Rakamin Academy X id/x partners

Prediction Model

ID/X Partners - Data Scientist

Presented by
Hamzah Mulyana

Cirebon, Jawa Barat

hamzahmulyana88@gmail.com

Hamzah Mulyana

Hamzah Mulyana

Data Scientist

Data Scientist with under one year of experience and certified in Data Science & Machine Learning from Purwadhika Digital Technology School. Skilled in data collection, cleaning, processing, and visualization using Python (Pandas, Numpy, Matplotlib, Seaborn) and Microsoft Excel (Basic Formula, IF Condition, Lookup Value, Pivot Table), along with tools like Tableau, Google Looker Studio. Proficient in MySQL, Spreadsheets, and Microsoft Office, with statistical analytics capabilities to support data-driven decision making. Detail-oriented and committed to applying technical skills to solve business problems through data-driven approaches.

Courses and Certification

Linear Models in Machine Learning | Link	**May, 2025**
Purwadhika Digital Technology School | Link	**Oct, 2024**
Machine Learning | Link	**Aug, 2024**
Data Analysis | Link	**Jul, 2024**

About Company

ID/X Partners didirikan pada tahun 2002 oleh *ex-banker* dan *management consultants* yang memiliki pengalaman dalam *credit cycle and process management, scoring development, and performance management*. Pengalaman gabungan kami telah melayani perusahaan-perusahaan di wilayah Asia dan Australia di berbagai industri, khususnya jasa keuangan, telekomunikasi, manufaktur, dan ritel.

ID/X Partners menyediakan layanan konsultasi yang mengkhususkan diri dalam pemanfaatan solusi *Data Analytic and Decisioning* (DAD), dikombinasikan dengan pendekatan terpadu dalam manajemen risiko dan pemasaran, untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis mereka.

Dengan layanan konsultasi yang komprehensif dan solusi teknologi yang ditawarkan, id/x partners menjadi penyedia layanan *one-stop solution* bagi kebutuhan bisnis klien.

Project Portfolio

Rakamin Academy

id/x partners

Proyek ini bertujuan membangun model machine learning untuk memprediksi risiko kredit pada perusahaan multifinance guna meningkatkan akurasi penilaian kelayakan peminjam dan mengurangi potensi kredit macet. Dataset yang digunakan mencakup riwayat pinjaman beserta berbagai atribut terkait seperti profil nasabah, riwayat pembayaran, dan karakteristik pinjaman. Solusi ini diharapkan dapat membantu perusahaan dalam mengoptimalkan proses pengambilan keputusan pemberian pinjaman sekaligus meminimalkan kerugian akibat gagal bayar, dengan menghasilkan model prediktif yang akurat beserta insight bisnis yang relevan untuk mendukung strategi penyaluran kredit yang lebih efektif.

Link code here!

Project explanation video here!

Data Understanding

Rakamin Academy

id/x partners

Data Understanding

Data Size
- Data terdiri dari 466285 baris dan 74 kolom

Data Condition
- Data types: Object, Float, Int
- Missing Value: Lebih dari 10 kolom tidak memiliki Nilai
- Duplicated Data: Tidak ada duplikat data

Data Preparation

Data Preparation

Rakamin Academy

id/x partners

Handle Duplicated Data

- Tidak terdapat duplikat dalam data

Delete Columns

- Menghapus kolom yang tidak digunakan untuk analisis dan prediksi, seperti:

'id', 'member_id', 'funded_amnt', 'funded_amnt_inv','url', 'desc', 'zip_code', 'title', 'mths_since_last_delinq', 'mths_since_last_record', 'mths_since_last_major_derog', 'policy_code', 'application_type', 'annual_inc_joint', 'dti_joint', 'verification_status_joint', 'open_acc_6m', 'open_il_6m', 'open_il_12m', 'open_il_24m', 'mths_since_rcnt_il', 'total_bal_il', 'il_util', 'open_rv_12m', 'open_rv_24m', 'max_bal_bc', 'all_util', 'total_rev_hi_lim', 'inq_fi', 'total_cu_tl', 'inq_last_12m', 'pymnt_plan', 'out_prncp_inv', 'total_pymnt_inv', 'sub_grade', 'earliest_cr_line', 'emp_title', 'last_credit_pull_d'

Data Preparation

Rakamin Academy

id/x partners

Change Columns Value

- loan_status: Fully Paid, Current, In Grace Periode -> Diterima (1)
- verification_status: Source Verified -> Verified
- purpose: house -> home_improvement
- home_ownership: ANY -> OTHER

Fill NaN

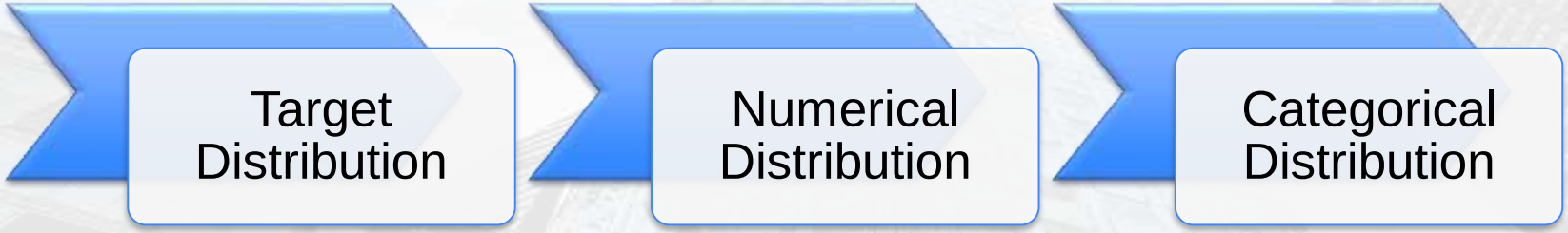
- Mengisi nilai dengan menggunakan metode modus(mode) untu kolom:

'delinq_2yrs', 'inq_last_6mths', 'open_acc', 'pub_rec', 'revol_util', 'total_acc', 'collections_12_mths_ex_med', 'acc_now_delinq'

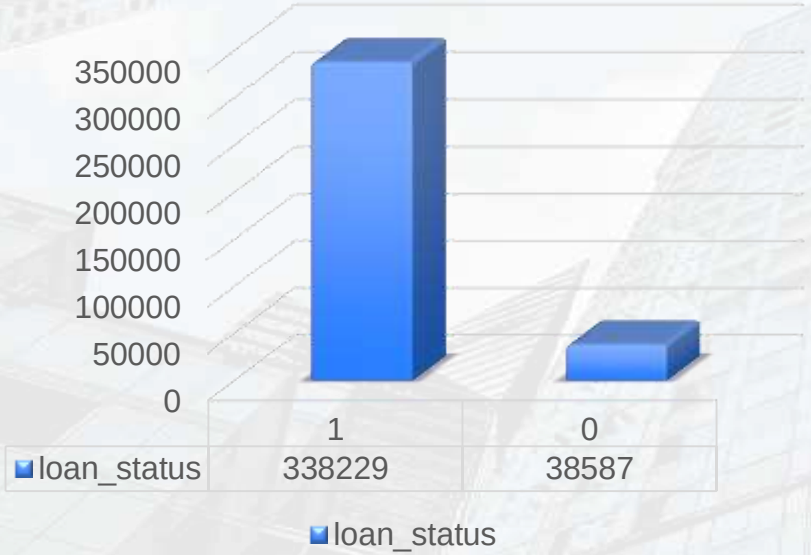
Drop NaN

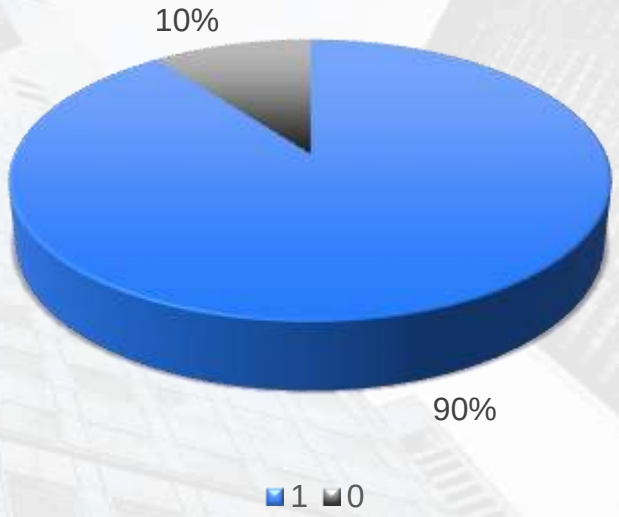
- Menghapus baris yg memiliki nilai NaN karena tidak ada acuan untuk mengisinya

Exploratory Data Analysis



Exploratory Data Analysis

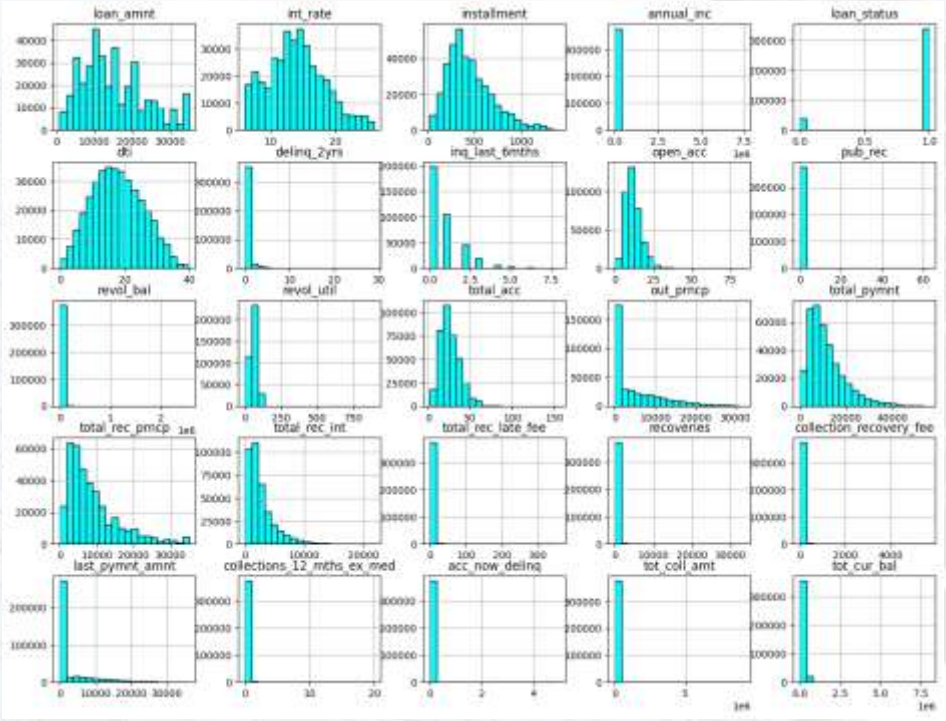




Exploratory Data Analysis

Dapat dilihat pada visualisasi disamping, untuk data numerical setiap fitur/kolom memiliki distribusi data yang tidak normal, maka dari itu analisis statistiknya menggunakan Non Parametrik.

Setalah dilakukan uji Non Parametrik ada kolom yang tidak akan dimasukan dalam Feature Engineering.



Exploratory Data Analysis

Rakamin Academy

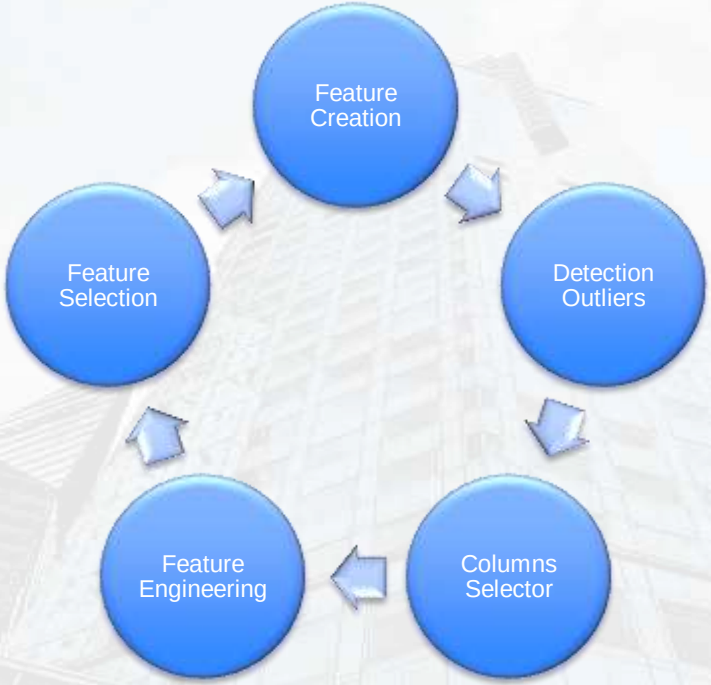
id/x partners

Categorical Feature

- Untuk categorical feature, hasilnya sebagai berikut:

term -> 36 months
grade -> B
emp_length -> 10+ years
home_ownership -> mortgage
verification_status -> verified
purpose -> debt_consolidation

Feature Engineering



Feature Engineering

Rakamin Academy

id/x partners

Feature Creation

- Membuat fitur tambahan untuk analisis, kasus kali ini yaitu memisahkan bulan dan tanggal kedalam kolom berbeda

Outliers Detection

- Untuk kolom yang memiliki nilai outliers tinggi, maka akan dilakukan penangan dalam Outliers Detection ini, yaitu akan diubah ke rentang percentil 10 sampai percentil 90

Columns Selector

- Untuk kolom yang tidak akan dilakukan apa-apa maka Columns Selector ini dapat digunakan, alasannya karena ada kolom yang tidak perlu dilakukan feature engineering maka dari itu dibuatlah Columns Selector ini.

Feature Engineering

Rakamin Academy

id/x partners

Feature Engineering

- **Onehot Encoding, Ordinal Encoding, Scaling (Robust), Simple Imputer**

Feature Selection

- **Memilih feature yang berpengaruh untuk dilakukan modeling, pemilihan feature ini menggunakan Feature Importance pada algoritma Random Forest. Didapatkan hasil feature-feature yang akan diambil sebagai berikut:**

'mnth_last_pymnt_d', 'mnth_next_pymnt_d', 'loan_amnt', 'total_rec_prncp', 'collection_recovery_fee', 'last_pymnt_amnt', 'recoveries', 'total_pymnt', 'out_prncp', 'installment', 'total_rec_int', 'total_rec_late_fee', 'int_rate', 'day_next_pymnt_d', 'day_last_pymnt_d'

Data Modeling

Rakamin Academy

id/x partners

- **Linear**
  - Logistic Regression
- **Ensemble**
  - Decision Tree
  - Random Forest
  - AdaBoost Classifier

Data Modeling

Before Tunning (Cross Validation)

Nama Model	Precision	
	Train	Test
Logistic Regression	98%	98%
Decision Tree	99%	99%
Random Forest	99%	99%
AdaBoost Classifier	99%	99%

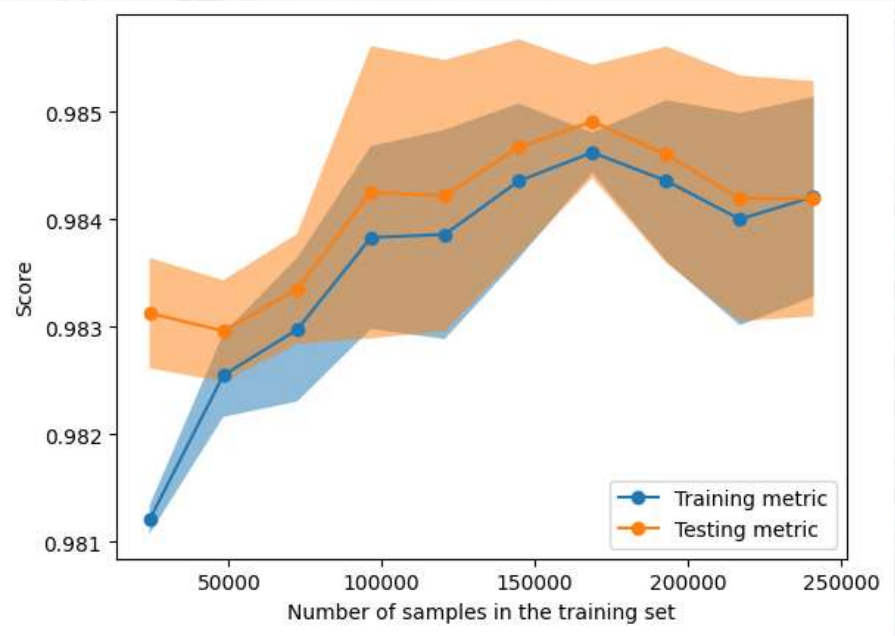
Data Modeling

After Tunning (Hyperparameter Tunning)

Nama Model	Precision	
	Train	Test
Logistic Regression	95%	95%
Decision Tree	99%	99%
Random Forest	99%	99%
AdaBoost Classifier	99%	99%

Evaluation

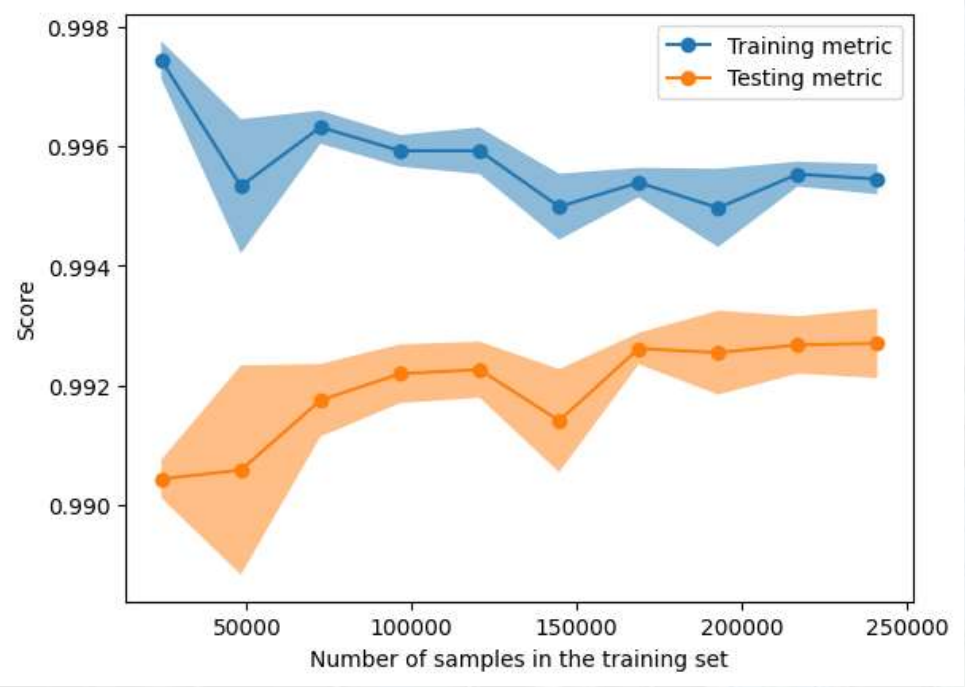
Logistic Regression



Dapat dilihat Learning Curve disamping, data training dan data testing menunjukan hasil yang bagus. Hasil dari data training dan data testing tidak terlalu jauh.

Evaluation

Decision Tree



Dapat dilihat Learning Curve disamping, data training dan data testing menunjukan hasil yang tidak terlalu bagus (indikasi adanya overfitting). Hasil dari data training dan data testing jauh.

Conclusion

Logistic Regression

- Dengan precision 95%, model dapat memprediksi bahwa orang yang diprediksi layak mendapatkan pinjaman memang benar layak (true positive). Ini menunjukkan bahwa model sangat baik dalam meminimalkan Type I Error (False Positive) hingga 5%. Threshold yang digunakan sangat ketat, yaitu 0.16, sehingga hanya prediksi dengan probabilitas tinggi yang diklasifikasikan sebagai 'layak'.

Decision Tree

- Dengan precision 98%, model dapat memprediksi bahwa orang yang diprediksi layak mendapatkan pinjaman memang benar layak (true positive). Ini menunjukkan bahwa model sangat baik dalam meminimalkan Type I Error (False Positive) hingga 2%. Threshold yang digunakan sangat ketat, yaitu 0.05, sehingga hanya prediksi dengan probabilitas tinggi yang diklasifikasikan sebagai 'layak'.

Thank You