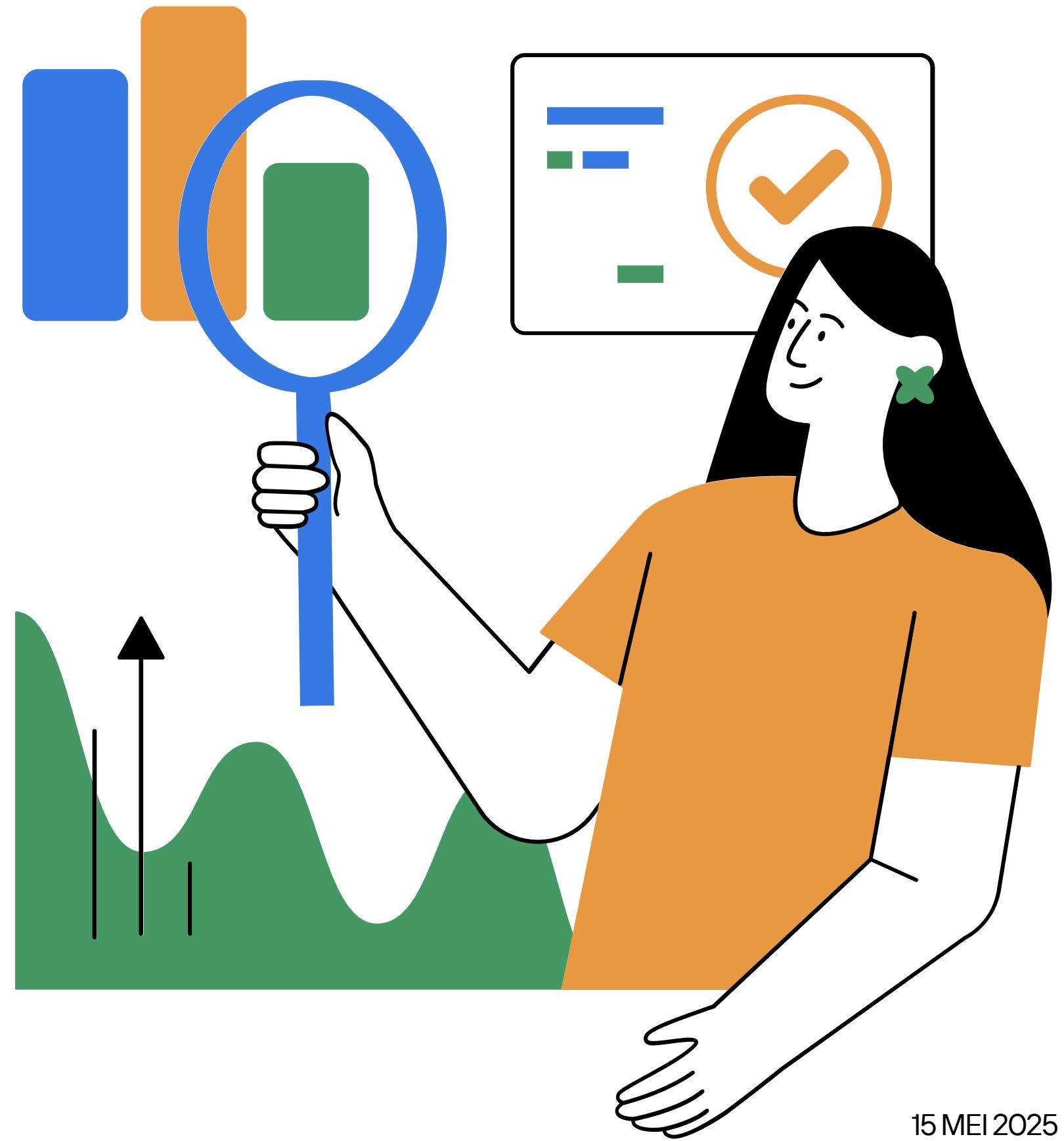
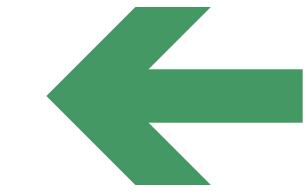


Prediction Model

ID/X Partners - Data
Scientist

Presented by:
Anis Widya Astuti





Anis Widya Astuti

Robotics and Artificial Intelligence Engineering student at Universitas Airlangga, currently completing final academic requirements. Experienced in AI-driven projects such as object detection, RPA automation, and sensor-based data processing. Previously worked as an AI Engineer Intern at PLN Icon Plus, focusing on computer vision and model optimization. Skilled in Python and data handling, with a strong interest in turning data into smart, real-world solutions.



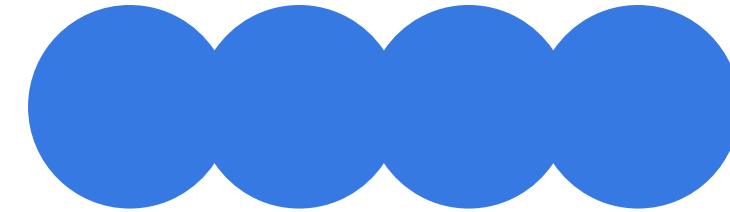
Surabaya, East Java



aniswidayastuti81@gmail.com



linkedin.com/in/aniswidyaastuti

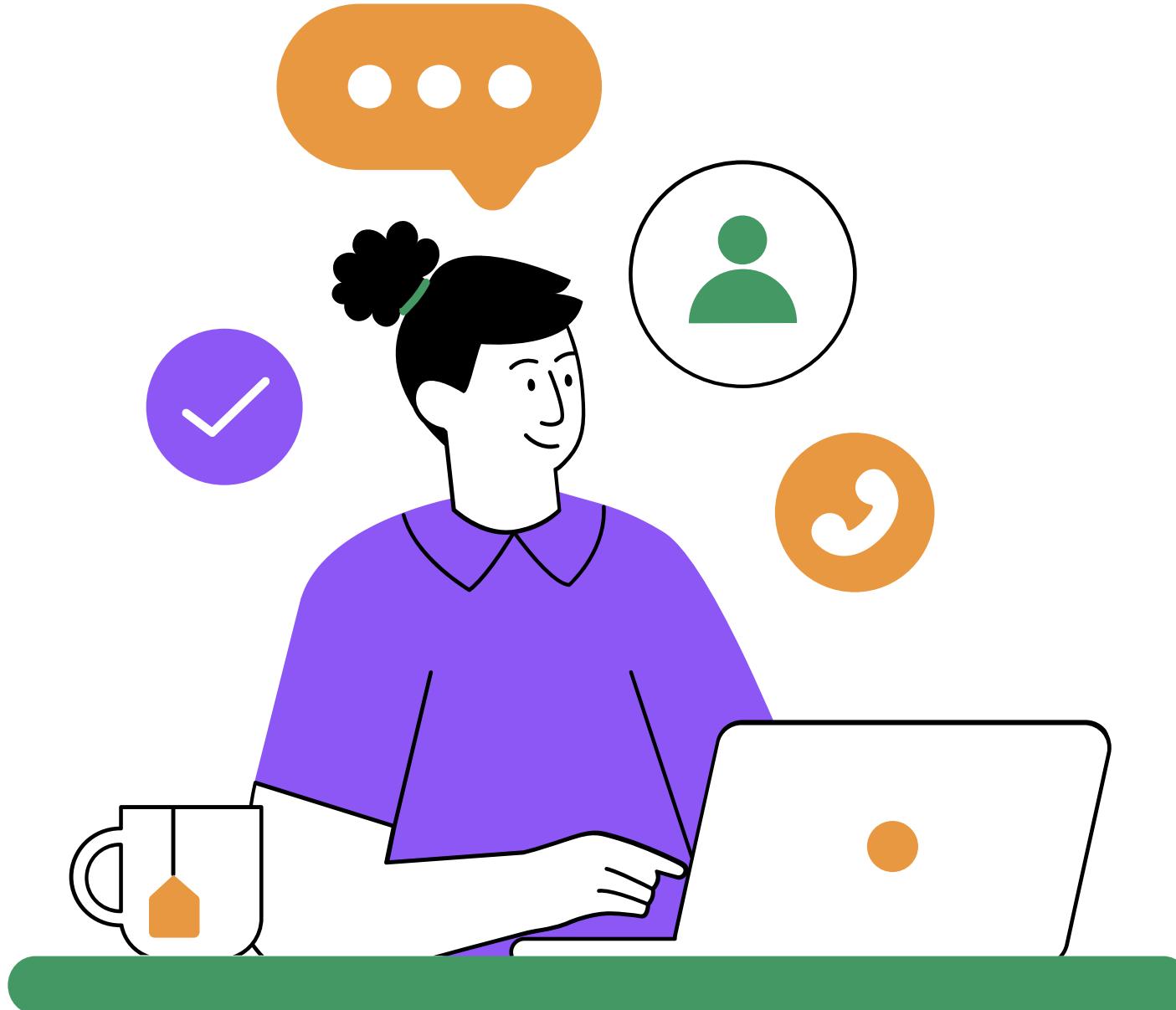


Course and Certification

- **Generative AI with Large Language Models by AWS and DeepLearning.AI**
- **Reinforcement Learning for Human Feedback by DeepLearning.AI**
- **DIGIFEST - Data Science #36**
- **ASEAN Data Science Explorer**



About ID/X Partner



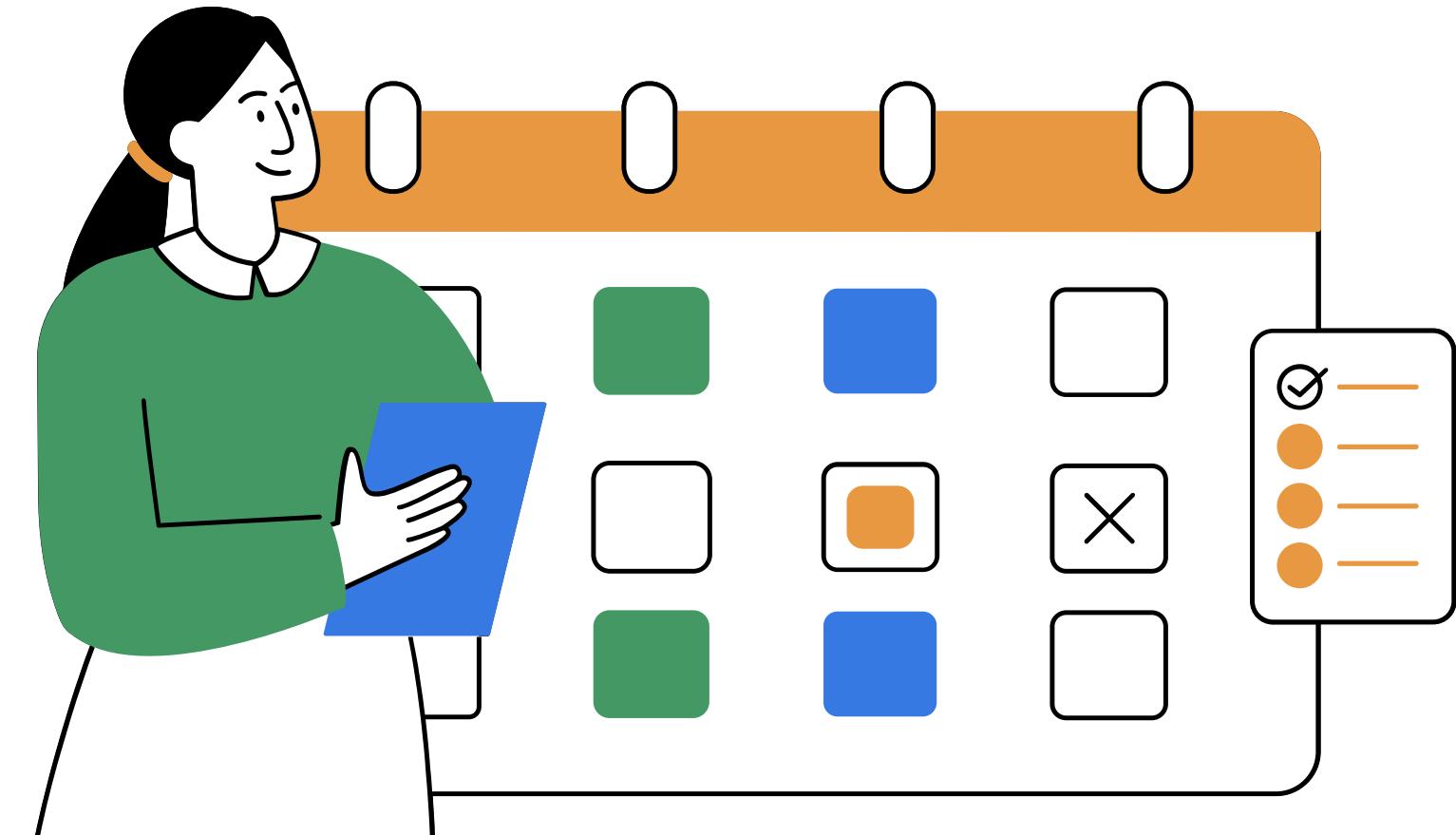
id/x partners is a technology consulting firm specializing in data analytics and decisioning solutions. With a strong focus on financial services, they help businesses optimize profitability and decision-making through AI, machine learning, and digital transformation. As an IBM partner, id/x partners combines deep domain expertise with innovative technologies to deliver impactful, data-driven strategies.



Tools & Technologies

Problem Statement

Platform peer-to-peer lending menghadapi risiko kerugian besar ketika peminjam gagal membayar ("Charged Off"). Dengan data historis pinjaman periode 2007–2014—termasuk jumlah pinjaman, tenor, suku bunga, pendapatan, dan grade peminjam—kita membangun model klasifikasi untuk memprediksi kegagalan bayar. Tujuannya adalah meningkatkan recall pada kelas "default" agar lebih sedikit pinjaman macet, sambil mempertahankan kemampuan diskriminasi (ROC-AUC) yang tinggi, sehingga dapat mendukung keputusan underwriting dan mitigasi risiko secara proaktif.



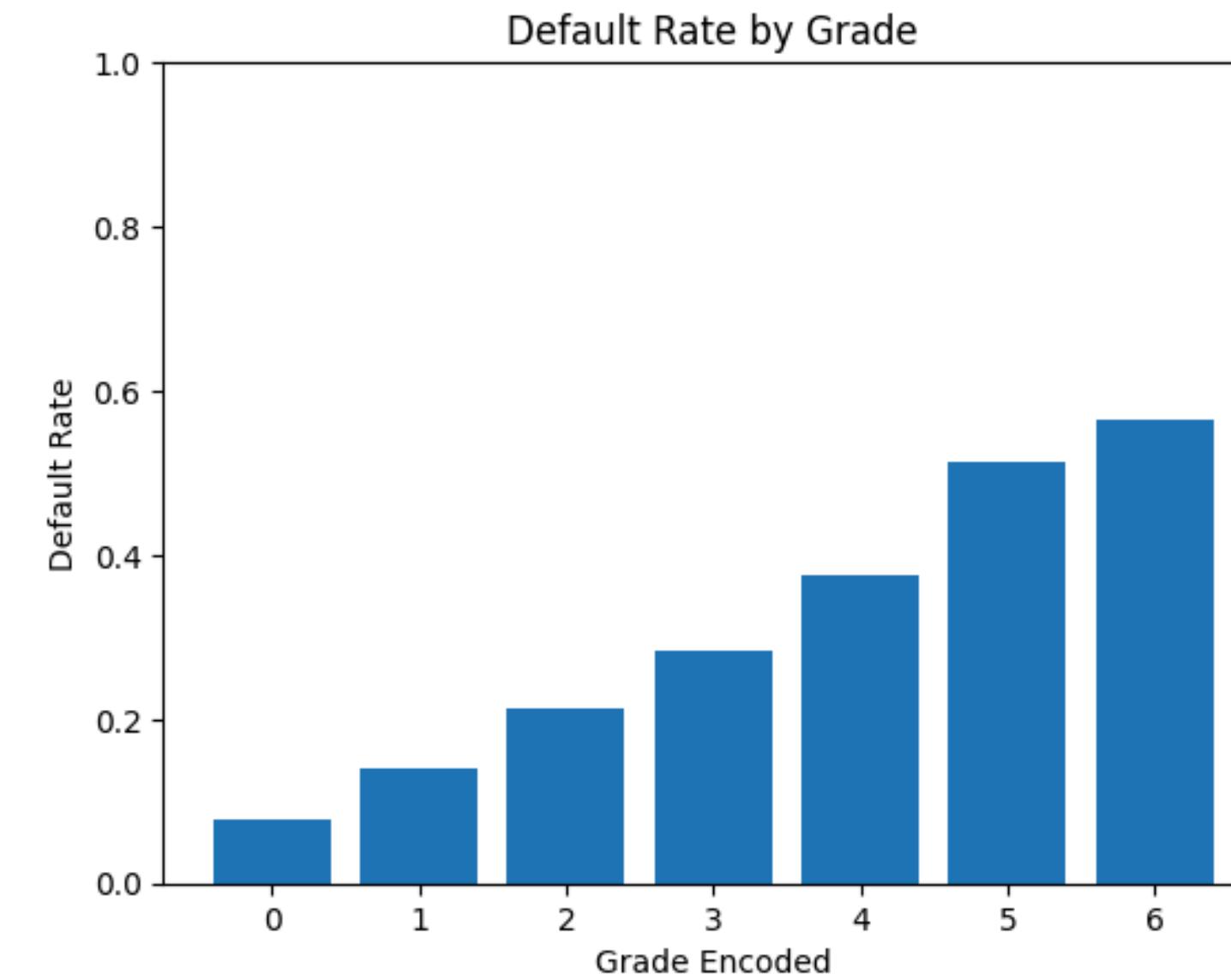
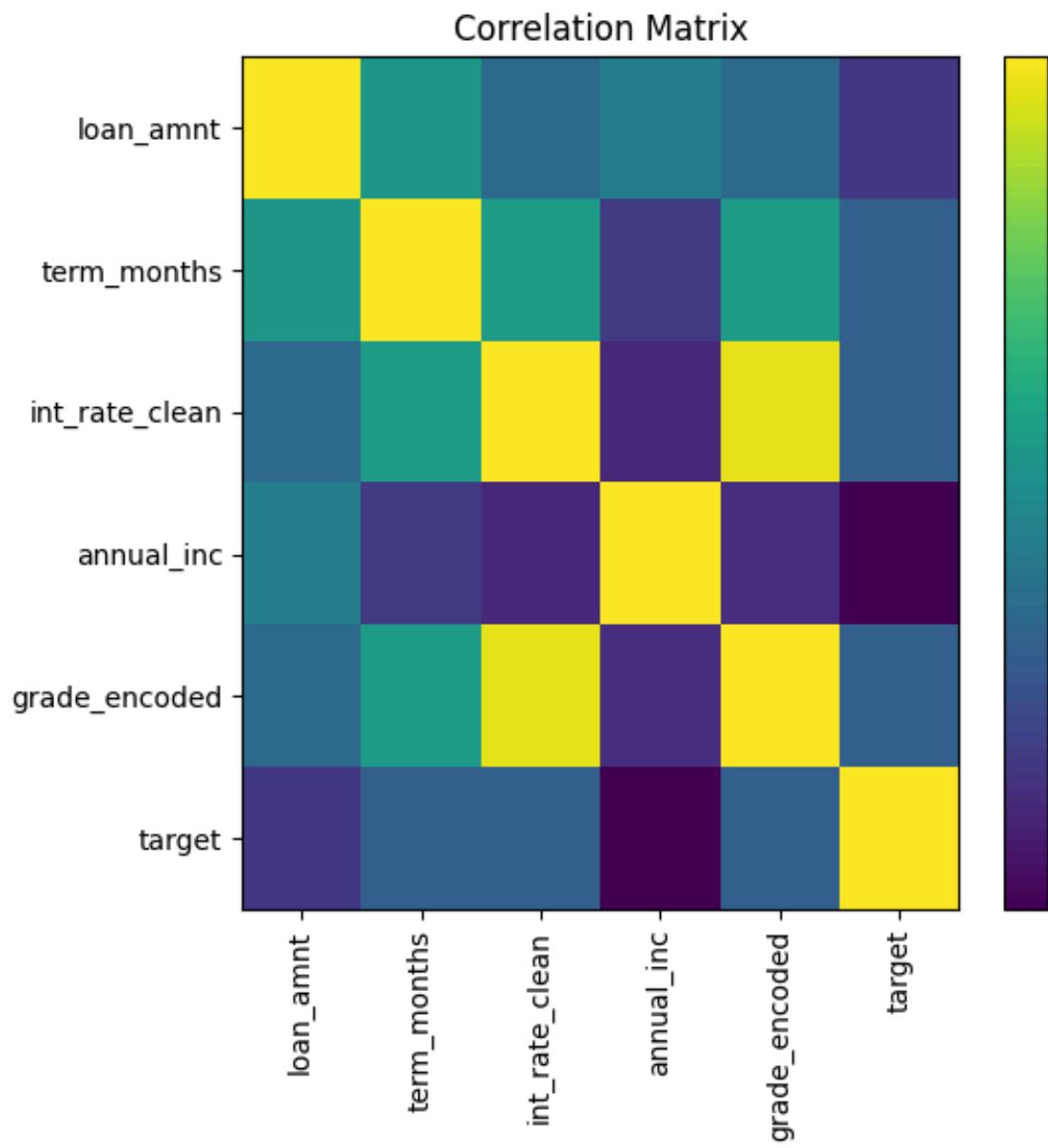
Data Overview

```
Data shape: (8580, 75)
Unnamed: 0          int64
id                  int64
member_id           int64
loan_amnt           int64
funded_amnt         int64
...
all_util            float64
total_rev_hi_lim   float64
inq_fi              float64
total_cu_tl          float64
inq_last_12m        float64
Length: 75, dtype: object
```

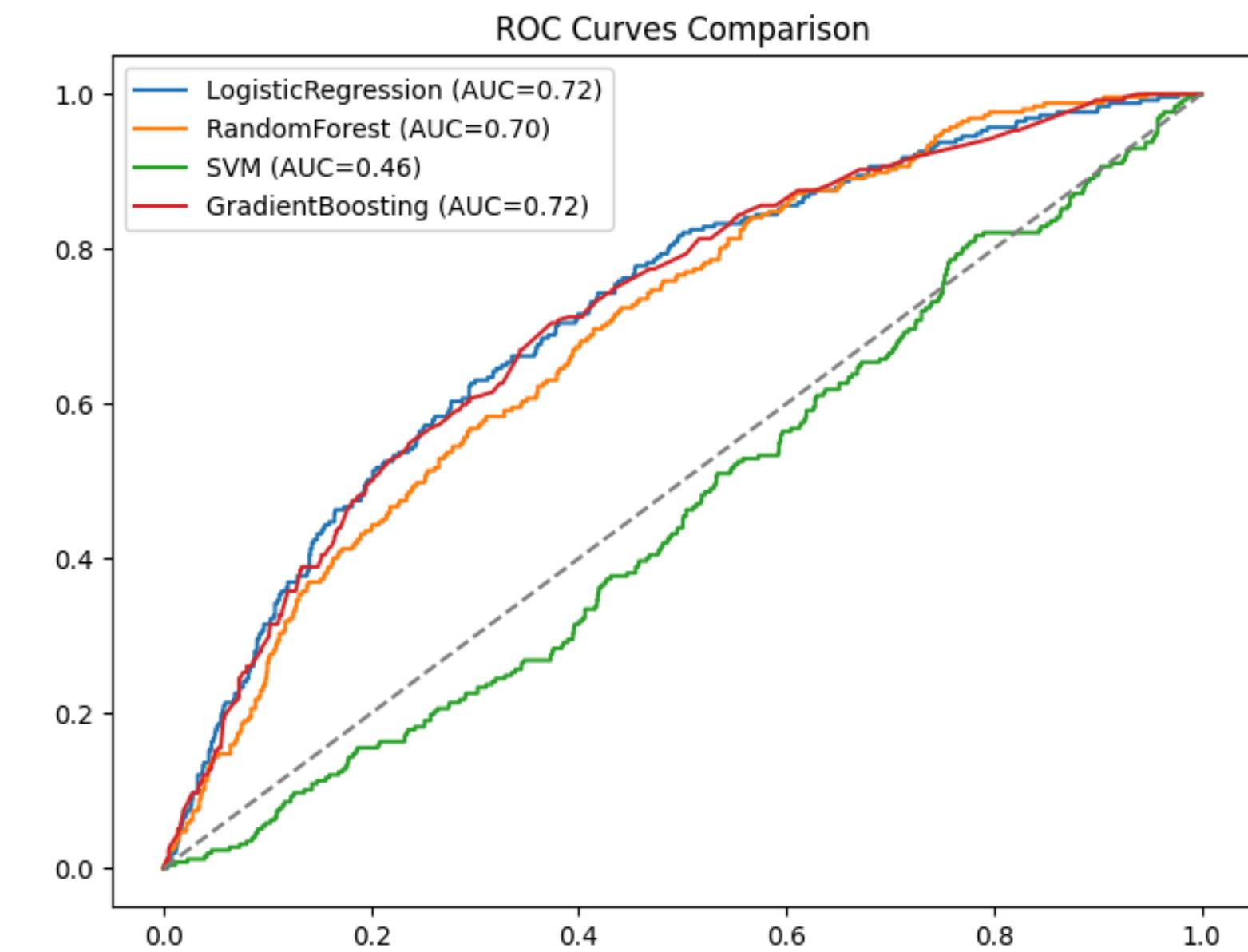
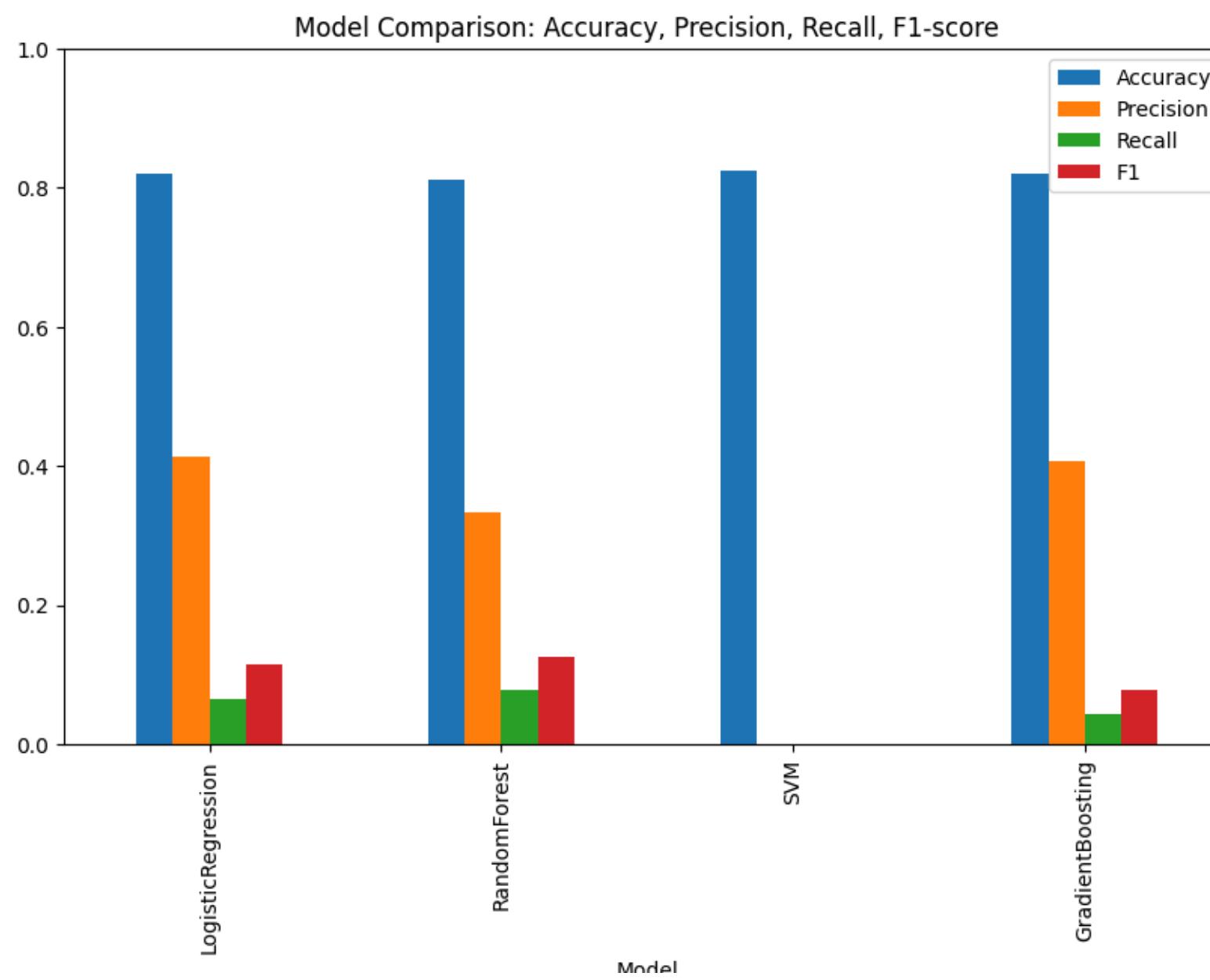
Pre-processing

Definisi Target
Imputasi
Cleaning
Outlier
Encoding
Split

Exploratory Data Analysis (EDA)

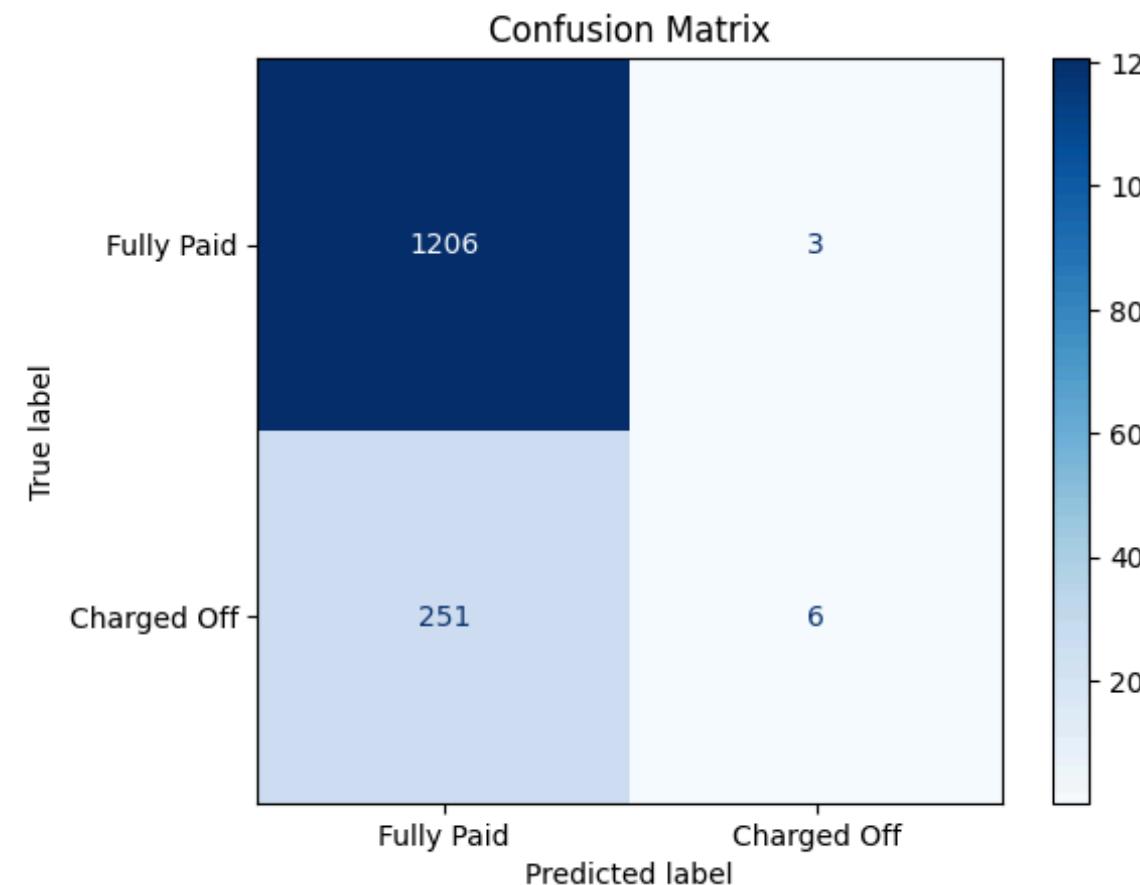


Modeling Approach



Key Insight

```
Classification Report for GB:  
precision    recall    f1-score   support  
  
          0       0.83      1.00      0.90     1209  
          1       0.67      0.02      0.05      257  
  
accuracy                           0.83     1466  
macro avg       0.75      0.51      0.47     1466  
weighted avg     0.80      0.83      0.75     1466
```



Predicted default proportion: 0.01, actual: 0.18

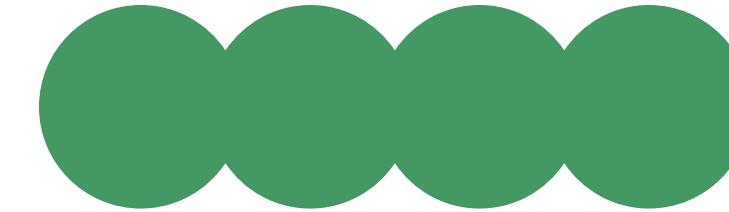
Top 3 features driving default: {'term_months': 0.541381024288182, 'int_rate_clean': 0.26104640563970044, 'annual_inc': 0.16155514334579277}

DEPLOYMENT

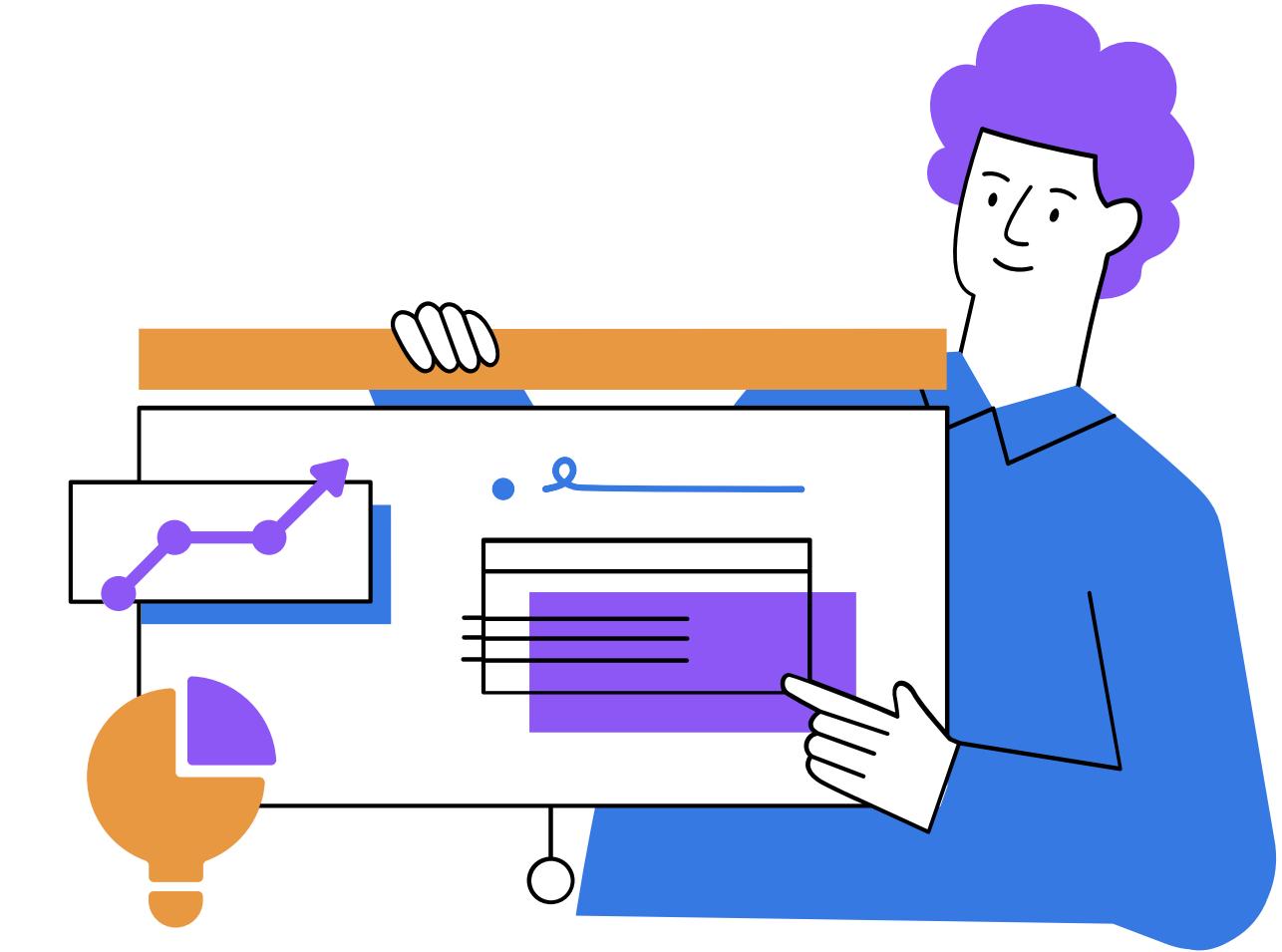
✓ DEPLOYMENT

```
[13] final_model = best_models['GradientBoosting']
     final_model.fit(x, y)
     joblib.dump(final_model, 'final_credit_risk_model.pkl')
     print("Saved final_credit_risk_model.pkl for deployment")
```

→ Saved final_credit_risk_model.pkl for deployment



CONCLUSION



Meskipun model Gradient Boosting menunjukkan akurasi keseluruhan sebesar 83 %, ia gagal mengenali mayoritas kasus gagal bayar: recall untuk kelas "Charged Off" hanya 2 %, artinya dari 257 peminjam yang benar-benar macet hanya 6 yang berhasil diprediksi. Kontrasnya, model sangat andal mengidentifikasi peminjam yang lunas (recall 100 %), tetapi proporsi prediksi default (1 %) jauh di bawah proporsi aktualnya (18 %), menandakan bias kuat ke kelas mayoritas.

Analisis feature importance mengungkap tenor pinjaman (term_months), suku bunga (int_rate_clean), dan pendapatan tahunan (annual_inc) sebagai tiga faktor paling berpengaruh. Untuk meningkatkan kemampuan deteksi default, disarankan menerapkan teknik penanganan ketidakseimbangan kelas—seperti SMOTE atau pengaturan `class_weight='balanced'`—serta menyesuaikan threshold probabilitas. Selain itu, menambah variabel seperti debt-to-income ratio atau skor kredit historis dapat memperkaya sinyal risiko dan meningkatkan performa model pada kelas minoritas.

Thank You

- aniswidayastuti81@gmail.com
- <https://github.com/aniswidyaastuti>

