

CREDIT SCORING

END - TO - END SOLUTION

By Ricardo Zefanya Damarjanan



TABLE OF CONTENTS



- BACKGROUND
- DATA UNDERSTANDING
- CHARACTERISTICS OF THE BORROWER
- PREPROCESSING
- MODELLING
- EVALUATION METRICS
- TABLE OF MODEL PERFORMANCES
- RESULT ANALYSIS
- GET IN TOUCH

BACKGROUND

Credit risk is the risk of loss associated with the possible failure of a counterparty to fulfill its obligations; or the risk that the debtor will not repay the debt. To reduce this credit risk, we need to have an assessment process using credit ratings and credit ratings for prospective borrowers where this assessment determines the lending institution to potential borrowers. By predicting credit risk from received and rejected loan data, we can also reduce this credit risk.



DATA UNDERSTANDING

- All data is of individual loan application type
- Data is taken from 2007 to 2014
- In the dataset there are 466,285 rows of data and 75 columns
- Each borrower can be marked whether they are a bad loan (0) or good loan (1)
- In this analysis, it is determined if the late payment is more than 30 days and worse than that as a bad loan marker.

CHARACTERISTICS OF THE BORROWER



LOAN STATUS

The loan status owned by the loan recipient and is the information that we will predict.



LOAN PURPOSE

The reason the debtor borrows money from the company



RANK LOAN

The loan rating is in the form of letters of the alphabet from A - G, the closer to G, the greater the interest rate



TOTAL LOAN

The amount of the loan received by the debtor.

LOAN STATUS



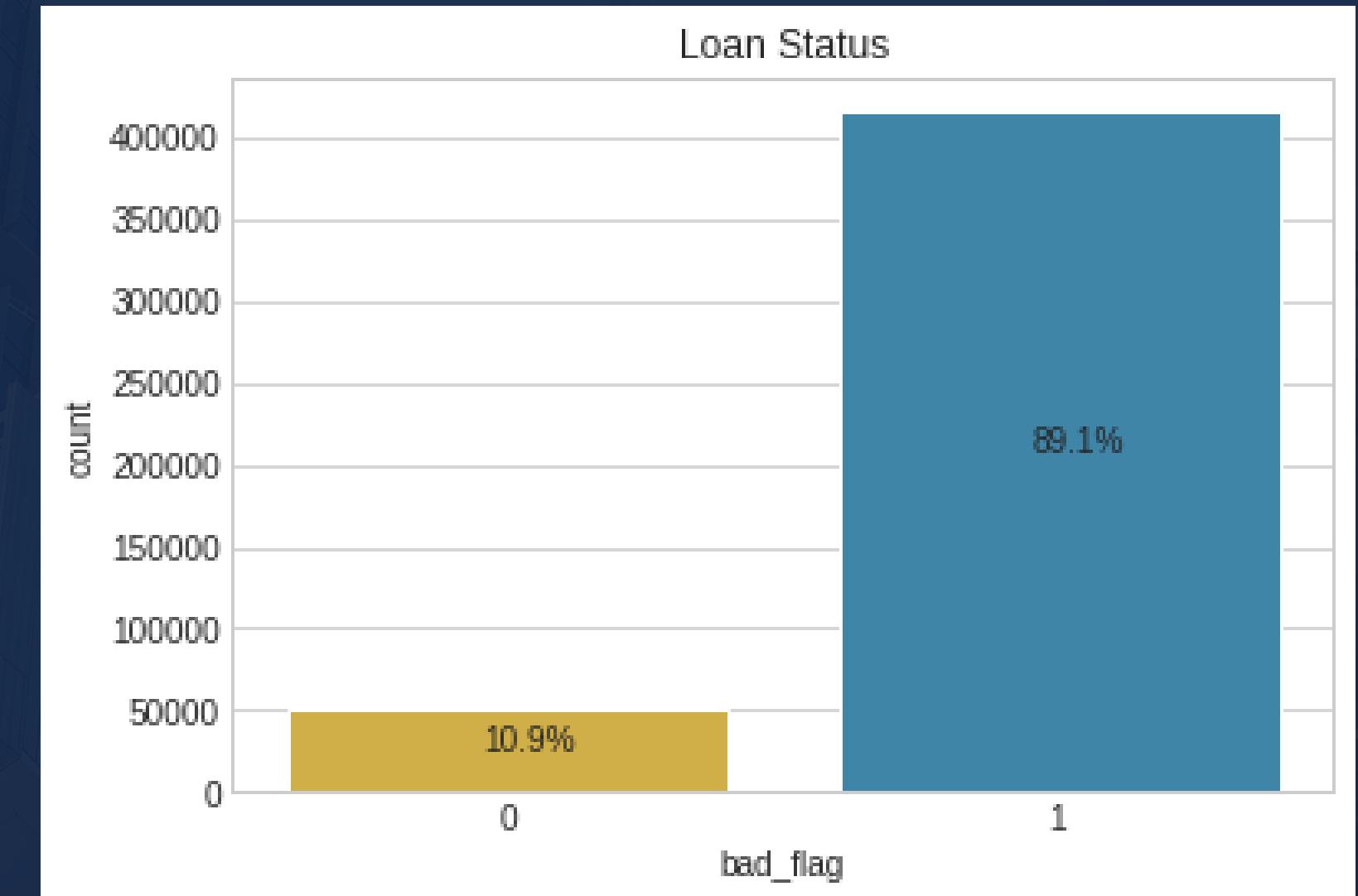
GOOD LOAN

There are **89.1%** Borrowers are in good status ('1') more than half the number of bad borrowers ('0')



BAD LOAN

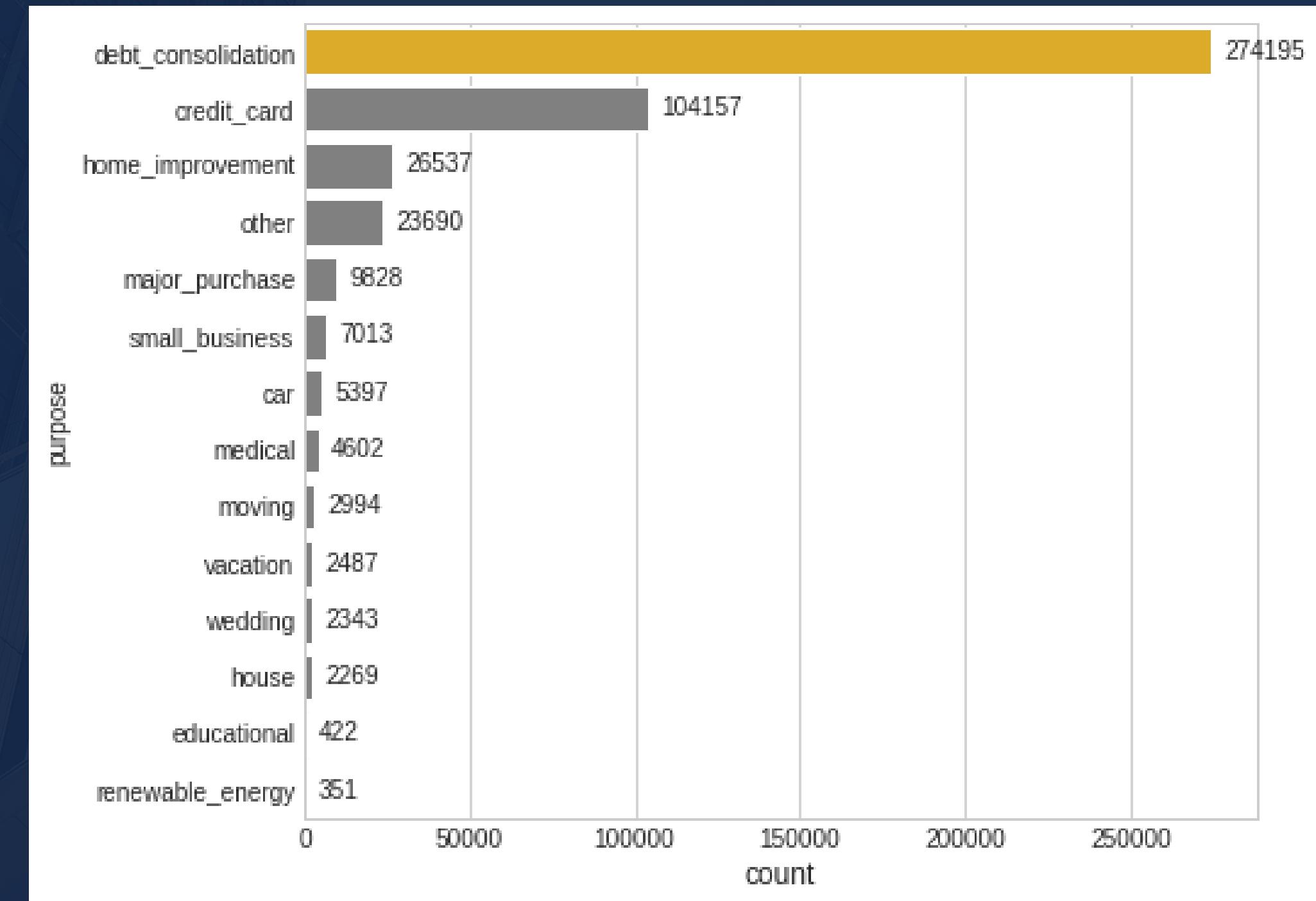
There are **10.9%** Borrowers are in Bad status ('0')



LOAN PURPOSE



Most loans are made for **debt consolidation** reasons



TOTAL LOAN

Loan Status	Mean	Std
Good Loan	14282	8274
Bad Loan	14596	8380

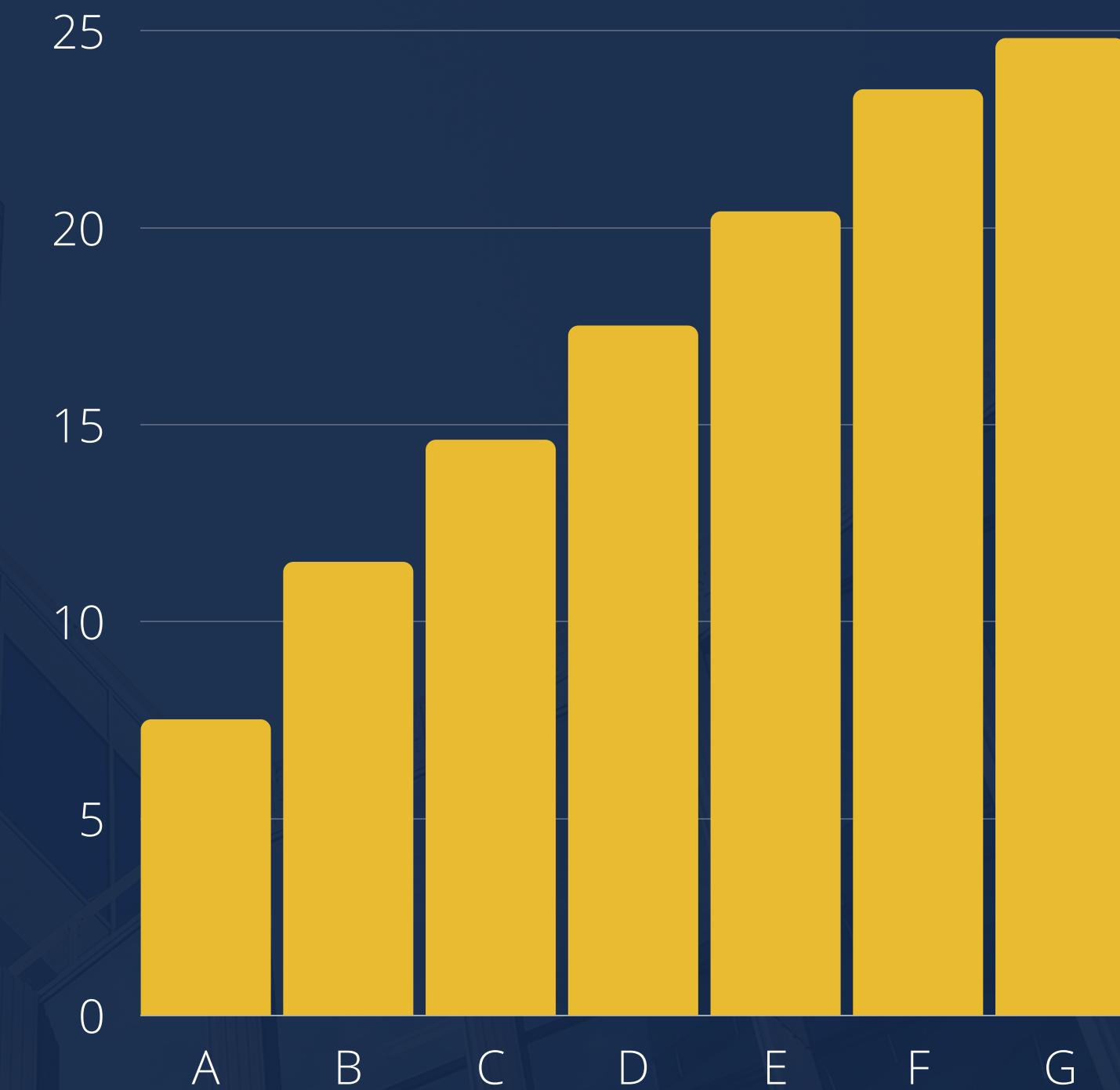


it turns out that the **Average and standard deviation** value of total credit in **bad loan status is greater than Good loan status.**

RANK LOAN



Each loan rating has its own interest rate. Starting from rank A with an average interest rate of **7.5%** to rank G with an average interest rate of **24.8%**

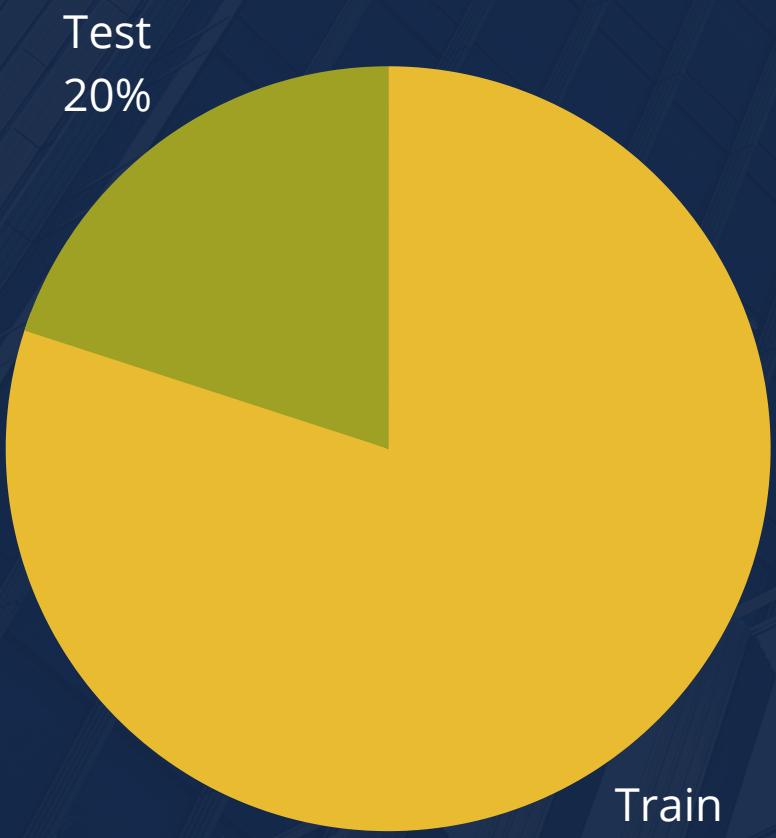


PREPROCESSING

- Handling missing value with `mean()`
- Create dummy variable in categorical column
- Normalize data using `StandardScaler()`

MODELING

Data Splitting

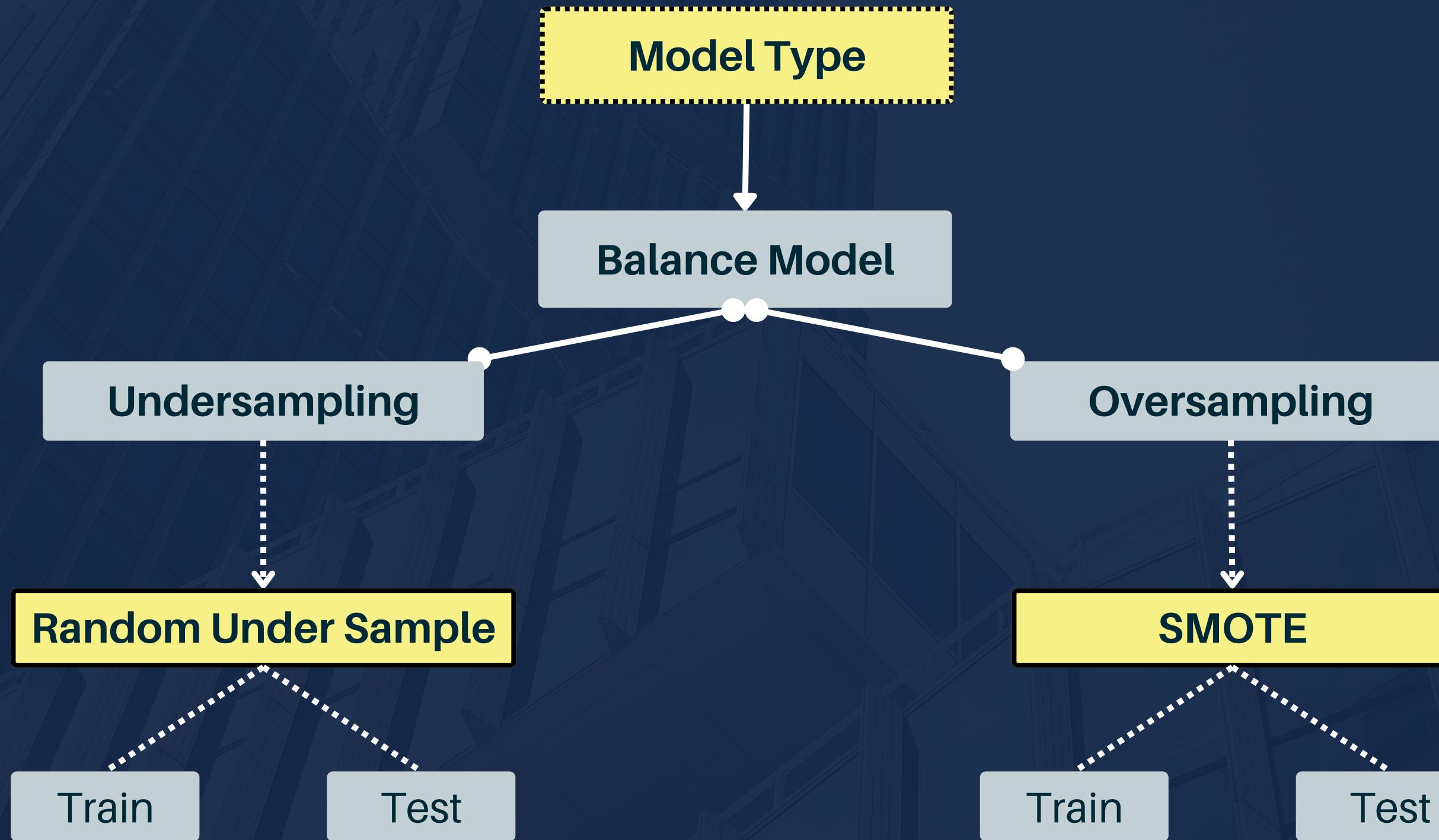


- Total of Train : **373028 Rows**
- Total of Test : **93257 Rows**

Model and Balance Model

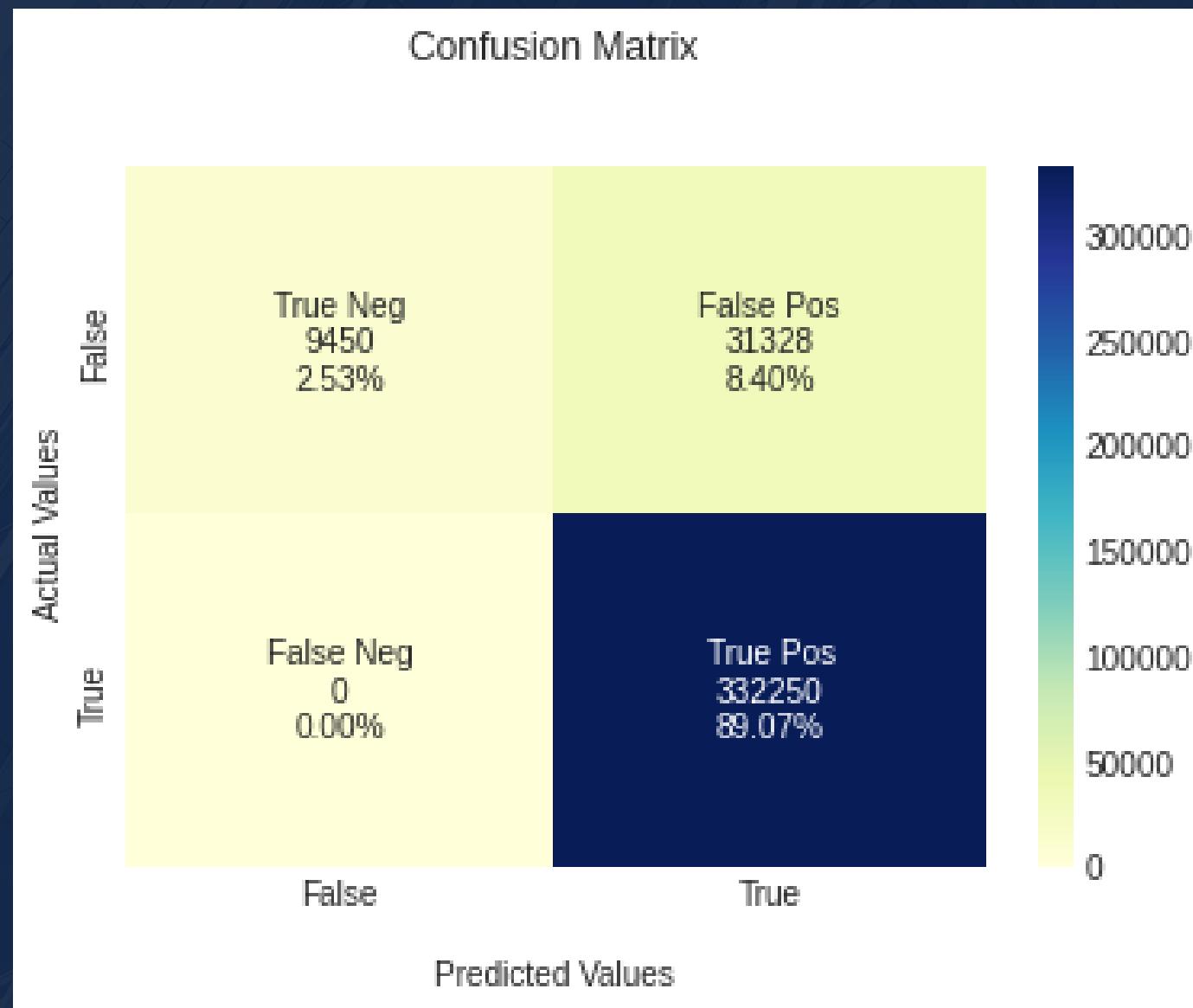
- There are 2 models that will be used are :
 - **Random Forest Classifier**
 - **Logistic Regression**
- There are 2 balance models that will be used are :
 - **Random Under Sample**
 - **SMOTE**

MODELING FLOW

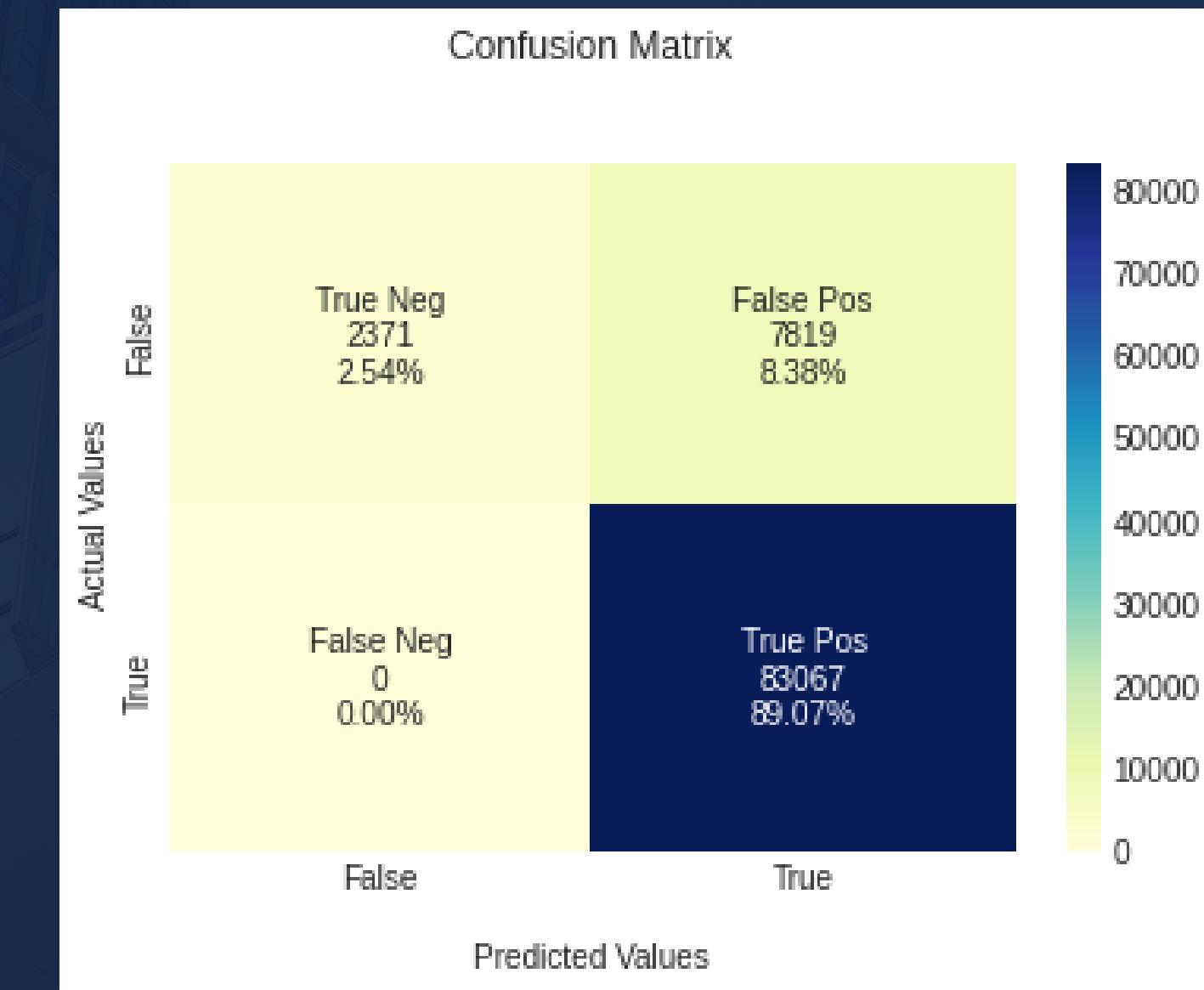


EVALUATION METRICS

Random Forest Classifier



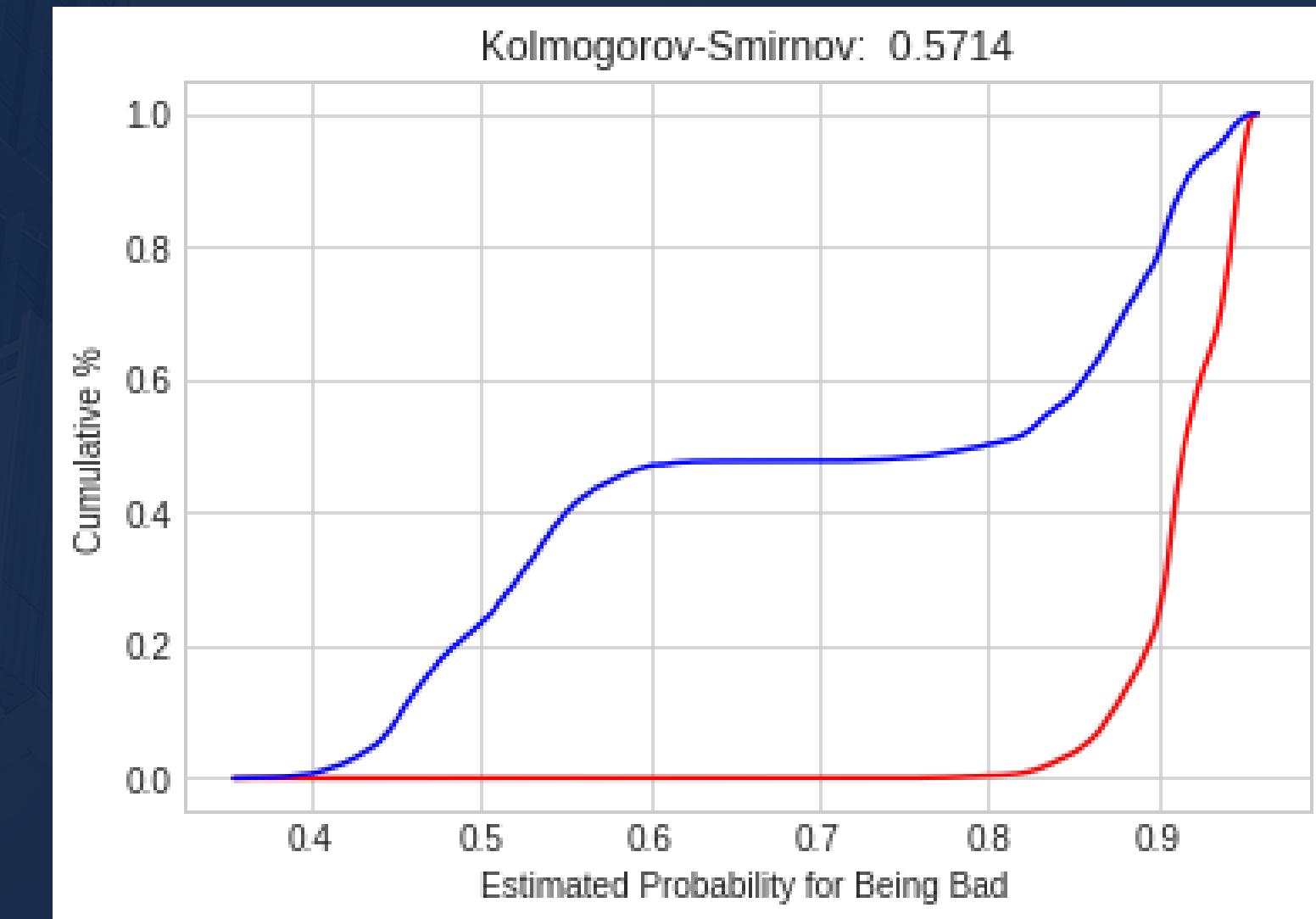
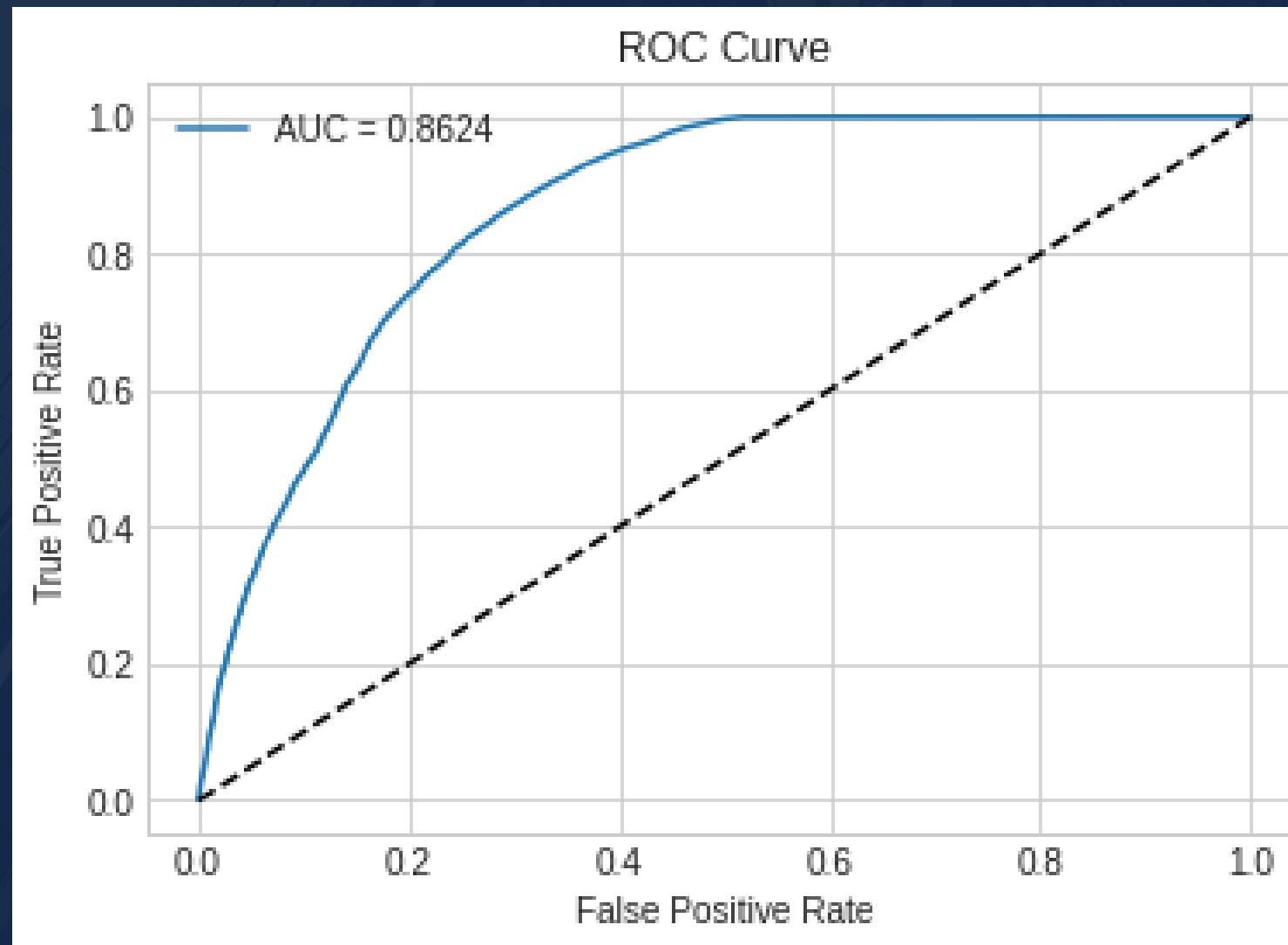
Train



Test

EVALUATION METRICS

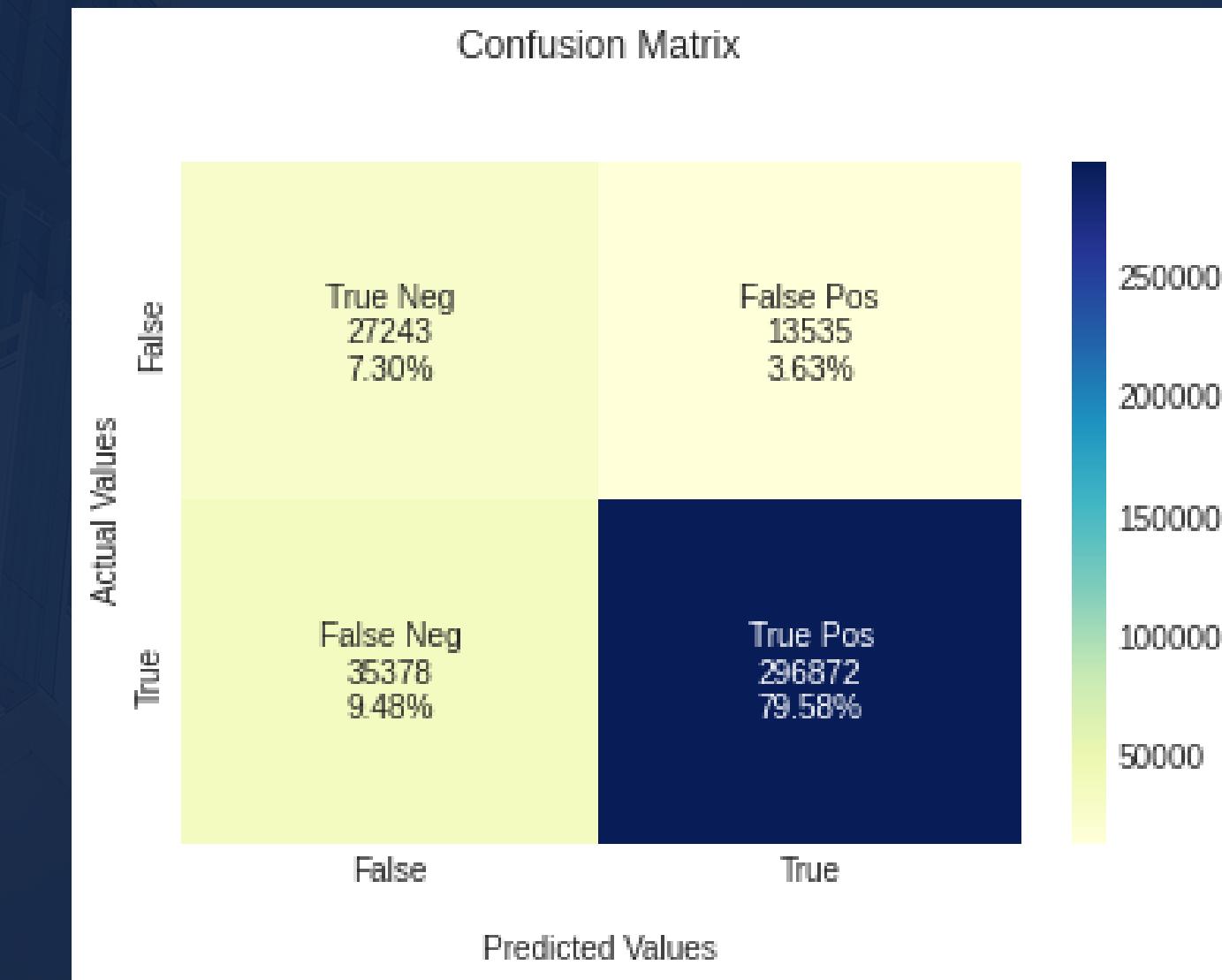
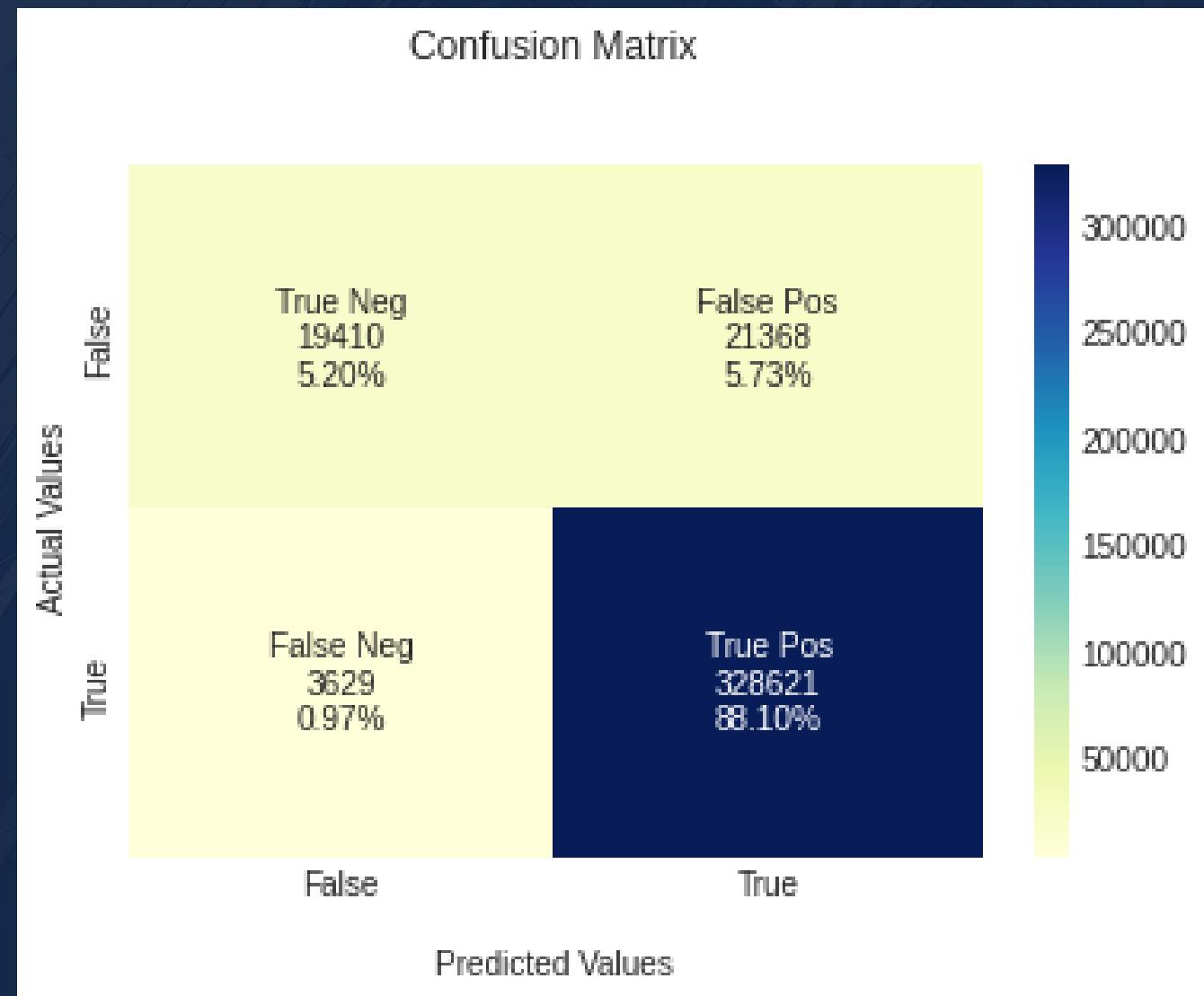
Random Forest Classifier



The model that was built resulted in the performance of **AUC = 0.857** and **KS = 0.56**. In the world of credit risk modeling, generally AUC above 0.7 and KS above 0.3 **are considered good performances**.

EVALUATION METRICS

Logistic Regression - Train

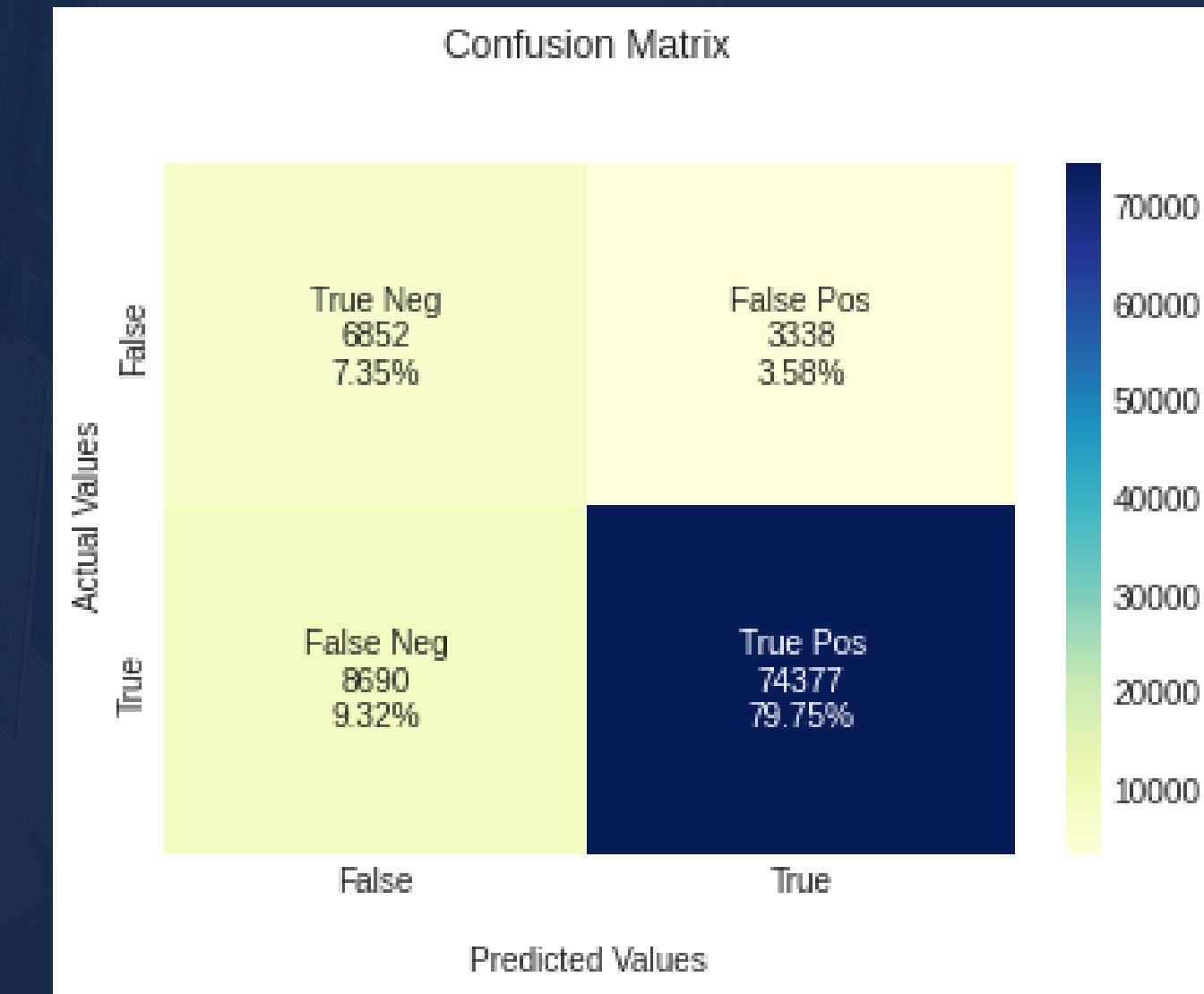
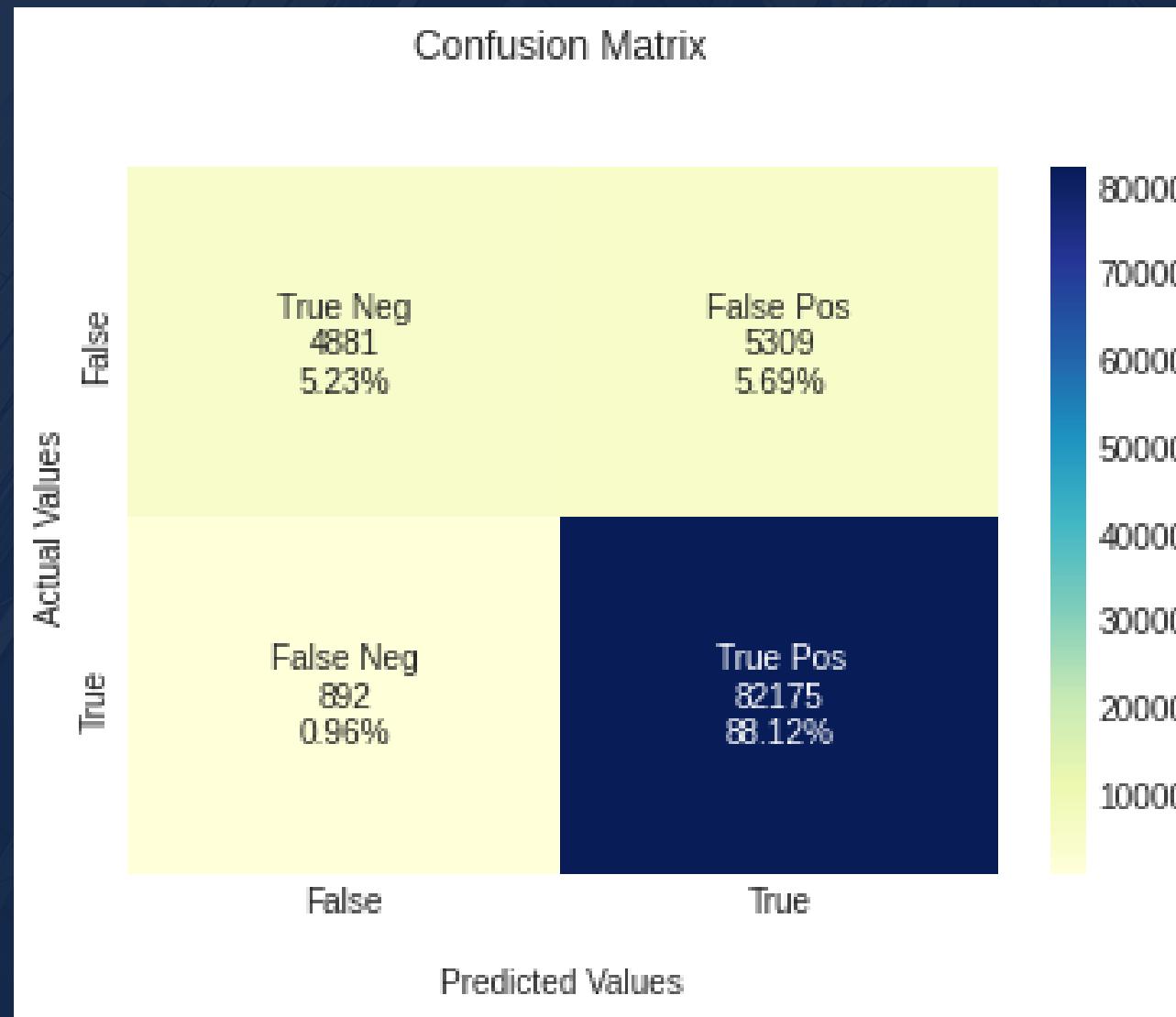


Undersampling

Oversampling

EVALUATION METRICS

Logistic Regression - Test



Undersampling

Oversampling

TABLE OF MODEL PERFORMANCE

	Model Name	Balance Model	Train/Test	Accuracy	Precision	F1-Score	Recall
Logistic Regression	Random Under Sample	Train	87%	96%	89%	92%	
		Test	87%	96%	90%	93%	
	SMOTE	Train	93%	94%	99%	96%	
		Test	93%	94%	99%	96%	
Random Forest Classifier		Train	92%	91%	100%	95%	
		Test	92%	91%	100%	96%	

RESULT ANALYSIS

Based on the model performance table, the **Logistic Regression** Model has a fairly high accuracy compared to others so that we can predict the pattern of borrowing well, with an accuracy level of **93%**

PRECISION

94%

Possibility to predict **bad loan** status as **bad loan**

F1-SCORE

99%

It's possible not to mis-predict **bad loan** status as **good loan**.

RECALL

96%

The combination of precision and recall, the likelihood of how **consistent** the model is in **predicting loan status**.

GET IN TOUCH



+6281344530897



richarddamarjanan@gmail.com



<https://www.linkedin.com/in/richardo-damarjanan/>



<https://github.com/richardzefan>



THANK YOU

PROGRAM VIRTUAL INTERNSHIP

