Loan Default Analysis:

By Nagib Gonzalez

Overview:

- Chase Consumer Bank
- Data analysis needed to identify factors influencing loan defaults
- Goal is to develop marketing plan to minimize losses using data science techniques
- Binary classification problem to identify loan default risk

Data Understanding:

- Kaggle dataset with training and testing data
- 35 columns, including loan amount, interest rate, employment length, home ownership, and loan status
- 67,000 rows (clients)

Business Problem:

Will clients default on their loans or not?

Binary Classification:

0 = Not In Default 1 = In default

Focus on reducing false positives:

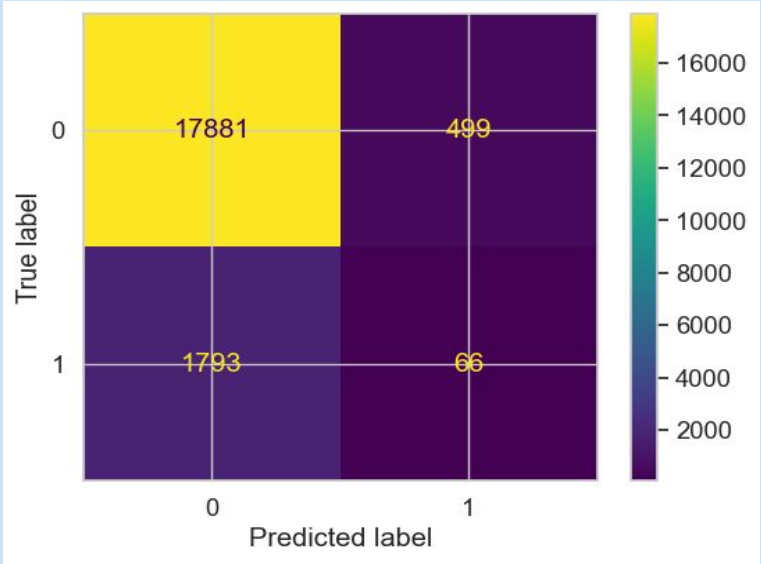
- Model predicts loan default when there is none
- Costly for bank and may harm relationships with borrowers

Results:

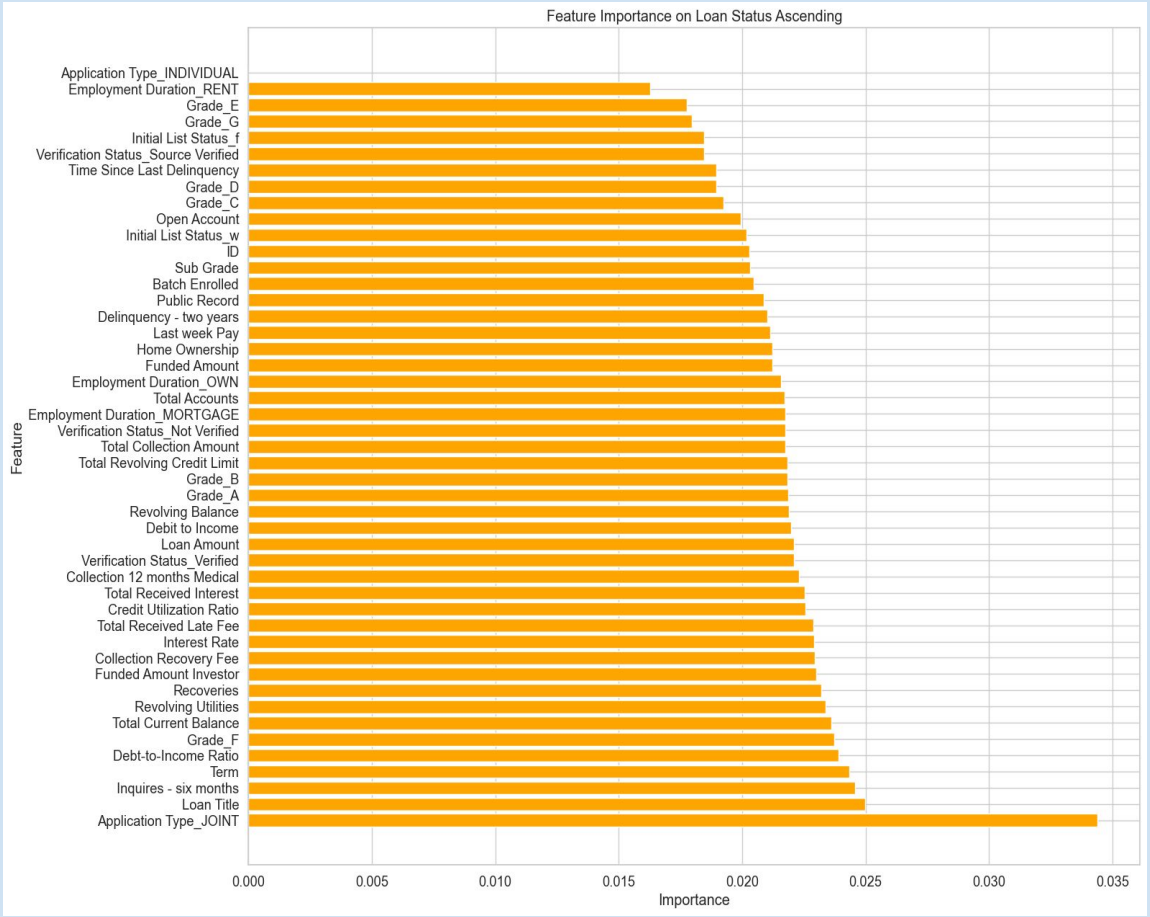
Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Final Model Results: XGBoost Hypertuned



Results:


Feature Importance on Loan Status Ascending
Application Type_INDIVIDUAL
Employment Duration_RENT
Grade_E
Grade_G
Initial List Status_f
Verification Status_Source Verified
Time Since Last Delinquency
Grade_D
Grade_C
Open Account
Initial List Status_w
ID
Sub Grade
Batch Enrolled
Public Record
Delinquency - two years
Last week Pay
Home Ownership
Funded Amount
Employment Duration_OWN
Total Accounts
Employment Duration_MORTGAGE
Verification Status_Not Verified
Total Collection Amount
Total Revolving Credit Limit
Grade_B
Grade_A
Revolving Balance
Debit to Income
Loan Amount
Verification Status_Verified
Collection 12 months Medical
Total Received Interest
Credit Utilization Ratio
Total Received Late Fee
Interest Rate
Collection Recovery Fee
Funded Amount Investor
Recoveries
Revolving Utilities
Total Current Balance
Grade_F
Debt-to-Income Ratio
Term
Inquires - six months
Loan Title
Application Type_JOINT
Feature
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
Importance

Recommendations:

1. Targeted advertising for joint application loans
2. Leverage the six-month inquiry feature
3. Highlight loan titles that perform well

Limitations And Further Analysis:

- Class imbalance affects model accuracy
- Historical data may not reflect current conditions
- Explore additional data sources or features
- Investigate reasons behind loan defaults
- More extensive hyperparameter tuning