

PREDICTION MODELING

ID/X Partners - Data Scientist

Presented By
Mhd. Arsyia Fikri



Hello,

I'm Mhd. Arsyia Fikri

Data Scientist

An enthusiastic person with various interests, quick to learn and adapt to new challenges. Experience public speaker that possess strong academic and technical skills in programming and data science through training, courses, laboratory instructor, and machine learning-focused final project. Motivated and actively looking for career opportunities in the data field, with a focus on data science.



Project Portfolio

Credit Risk Analysis

- ID/X Partners terlibat dalam proyek dengan perusahaan *multifinance* untuk meningkatkan penilaian dan pengelolaan risiko kredit.
- Tujuan: Mengoptimalkan keputusan bisnis dan mengurangi potensi kerugian dengan mengembangkan model machine learning untuk memprediksi risiko kredit.

Project explanation video is [here!](#)

Github Repository is also [here!](#)

About Company

id/x partners

- ID/X Partners didirikan pada tahun 2002 oleh mantan banker dan konsultan manajemen yang memiliki pengalaman luas dalam siklus kredit dan manajemen proses, pengembangan skor kredit, dan manajemen kinerja. Pengalaman gabungan kami telah melayani perusahaan-perusahaan di berbagai wilayah Asia dan Australia dan dalam berbagai industri, khususnya layanan keuangan, telekomunikasi, manufaktur, dan ritel.

About Company

id/x partners

- ID/X Partners menyediakan layanan konsultasi yang mengkhususkan diri dalam menggunakan solusi analitik data dan pengambilan keputusan (DAD) yang dikombinasikan dengan disiplin manajemen risiko terintegrasi dan pemasaran untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis.
- Layanan konsultasi komprehensif dan solusi teknologi yang ditawarkan oleh ID/X Partners membuatnya sebagai penyedia layanan lengkap yang dapat diandalkan.



Rakamin
Academy

X

id/x partners

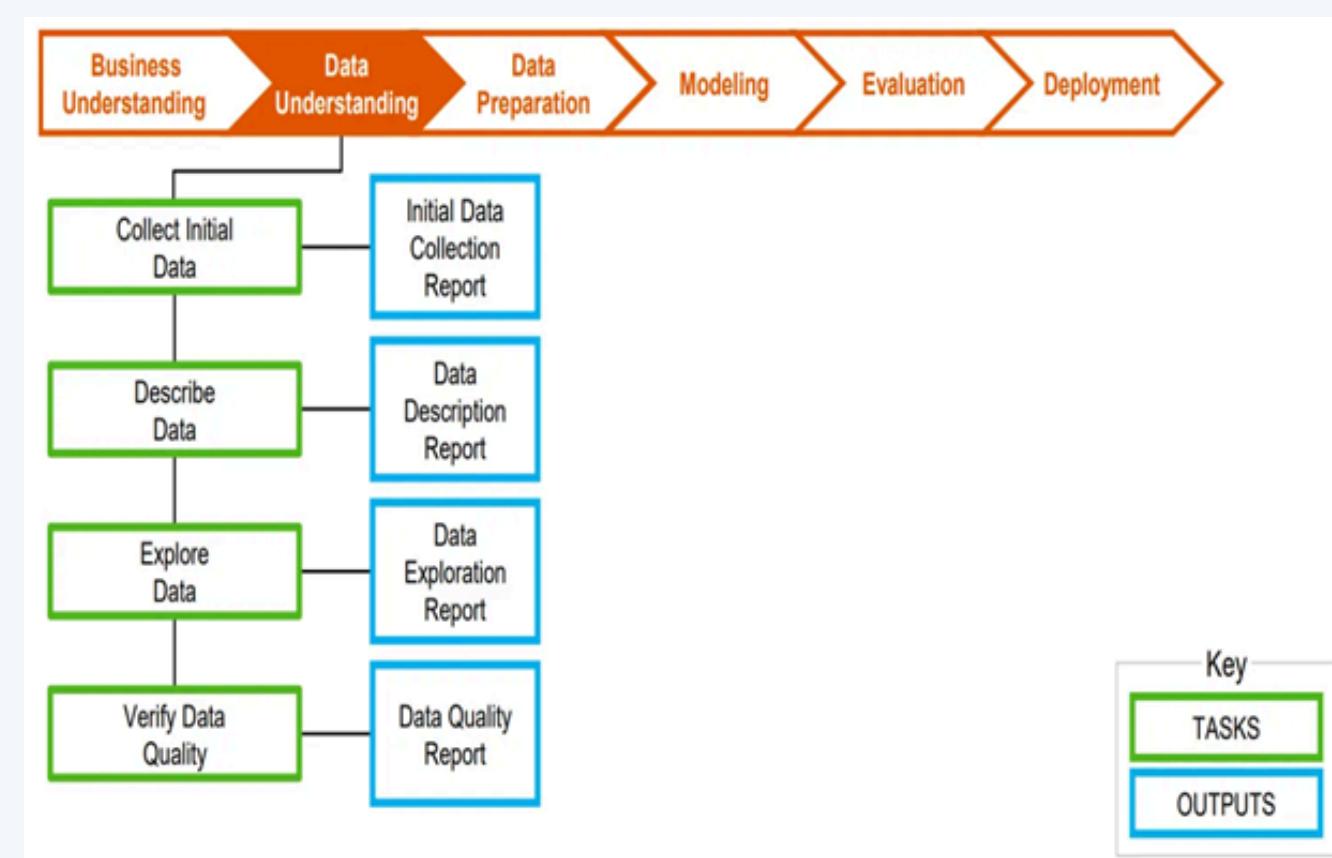
PREDICTION MODELING

Tahapan Pengerjaan

...

1. Data Understanding

Data Understanding adalah tahap kedua dalam proses CRISP-DM (Cross-Industry Standard Process for Data Mining) yang fokus pada pengumpulan dan penilaian kualitas data.



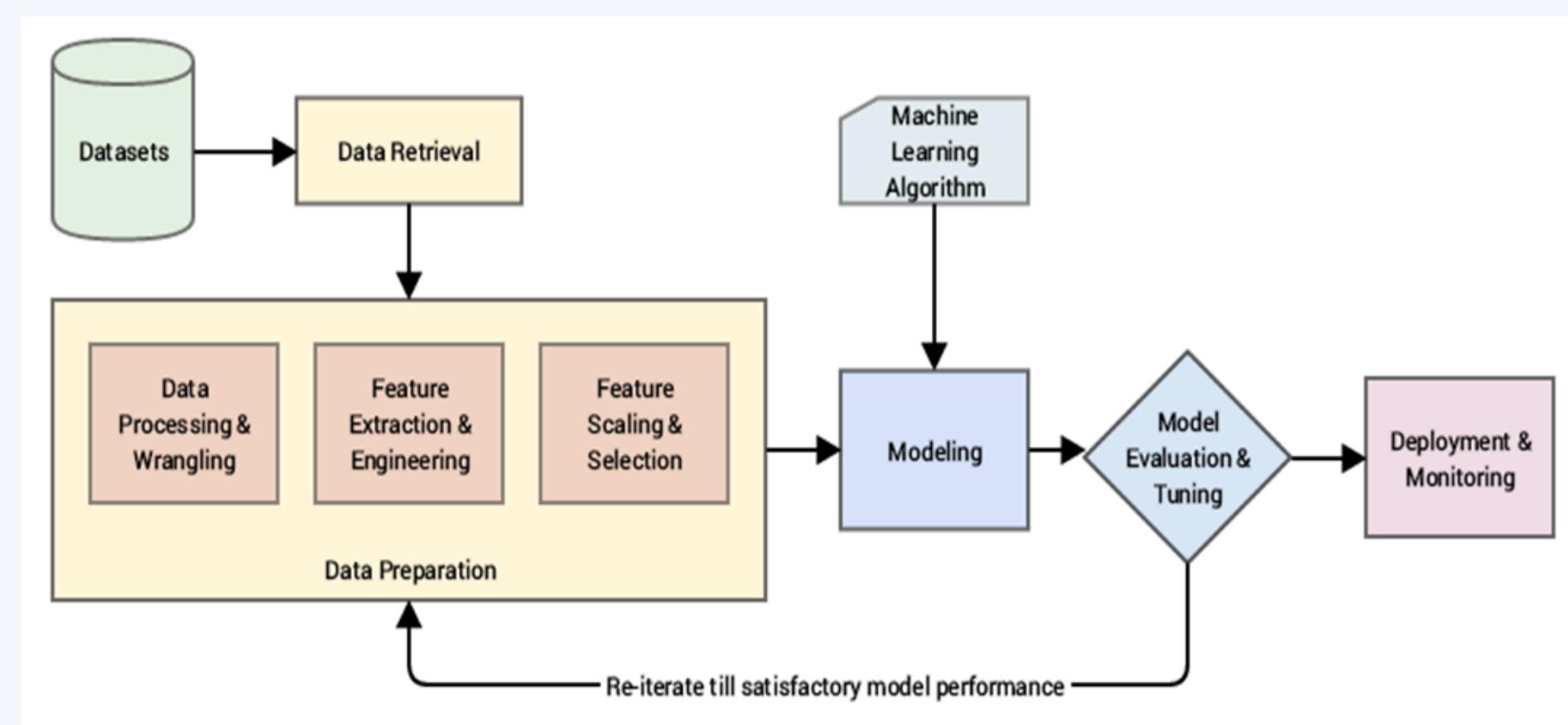
1. Data Understanding

Tahap ini melibatkan empat tugas utama:

1. Mengumpulkan Data Awal: Mengidentifikasi data yang tersedia, metode pengambilan, dan masalah yang mungkin dihadapi.
2. Mendeskripsikan Data: Memeriksa properti data yang diperoleh, termasuk format, kuantitas, dan isi dari setiap tabel atau dataset.
3. Menjelajahi Data: Menggunakan pertanyaan ilmu data untuk mendapatkan wawasan awal melalui kueri, visualisasi, dan laporan ringkasan.
4. Memverifikasi Kualitas Data: Memastikan data cukup bersih dan relevan untuk analisis yang akan dilakukan

2. Feature Engineering

Feature Engineering dalam ilmu data adalah proses kreatif yang melibatkan pemilihan, transformasi, dan pembuatan fitur baru dari data mentah untuk meningkatkan kinerja model pembelajaran mesin.



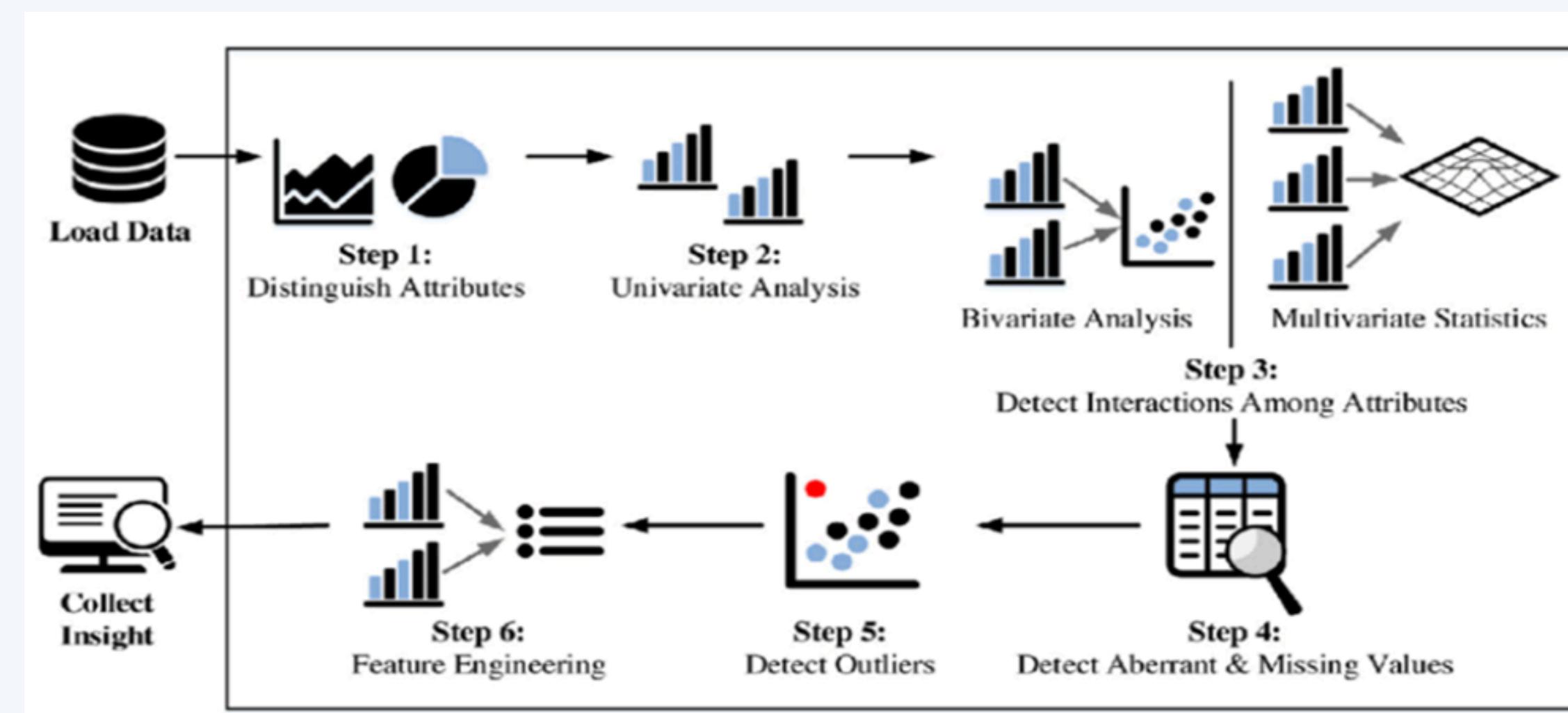
2. Feature Engineering

Feature Engineering termasuk di dalamnya:

1. Seleksi Fitur: Memilih fitur yang paling relevan dengan masalah yang sedang dihadapi.
2. Transformasi Fitur: Mengubah skala atau distribusi fitur untuk meningkatkan interpretasi oleh model.
3. Penciptaan Fitur: Menggabungkan atau memodifikasi fitur untuk menghasilkan informasi yang lebih berguna.
4. Ekstraksi Fitur: Mengidentifikasi dan mengekstrak informasi penting dari data mentah.

3. Exploratory Data Analysis

Exploratory Data Analysis (EDA) dalam ilmu data adalah proses analisis awal data untuk memahami karakteristik utama, menemukan pola, anomali, dan hubungan antar variabel.



3. Exploratory Data Analysis

Proses EDA biasanya meliputi:

1. Gambaran Umum Dataset: Memahami jumlah observasi, jenis fitur, dan data yang hilang.
2. Statistik Deskriptif: Meringkas data numerik melalui ukuran tendensi sentral dan dispersi.
3. Visualisasi Data: Menggunakan grafik dan diagram untuk menggambarkan distribusi dan hubungan data.
4. Evaluasi Kualitas Data: Memeriksa kebersihan dan konsistensi data.

4. Data Preparation

Proses Data Preparation yang dilakukan meliputi:

- Data Cleaning: Pembersihan data menyesuaikan format setiap kolom-kolom data
- Feature Selection: Memilih kolom-kolom yang memiliki keterkaitan dan signifikansi terhadap variabel target.
- Handling Missing Values: Dilakukan dengan berbagai metode, seperti menghapus kolom yang terlalu banyak data miss, menghapus baris data yang MCAR, serta imputasi KNN dan mean.
- Handling Anomaly: Melakukan analisa lebih lanjut terhadap data-data outliers ekstrim, dan menghapusnya.

5. Modeling

Tahap Modeling atau Pemodelan adalah tahapan yang dilakukan untuk melatih model machine learning untuk memprediksi data. Strategi yang dilakukan sebelum melakukan pemodelan adalah:

1. Feature Encoding: Dilakukan dengan 3 metode yang berbeda, yaitu: Ordinal Encoding, untuk data yang memiliki tingkatan di setiap kategorinya. Label Encoding, untuk data yang tidak memiliki tingkatan dan data yang memiliki kategori cukup banyak. Terakhir, adalah One Hot Encoding, untuk data-data yang tidak memiliki tingkatan dan tidak terlalu banyak kategori.
2. Selanjutnya adalah Train Test Split: Pembagian dataset menjadi dua set data, yaitu training dan testing dengan perbandingan 80:20

5. Modeling

Untuk tahap pelatihan model sendiri, dilakukan terhadap 5 jenis model klasifikasi yang berbeda: Logistic Regression, Random Forest, LightGBM, CatBoost, dan XGBoost.

Dilanjutkan dengan Hyperparameter Tuning model CatBoost untuk model dengan performa yang terbaik.

6. Evaluation

Hasil Evaluasi Training Model yang terbaik adalah:



CatBoost

CatBoost

Precision : 0.98

Recall : 0.90

ROC-AUC : 0.97

7. Conclusion

Kesimpulan yang dapat diambil adalah:

1. Riwayat kredit peminjam sangat menjamin dalam penentuan resiko kredit yang diajukan.
2. Semakin banyak riwayat kasus pinjaman bermasalah yang dilakukan, maka akan semakin besar juga kemungkinan ia akan melakukan default atau masalah lainnya.
3. Kestabilan ekonomi peminjam juga menjadi salah satu penentu buruk atau tidaknya pinjaman tersebut.
4. Agar lebih memaksimalkan profit, stakeholder dapat memfokuskan pada peminjam dengan kestabilan ekonomi yang baik dan memberikan tingkat suku bunga yang rendah (5 - 10%).



Thank you!

Let's Connect!

Contact Details

Phone: +62 822 7580 9719

LinkedIn: linkedin.com/in/arsya-fikri/

Email: mhdarsyafikri@gmail.com

GitHub: github.com/arsyfpro