

CREDIT RISK ANALYSIS USING XGBOOST AND LIGHTGBM MODELS

**Final Project Internship
ID/X Partners x Rakamin**

A predictive dashboard to identify high-risk customers for early credit default detection

Angga Yulian Adi Pradana

Github Link: <https://github.com/anggapradanaa/Credit-Risk-Analysis-Using-XGBoost-and-LightGBM-Models>



PROJECT SUMMARY



OBJECTIVE

Membangun sistem prediksi risiko kredit secara dini menggunakan algoritma machine learning XGBoost dan LightGBM untuk mengidentifikasi potensi gagal bayar



DATASET

- 466285 aplikasi kredit
- 75 fitur
- Periode pengajuan: 2007-2014
- Target: Default vs Non-Default

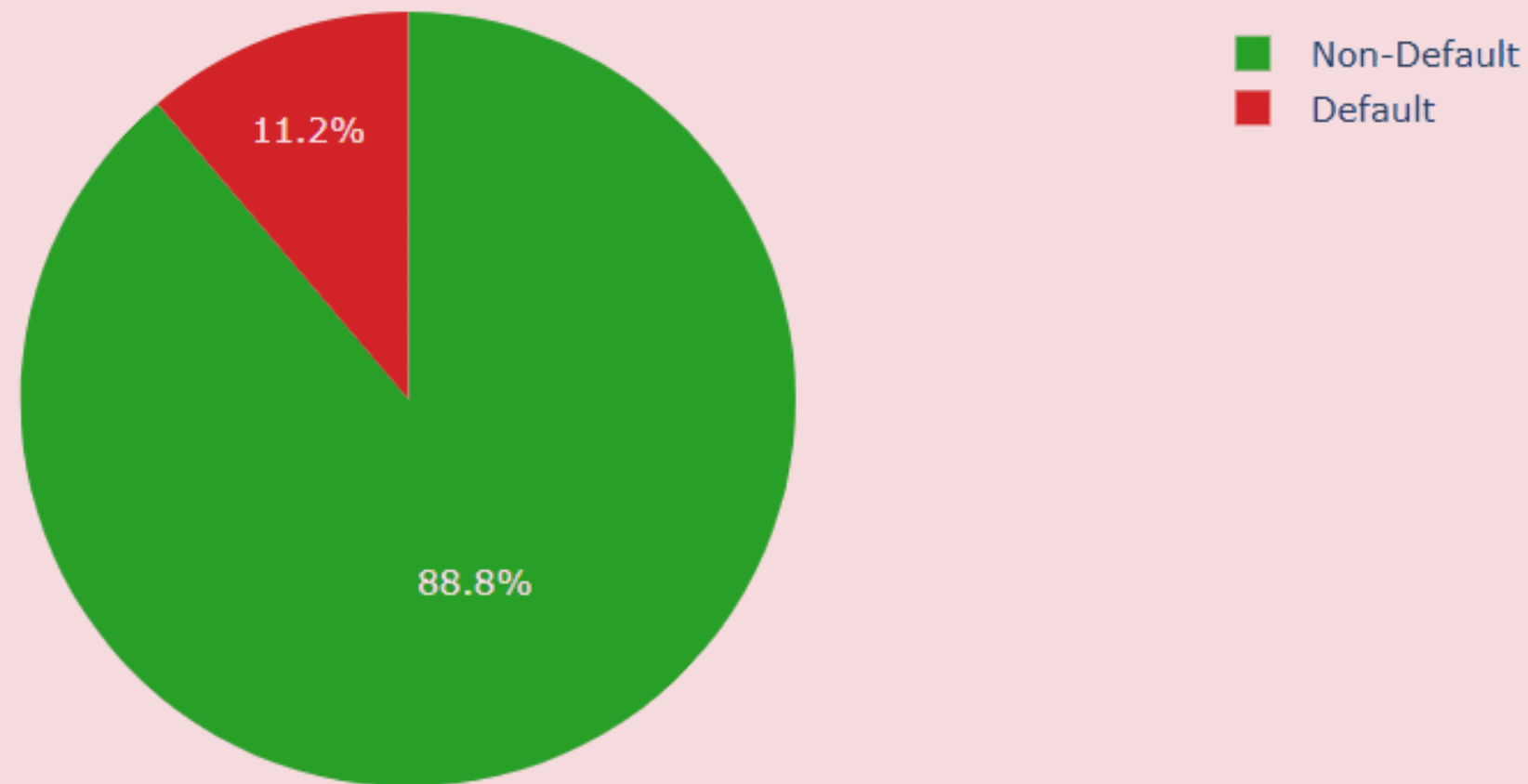


SUCCESS METRICS

- ROC AUC
- Classification Report
- Feature Importance
- Interactive Dashboard

DATA DISTRIBUTION OVERVIEW

Distribusi Kolom Target



KEY INSIGHTS:

- Non-Default: 88.8%
- Default: 11.2%
- Action: Implementasi SMOTE untuk balancing

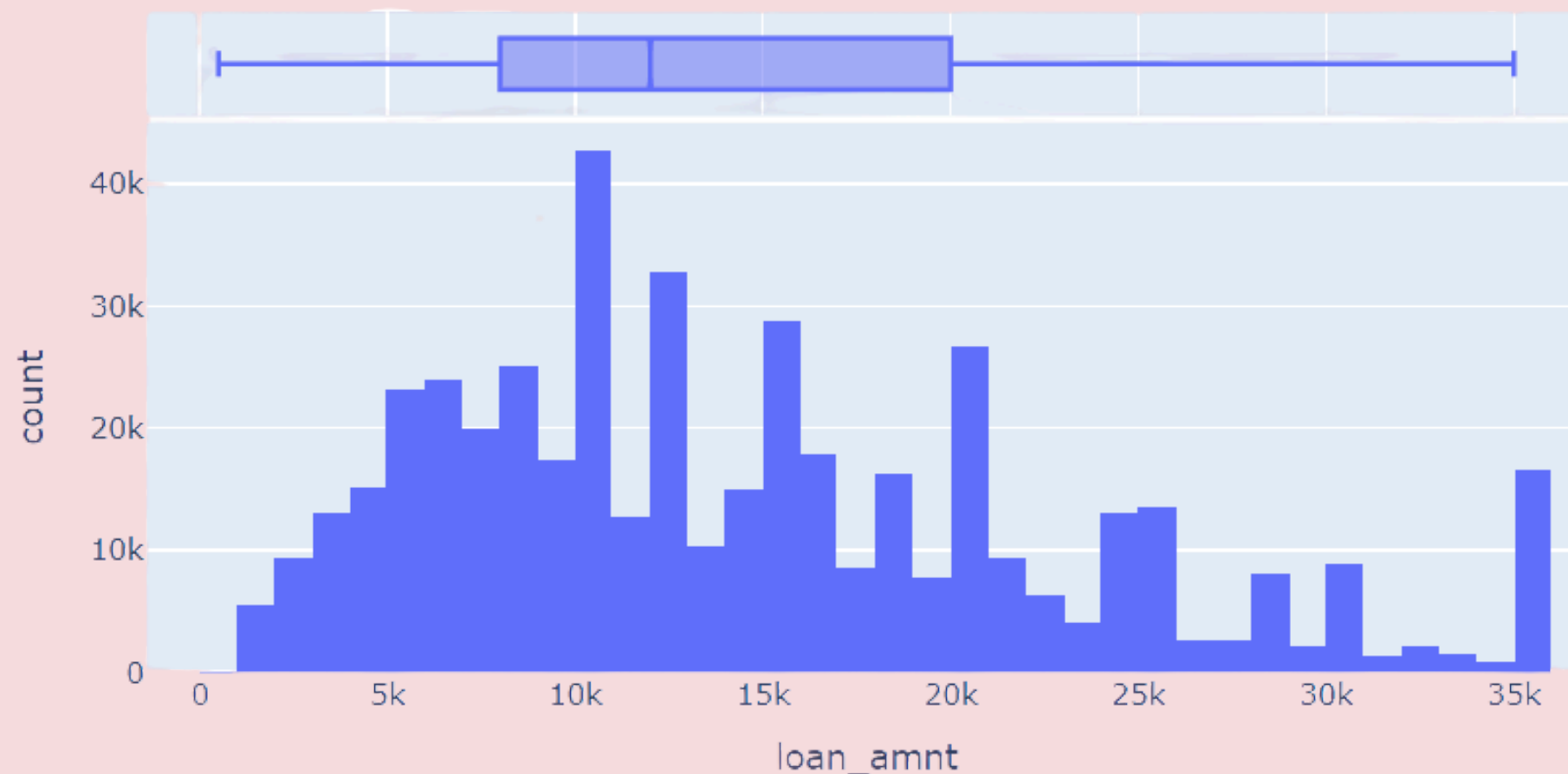


BUSINESS IMPLICATION:

- Default rate 11.2% tetap signifikan
- Deteksi dini tetap diperlukan
- Risk-based pricing tetap relevan

LOAN AMOUNT ANALYSIS

Histogram-Distribusi Loan Amount



STATISTICAL SUMMARY

- Median: \$12K
- Q1: \$8K
- Q3: \$20K
- Min: \$0.5K
- Max: \$35K
- Range: \$0.5K – \$35K
- Distribution: Positively skewed

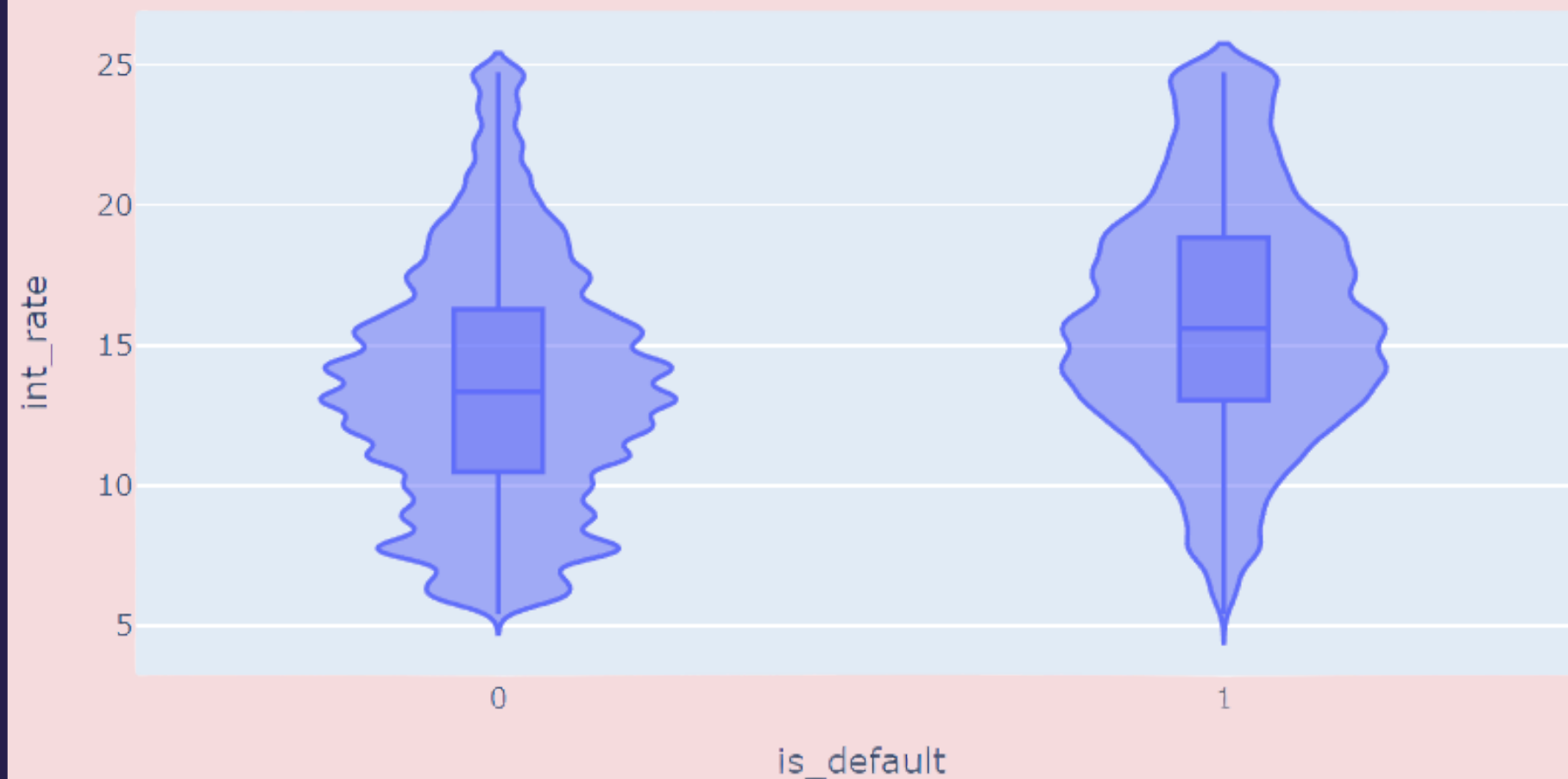


PATTERN INSIGHTS

- Pinjaman terbanyak di rentang \$5K–\$20K, dengan puncak di \$10K dan \$15K
- \$35K cukup menonjol, tapi jarang diajukan
- Pinjaman kecil (< \$5K) relatif sedikit
- Preferensi umum: \$10K–\$15K (sweet spot)
- Tidak ditemukan outlier ekstrem (max = upper fence)

INTEREST RATE VS DEFAULT RISK

Violin Plot-Interest Rate vs Default Status



STATISTICAL SUMMARY

- Median (Non-Default): 13.35%
- Median (Default): 15.61%
- Range: 5.42% – 24.74%
- Distribusi: Bunga peminjam default lebih tinggi

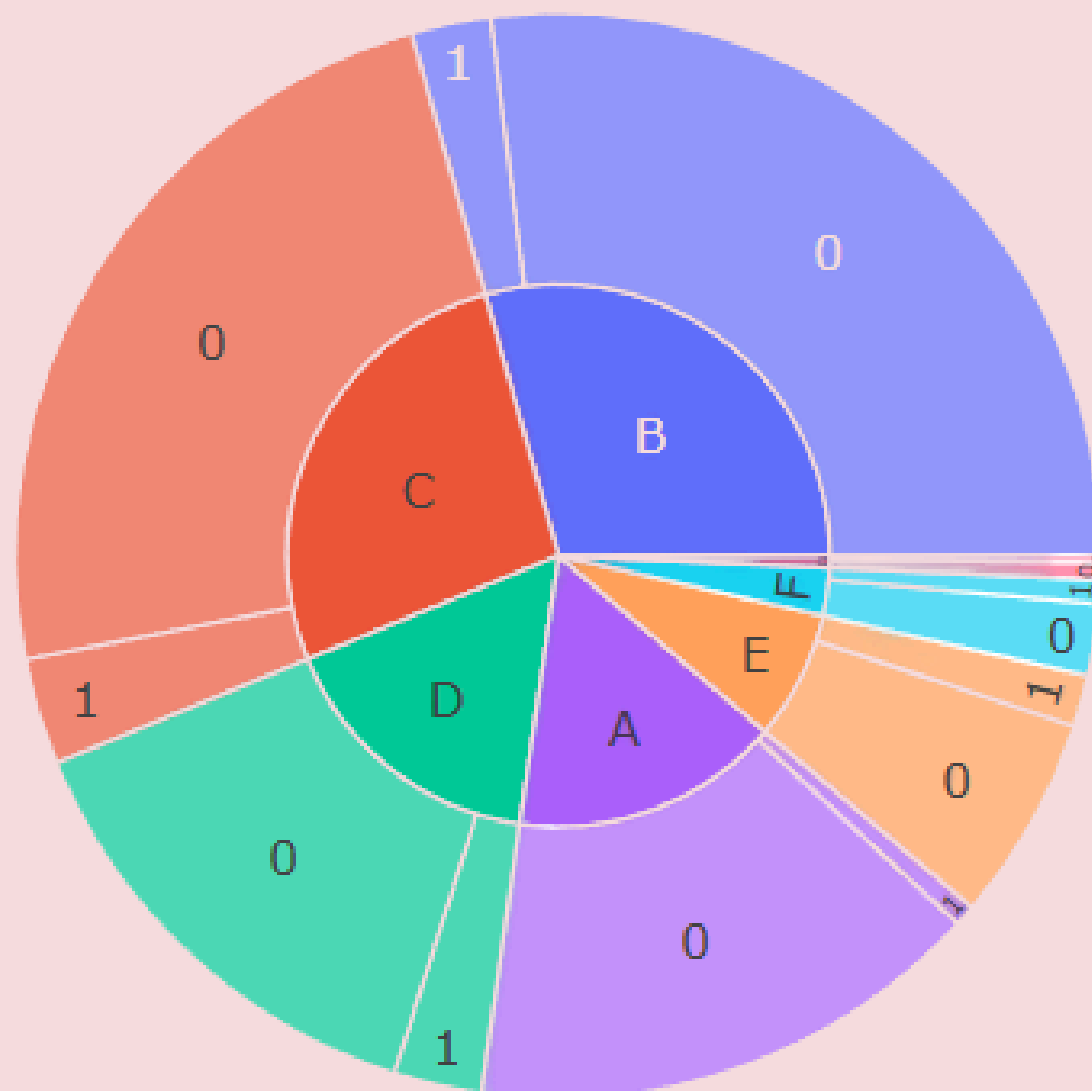


PATTERN INSIGHTS

- Borrower default konsisten punya bunga lebih tinggi
- Median bunga naik ~2.3% dari non-default ke default
- Titik kritis risiko tampak di >15%
- Zona risiko tinggi: bunga >20%
- Bunga efektif jadi prediktor kuat default

DATA DISTRIBUTION OVERVIEW

Distribusi Grade

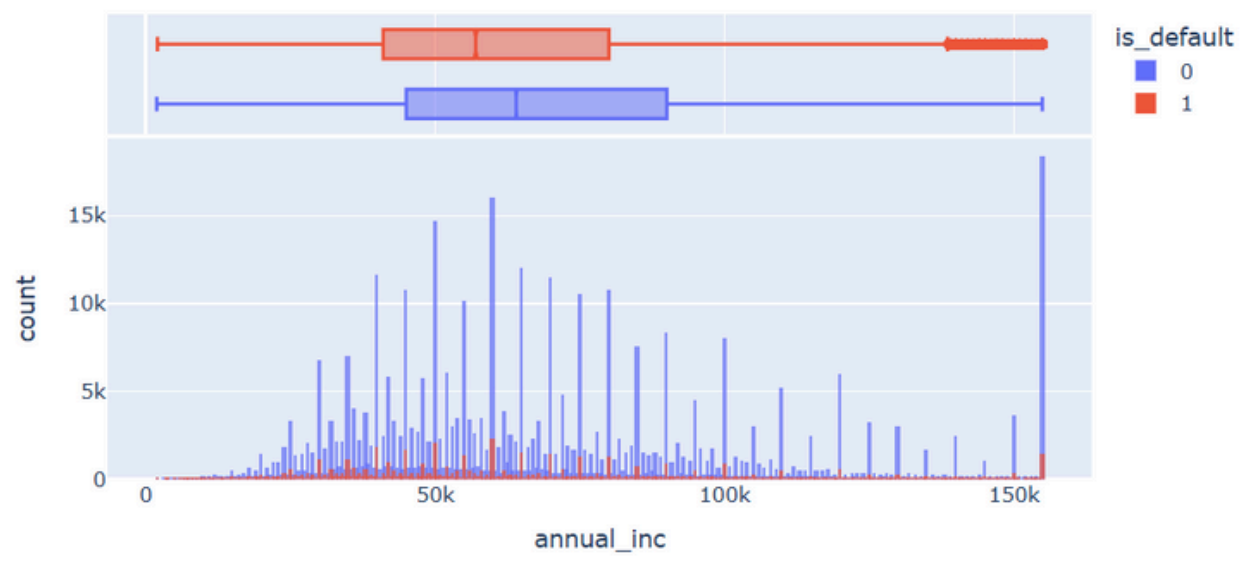


KEY INSIGHTS:

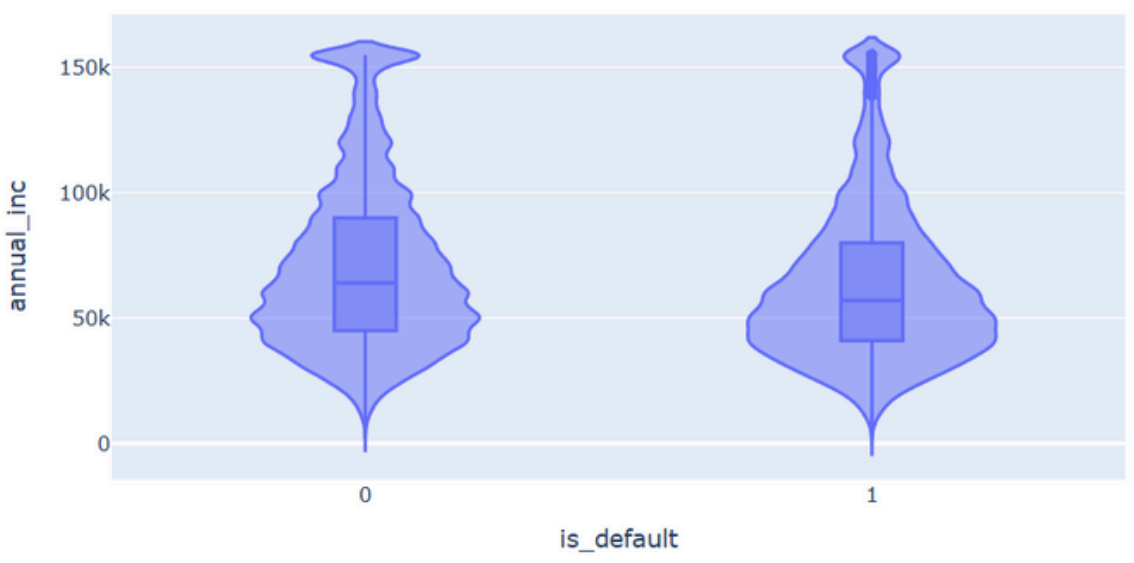
- Grade A dan B didominasi oleh peminjam dengan status non-default, mengindikasikan profil risiko yang relatif rendah.
- Grade C dan D menunjukkan peningkatan proporsi peminjam yang mengalami default, mencerminkan risiko yang lebih tinggi.
- Meskipun jumlahnya lebih kecil, peminjam pada Grade E ke bawah memiliki proporsi default yang signifikan, sehingga perlu perhatian khusus.
- Secara keseluruhan, terdapat pola yang konsisten: semakin rendah grade, semakin tinggi kecenderungan untuk gagal bayar, menjadikan grade sebagai indikator risiko yang relevan.

DEFAULT RISK BY ANNUAL INCOME AND DTI

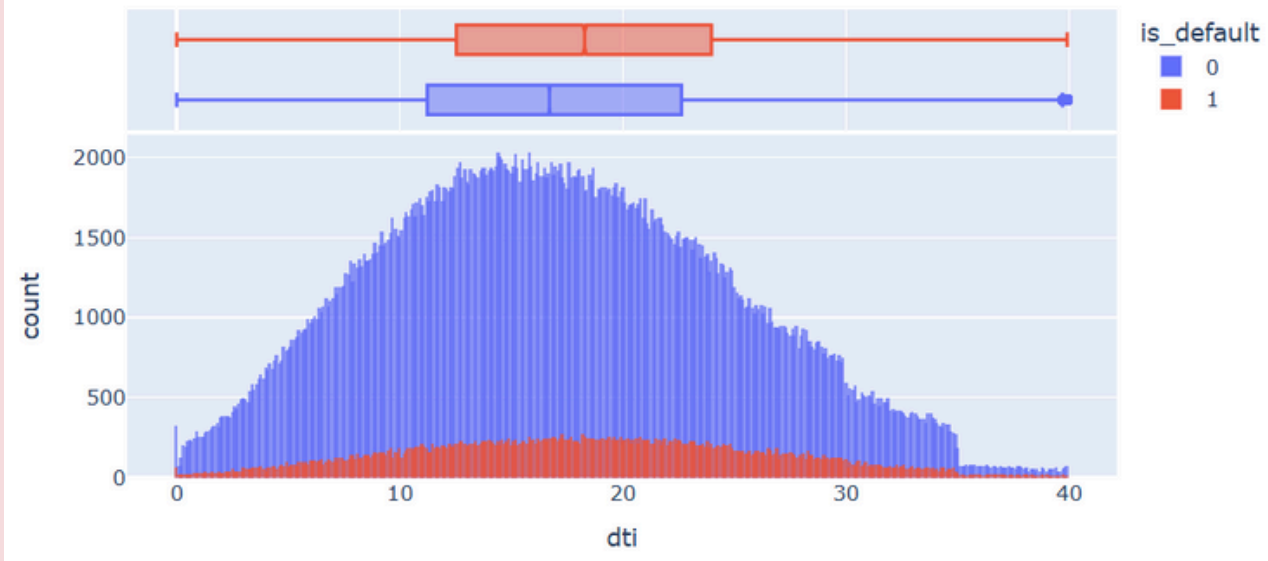
Distribusi annual_inc berdasarkan Status Default



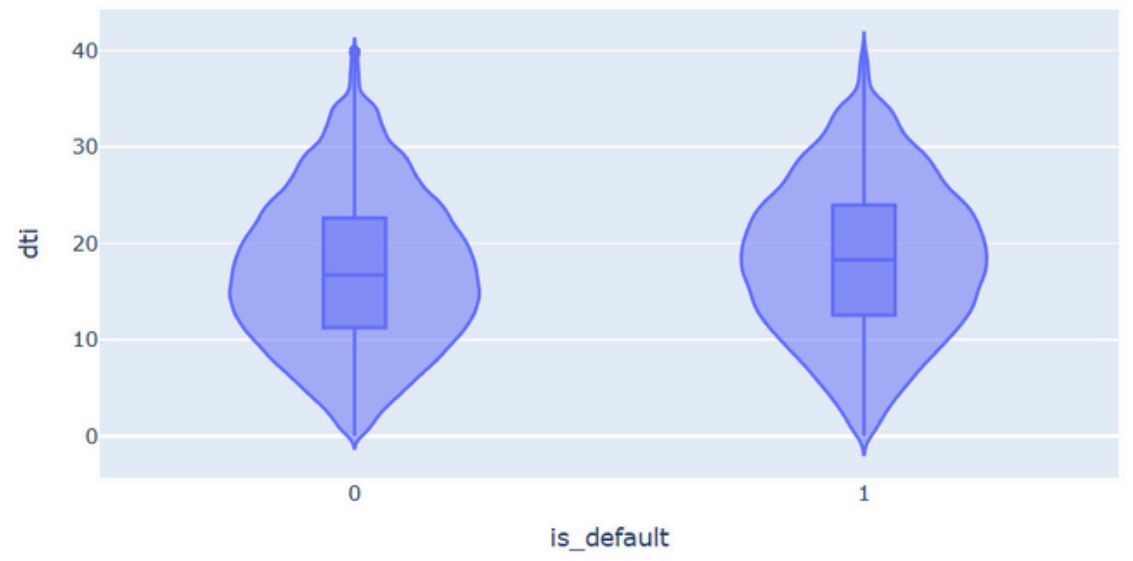
Violin Plot annual_inc berdasarkan Status Default



Distribusi dti berdasarkan Status Default

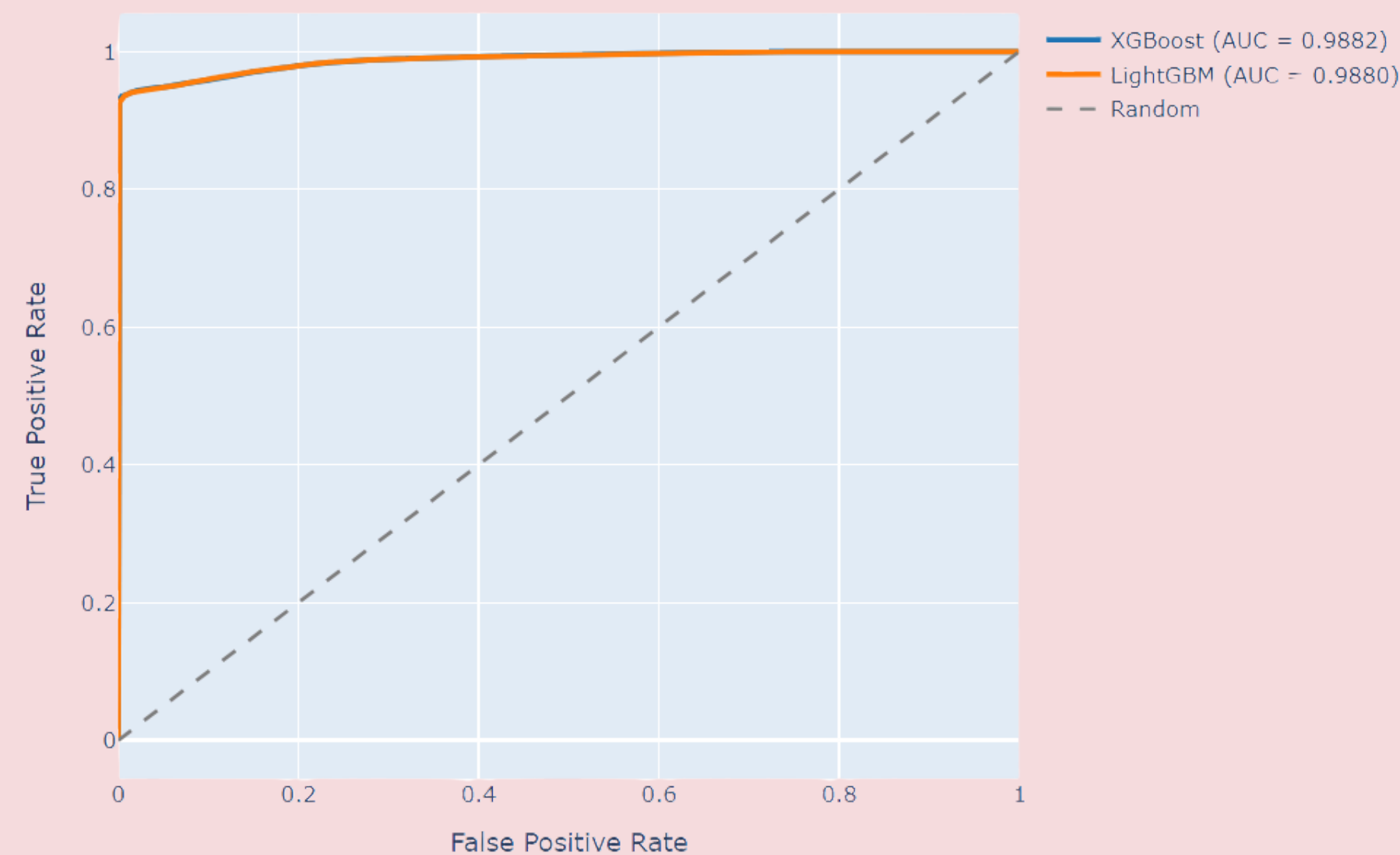


Violin Plot dti berdasarkan Status Default



MODEL PERFORMANCE BATTLE

ROC Curve Comparison



| XGBoost | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 1.00 | 0.99 | 82820 |
| 1 | 0.98 | 0.93 | 0.96 | 10437 |
| accuracy | | | 0.99 | 93257 |
| macro avg | 0.99 | 0.96 | 0.98 | 93257 |
| weighted avg | 0.99 | 0.99 | 0.99 | 93257 |

| LightGBM | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 1.00 | 0.99 | 82820 |
| 1 | 0.97 | 0.93 | 0.95 | 10437 |
| accuracy | | | 0.99 | 93257 |
| macro avg | 0.98 | 0.96 | 0.97 | 93257 |
| weighted avg | 0.99 | 0.99 | 0.99 | 93257 |

- XGBoost menang tipis di semua metrics
- Kedua model menunjukkan excellent performance

CONFUSSION MATRIX ANALYSIS

XGBoost Confussion Matrix

| | | |
|--------------------|-----------------------|-------------------|
| Actual Default | 708 | 9729 |
| Actual Non-Default | 82626 | 194 |
| | Predicted Non-Default | Predicted Default |

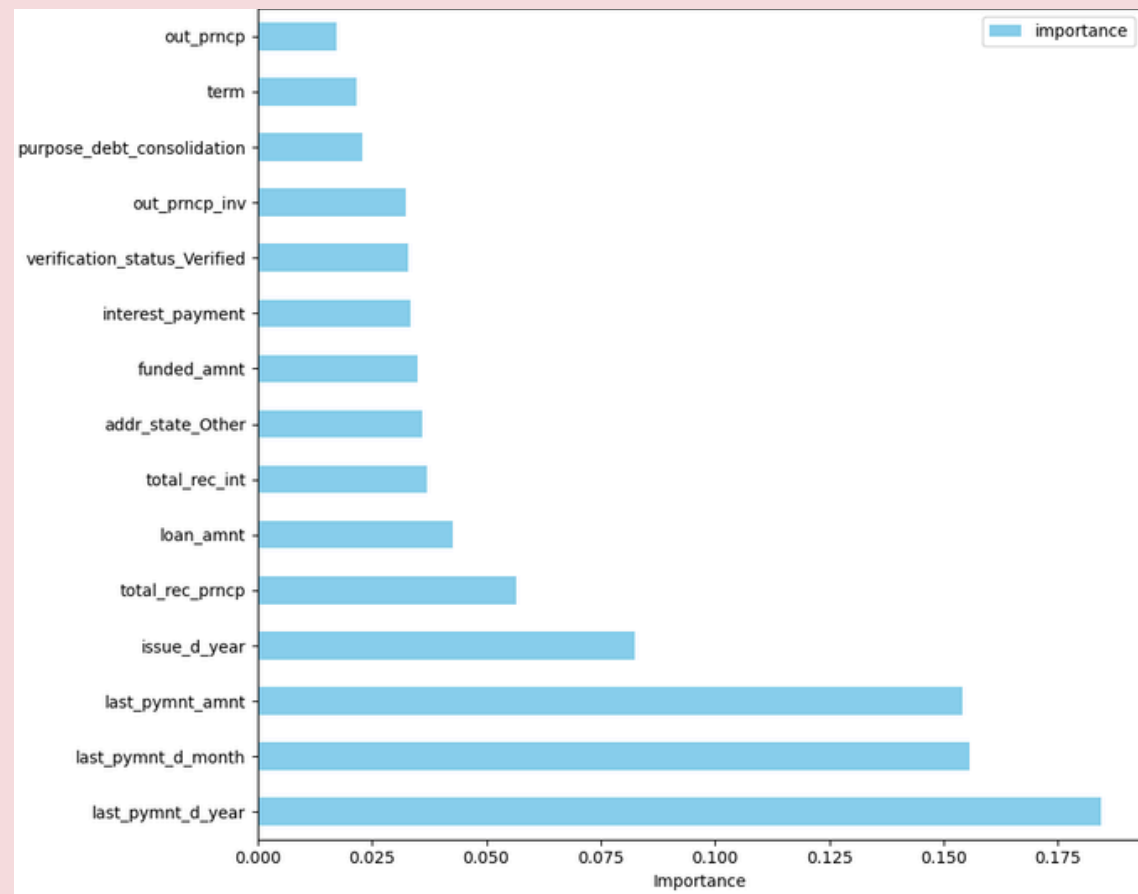
LightGBM Confussion Matrix

| | | |
|--------------------|-----------------------|-------------------|
| Actual Default | 727 | 9710 |
| Actual Non-Default | 82553 | 267 |
| | Predicted Non-Default | Predicted Default |

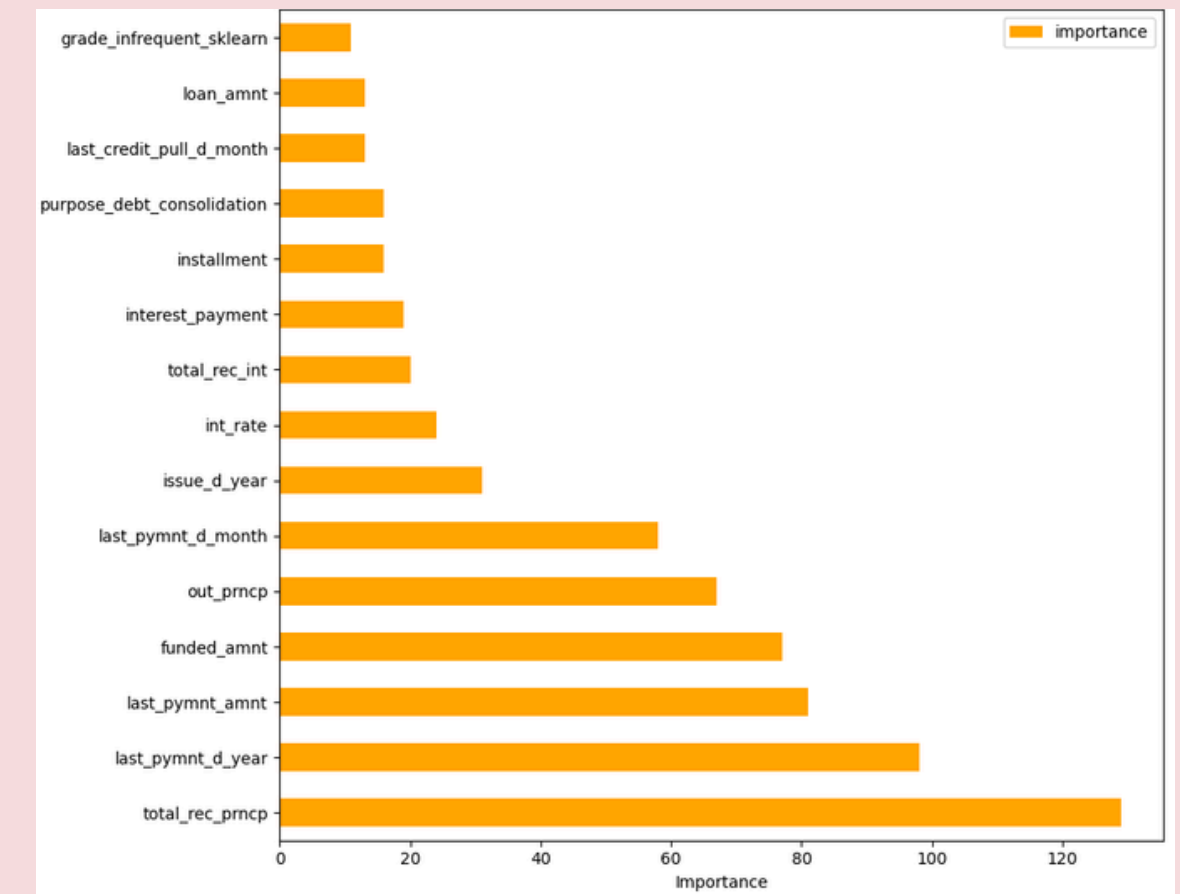
- LightGBM mendeteksi lebih banyak nasabah berisiko gagal bayar (TP = 727) dibanding XGBoost (TP = 708), menjadikannya lebih andal untuk kebutuhan deteksi dini.
- False negative lebih rendah pada LightGBM (9710 vs 9729), artinya LightGBM sedikit lebih baik dalam mengurangi risiko kelolosan nasabah gagal bayar.
- XGBoost lebih presisi namun lebih konservatif, dengan false positive lebih rendah (194 vs 267), namun berpotensi melewatkan lebih banyak nasabah default.

FEATURE IMPORTANCE

XGBoost Feature Importance



LightGBM Feature Importance



- Riwayat pembayaran merupakan sinyal utama dalam prediksi gagal bayar, di mana tanggal dan jumlah pembayaran terakhir terbukti sangat menentukan
- Pelunasan pokok pinjaman juga menjadi indikator kuat, di mana total pelunasan dan sisa pokok pinjaman yang kecil cenderung menunjukkan risiko gagal bayar yang rendah
- Nominal pinjaman dan beban bunga turut memperkuat prediksi risiko kredit
- Dibandingkan XGBoost, model LightGBM mengevaluasi lebih banyak aspek tambahan seperti cicilan bulanan, kategori peminjam, dan data penarikan kredit terakhir
- XGBoost cenderung lebih fokus dan presisi, hanya mengandalkan sedikit fitur utama dengan kontribusi besar terhadap keputusan model

CONCLUSION


Sistem prediksi risiko kredit berhasil dibangun menggunakan dua algoritma andal: XGBoost dan LightGBM

Kedua model mencapai akurasi sangat tinggi sebesar 99%, menunjukkan kemampuan klasifikasi yang sangat baik terhadap nasabah default dan non-default

XGBoost menunjukkan performa lebih presisi terhadap nasabah default dengan precision 0.98, serta false positive lebih rendah (194) dibanding LightGBM

LightGBM memiliki recall lebih tinggi (0.93 vs 0.93 dengan TP = 727 vs 708), menandakan kemampuan lebih baik dalam mendeteksi lebih banyak nasabah berisiko (true positive lebih tinggi). Namun, LightGBM juga menghasilkan lebih banyak false positive (267) dibanding XGBoost.

RECOMMENDATION



Gunakan LightGBM bila fokus utama adalah deteksi dini risiko dan minimalkan nasabah gagal bayar yang lolos, meski dengan sedikit peningkatan false positive

Gunakan XGBoost jika prioritas adalah mengurangi kesalahan deteksi pada nasabah sehat dan menjaga presisi keputusan kredit

Implementasikan dashboard interaktif sebagai alat bantu analis kredit untuk memonitor prediksi dan membuat keputusan lebih cepat dan tepat

Evaluasi berkala tetap diperlukan untuk memastikan model tetap relevan terhadap dinamika dan pola data nasabah terbaru



THANK YOU

Contact:

Email: anggayulianadipradana@gmail.com

Phone/WA: +62 823 6103 7618

LinkedIn: [Angga Yulian Adi Pradana](#)

Github: [anggapradanaa](#)