_bc                  0 non-null       float64
 70  all_util                    0 non-null       float64
 71  total_rev_hi_lim            396009 non-null  float64
 72  inq_fi                      0 non-null       float64
 73  total_cu_tl                 0 non-null       float64
 74  inq_last_12m                0 non-null       float64
dtypes: float64(46), int64(7), object(22)
memory usage: 266.8+ MB
```

Description
Dataset mengandung kebiasaan pinjam customer

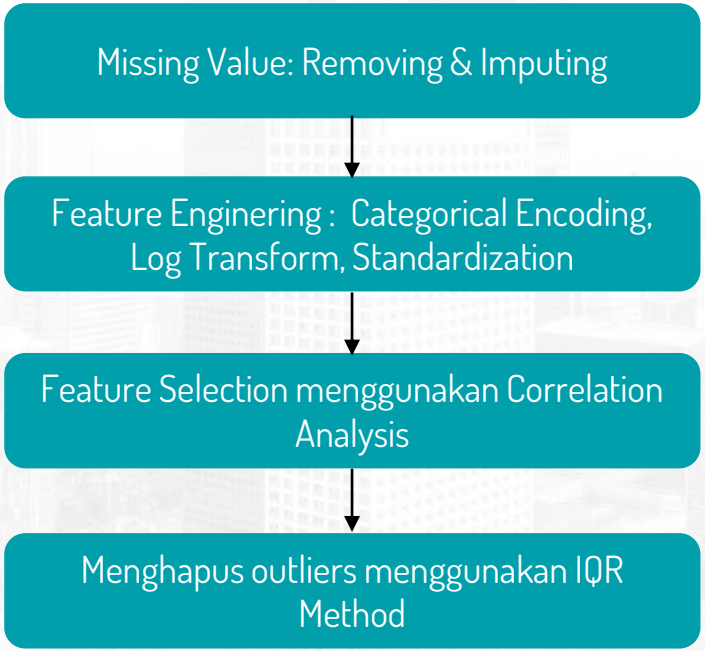
Shape
75 Columns

Dtype
Float64 (46 features), int64(7 features), Object (22 features)

Missing Value
Banyak Missing Value, sehingga perlu drop 17 features

Untuk selengkapnya, dapat melihat jupyter notebook disini
https://colab.research.google.com/drive/1uuPBOEkWOLnxp11-3tDSufuNeSS3GRFY?usp=sharing

Data Preparation



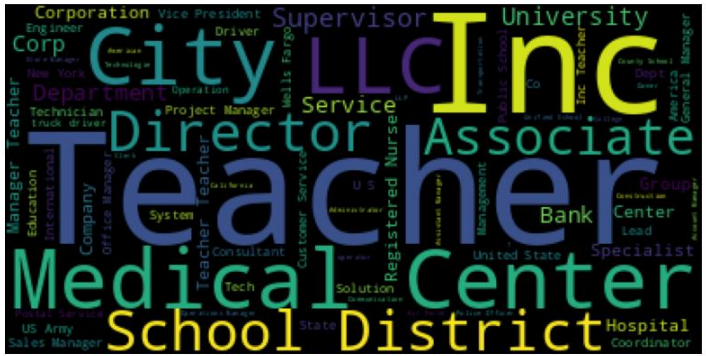
Untuk selengkapnya, dapat melihat jupyter notebook disini
https://colab.research.google.com/drive/1uuPBOEkWOLnxp11-3tDSufuNeSS3GRFY?usp=sharing

Exploratory Data Analysis

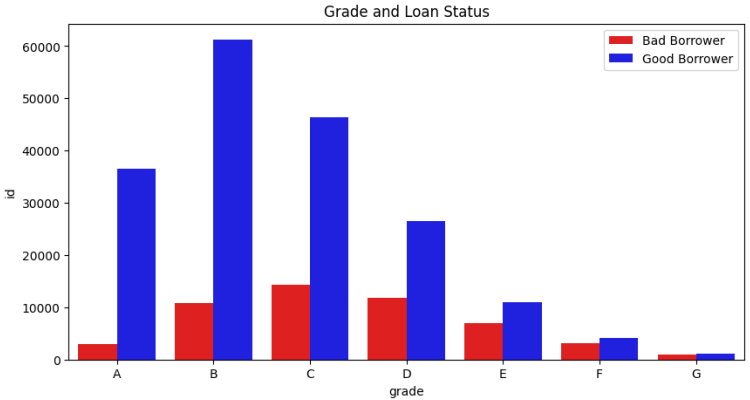
Target Variable

- Good Borrower (1) : **Fully Paid**, Does not meet the credit policy. Status:Fully Paid
- Bad Borrower (0) : **Charged Off**, Does not meet the credit policy. Status: **Charged Off**, **Default**, **Late (31-120 days)**
- Undetachable Loan (-1) : **Current**, **In Grace Period**, **Late (16-30 days)**

Status Pekerjaan Nasabah



Bagaimana dengan mengklasifikasikan nilai peminjam kita dan status pinjamannya



untuk selengkapnya, dapat melihat jupyter notebook disini
https://colab.research.google.com/drive/1uuPBOEkWOLnxp11-3tDSufuNeSS3GRFY?usp=sharing

Modeling

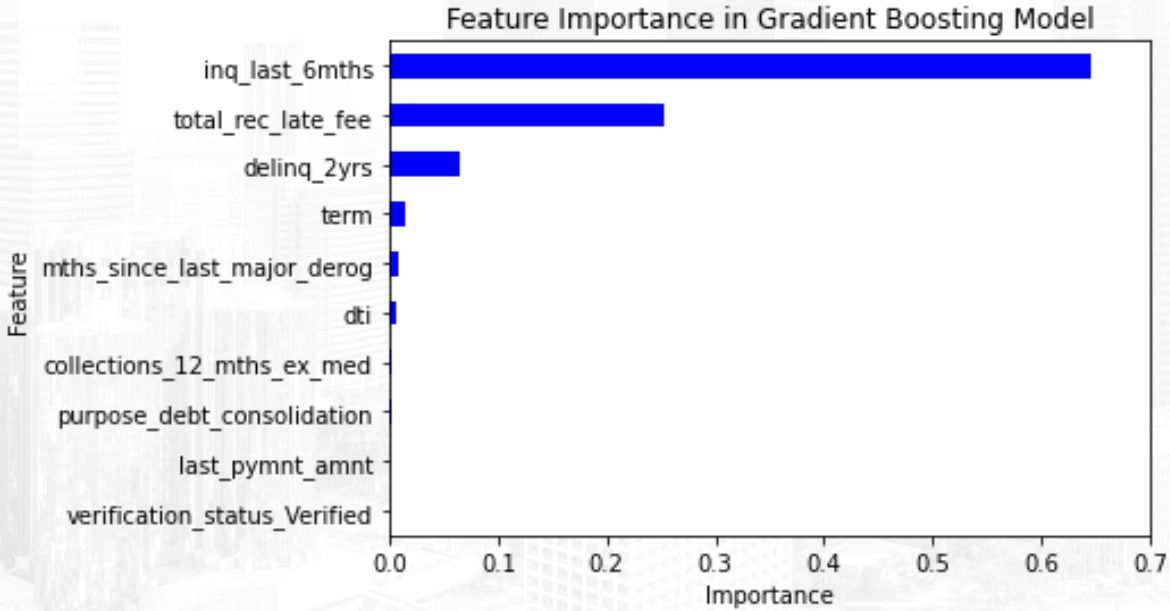
- Training Test Split (70% Training, 30% Testing)
- Modeling menggunakan SMOTE (Synthetic Minority Over-sampling Technique) karena metode oversampling yang menghasilkan sampel sintetis untuk kelas minoritas dengan cara menggabungkan sampel dari kelas minoritas yang ada. Tujuan utama dari SMOTE adalah untuk meningkatkan kinerja model pada kelas minoritas dengan meningkatkan jumlah sampel yang ada dalam kelas tersebut. Hal ini membantu mencegah model pembelajaran mesin menjadi bias terhadap kelas mayoritas dan meningkatkan kemampuannya untuk memprediksi kelas minoritas dengan akurasi yang lebih tinggi.
- Semua Langkah dihandle dengan pipeline

Evaluation Metric

Model	FN	Recall	ROC-AUC	KS
Random Forest	608	96	99,41	94,4
Gradient Boosting Trees	386	97	99,48	94,28
XGBoost	447	97	99,43	93,83
Voting Classifier	420	97	99,48	94,34

Untuk selengkapnya, dapat melihat jupyter notebook disini
https://colab.research.google.com/drive/1uuPBOEkWOLnxp11-3tDSufuNeSS3GRFY?usp=sharing

Top 10 Features Importance



Untuk selengkapnya, dapat melihat jupyter notebook disini
https://colab.research.google.com/drive/1uuPBOEkWOLnxp11-3tDSufuNeSS3GRFY?usp=sharing