Predicting Loan Defaults: A Data-Driven Approach to Credit Risk Analysis

BEE2041 - Data Science in Economics

Student Number - 720017170

Table of contents

Table 2: Summary Statistics

Variable	N	Mean	Median	SD	Min	Max
Age	5930.0	40.5	39.0	14.9	18.0	69.0
Income	5930.0	78831.0	76708.0	40289.1	15014.0	149944.0
LoanAmount	5930.0	134944.3	137330.5	70970.4	5000.0	249929.0
CreditScore	5930.0	568.9	568.0	158.7	300.0	849.0
MonthsEmployed	5930.0	55.6	54.0	34.8	0.0	119.0
InterestRate	5930.0	14.7	15.3	6.6	2.0	25.0
DTIRatio	5930.0	0.5	0.5	0.2	0.1	0.9

1. Introduction

2. Data

4. Conclusion

	Default	Stage	Count
0	0	Before Downsampling	225694
1	1	Before Downsampling	29653
2	0	After Downsampling	29653
3	1	After Downsampling	29653

2.1 Descriptive Statistics

2.2 Distribution Analysis

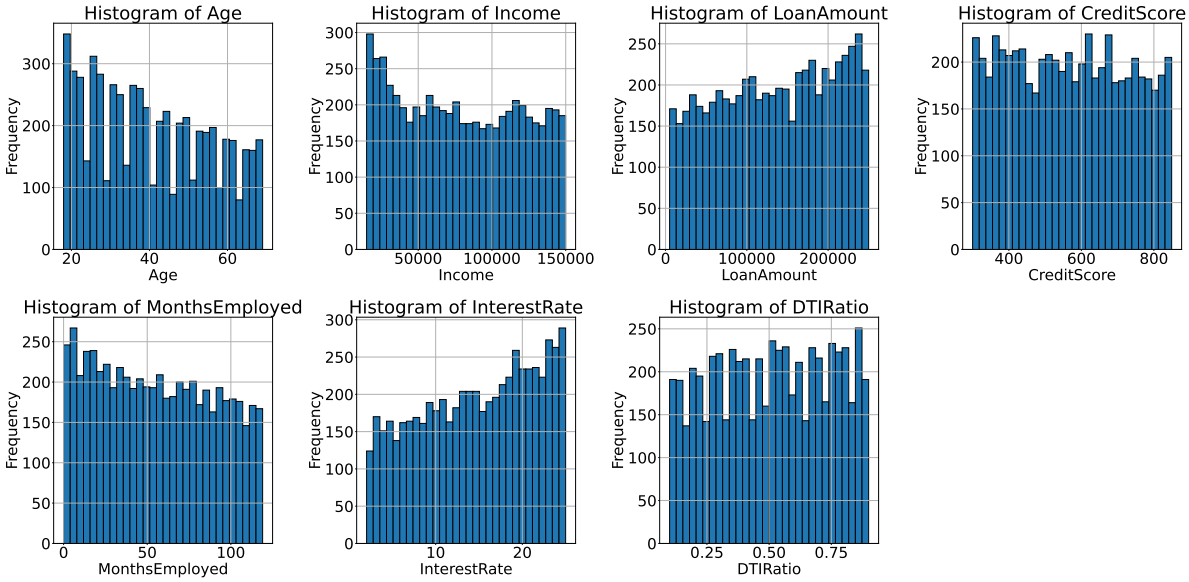

Figure ?: Histograms of all Numeric Variables

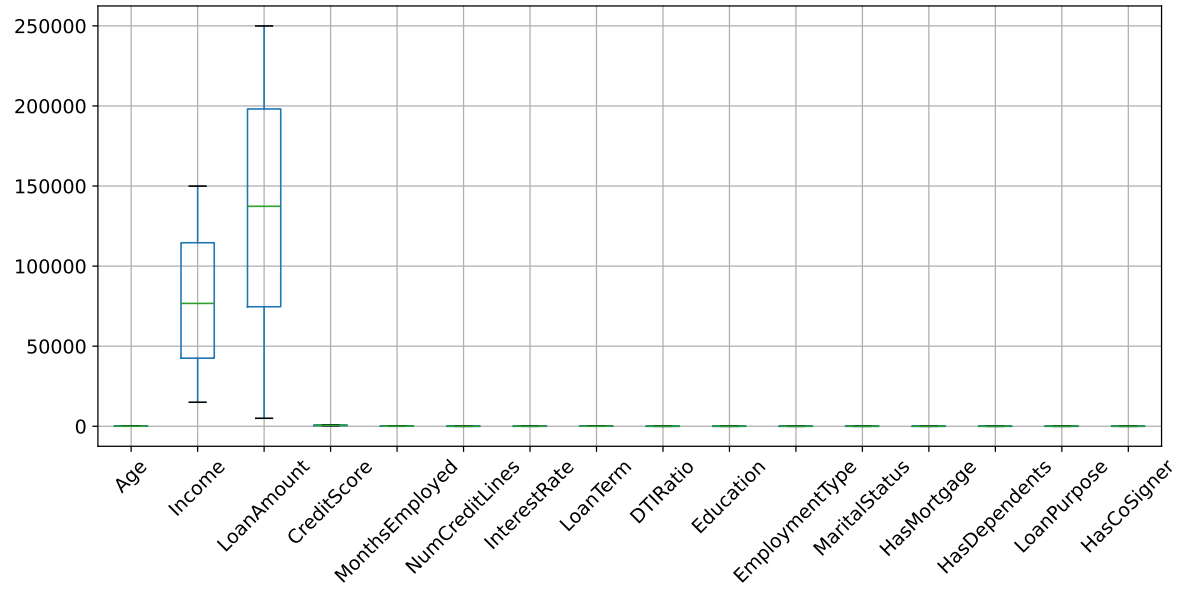

Figure ?: Box Plots of All Variables Before Normalisation

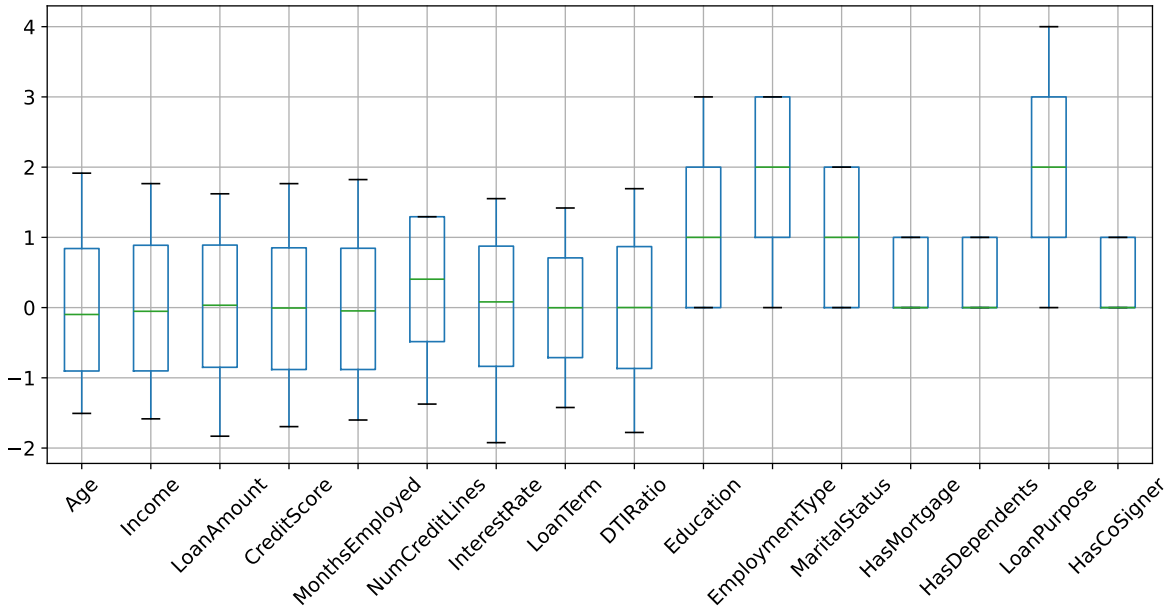

Figure ?: Box Plots of All Variables After Normalisation

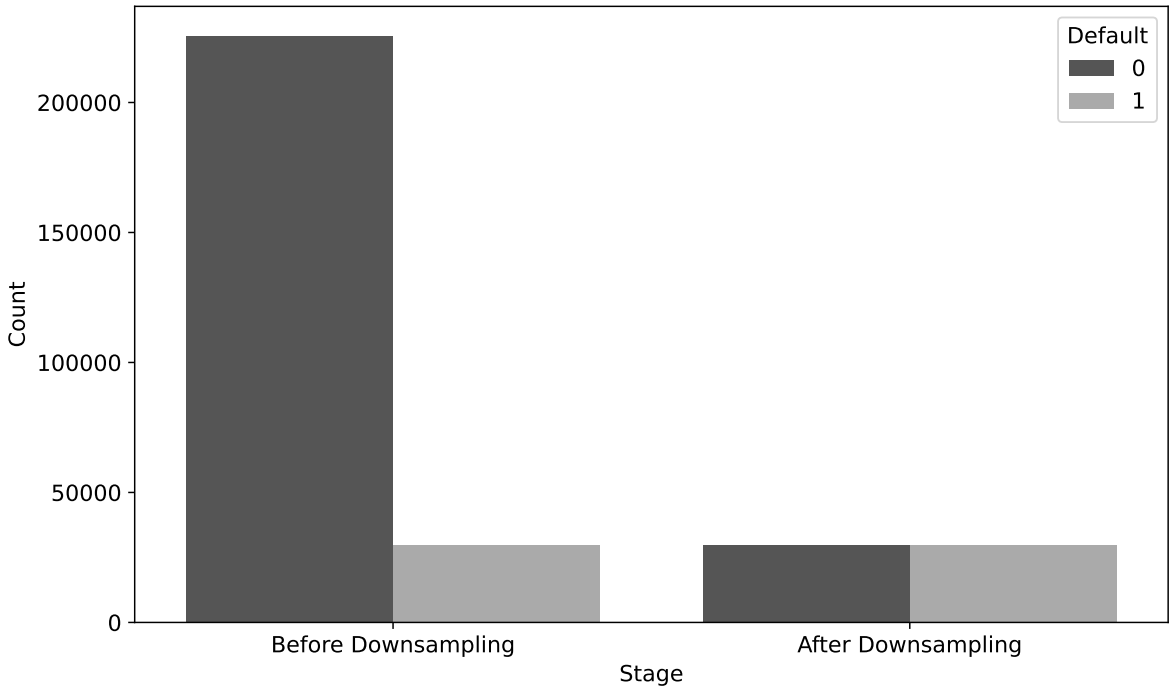

Figure ?: Distribution of Default Before and After Downsampling

2.3 Correlation Analysis

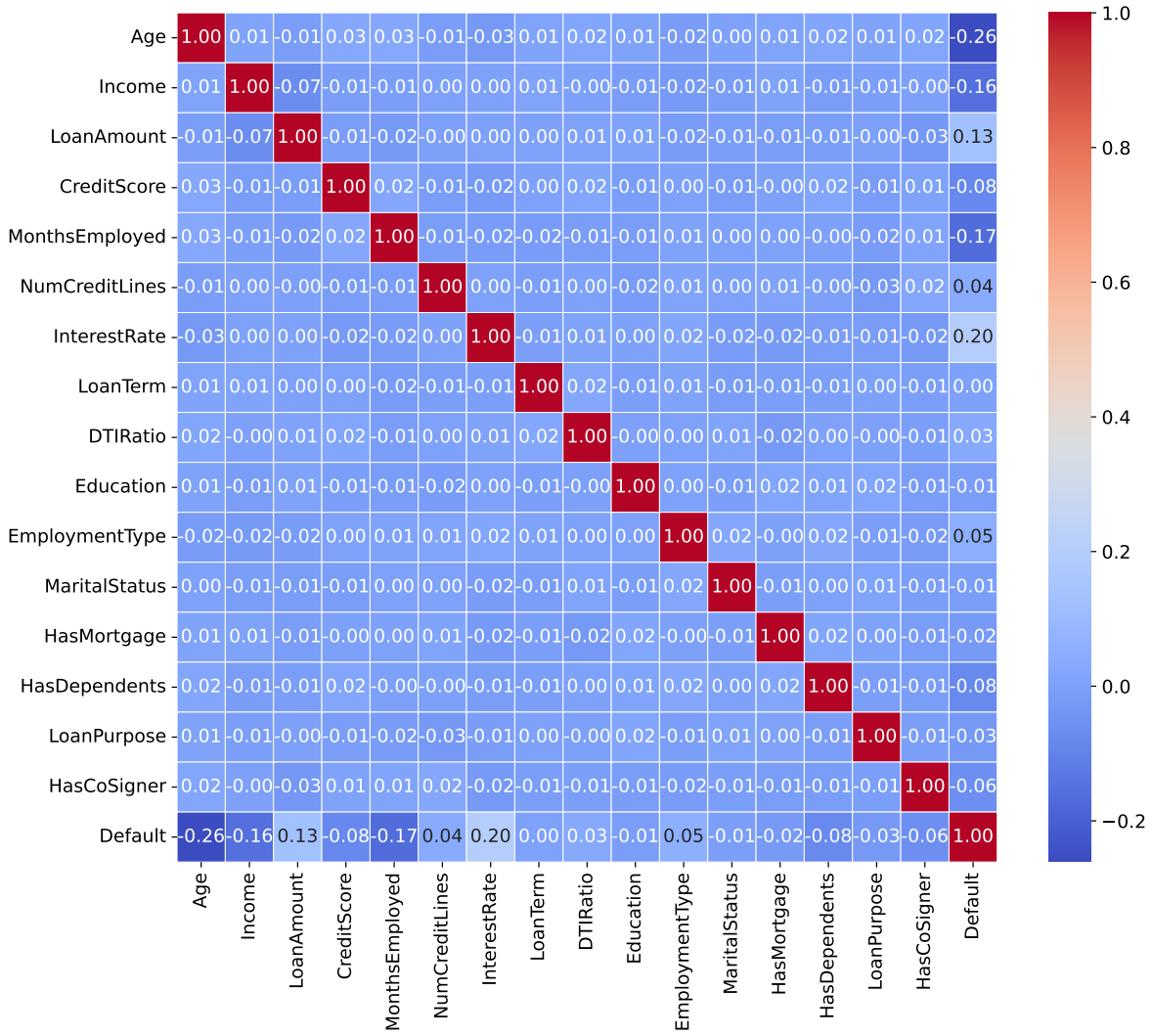

Figure ?: Correlation Plot of All Variables

3. Results and Discussion

**3.1 SVM

```
Fitting 5 folds for each of 16 candidates, totalling 80 fits
```

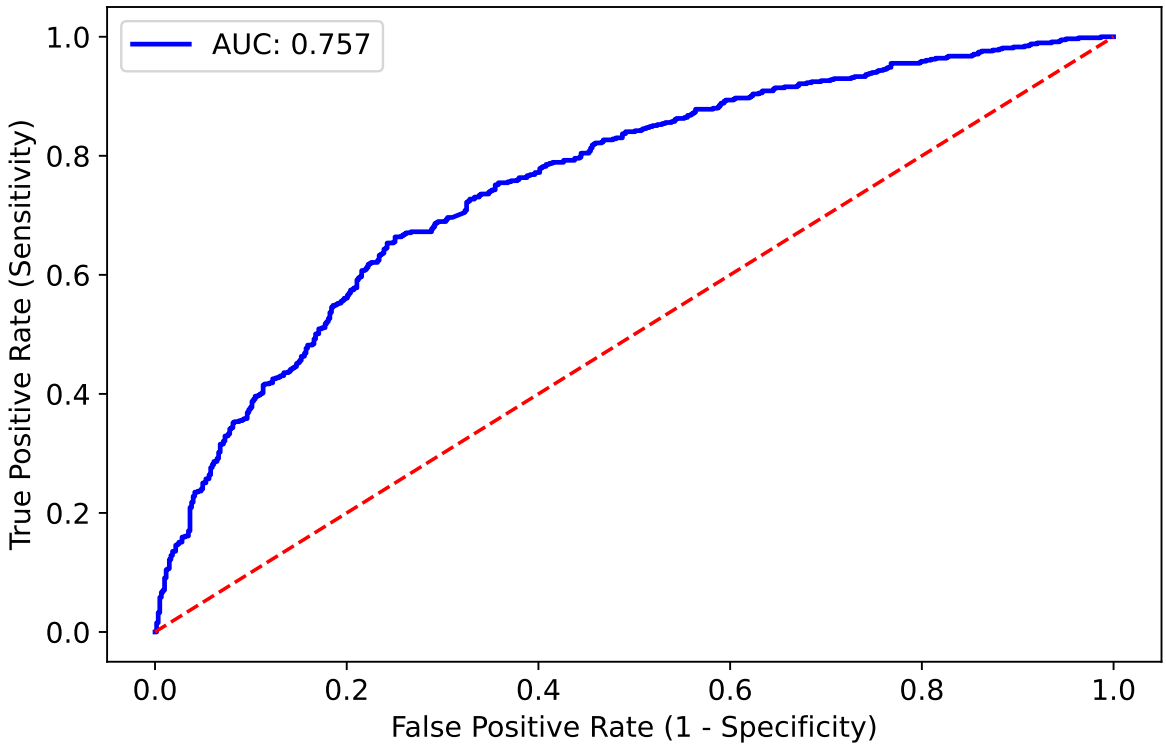

Figure ?: ROC Curve for SVM Model

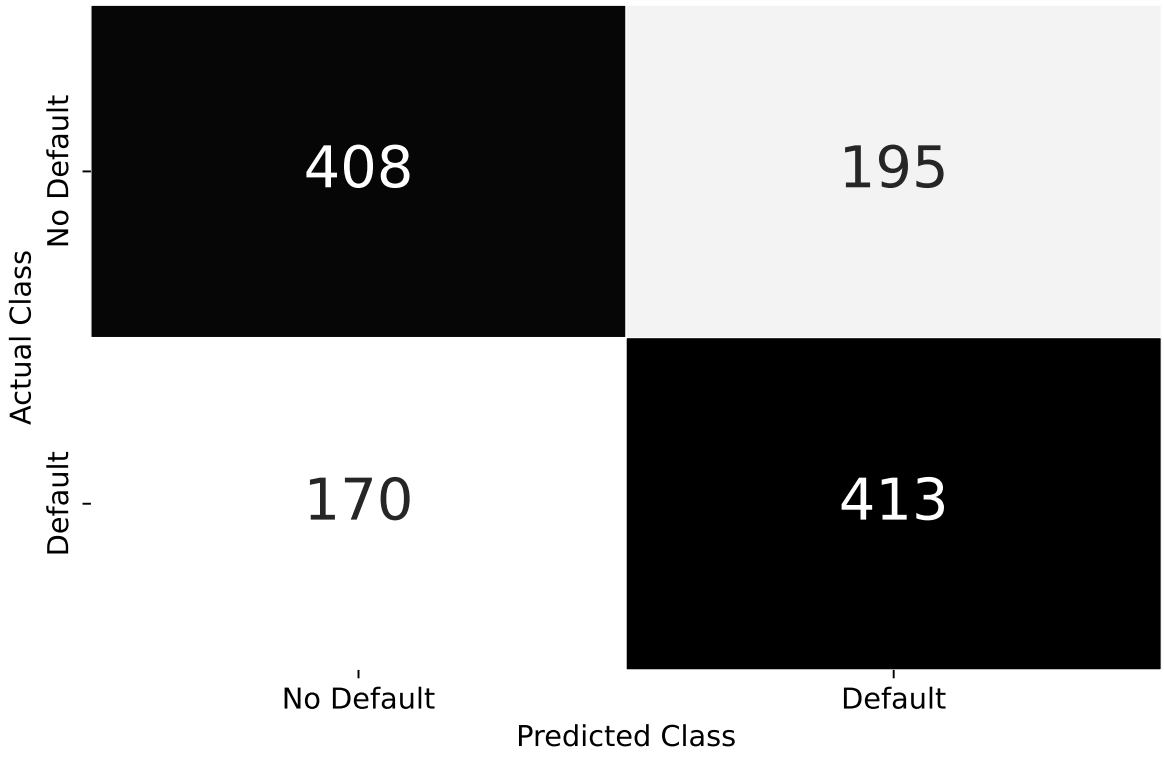

Figure ?: Confusion Matrix for SVM Model

****3.2 Naive Bayes**

```
Fitting 5 folds for each of 100 candidates, totalling 500 fits
```

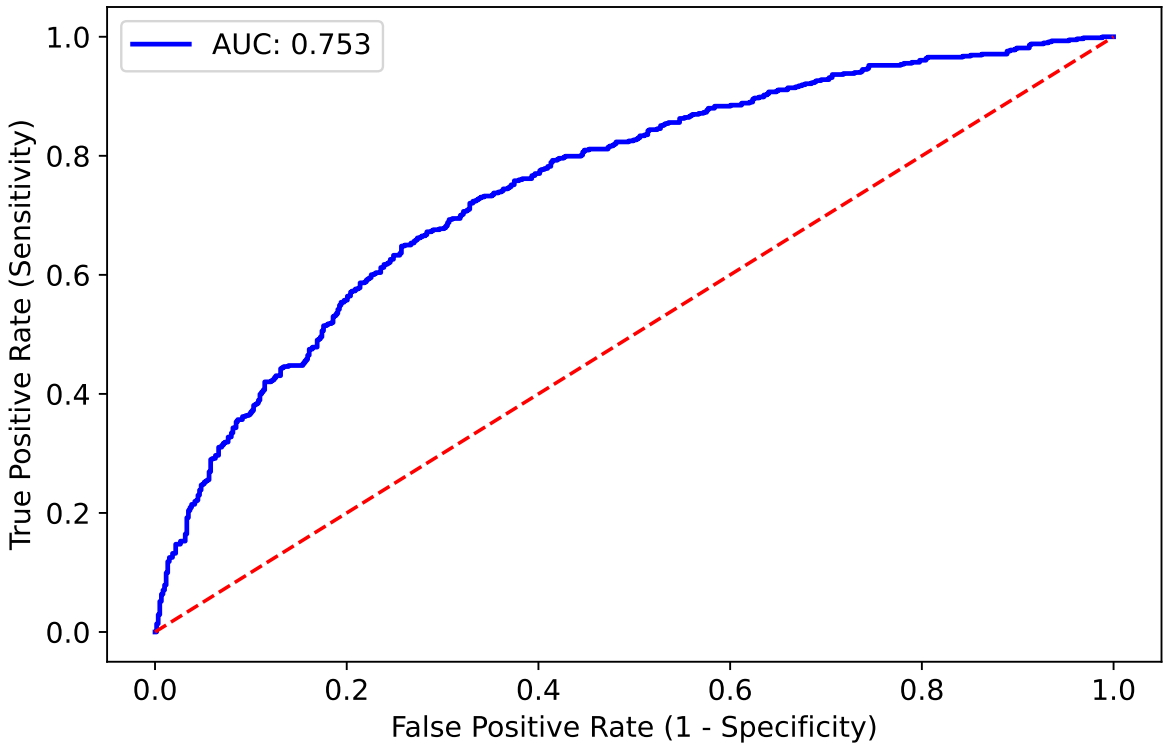

Figure ?: ROC Curve for Naive Bayes Model

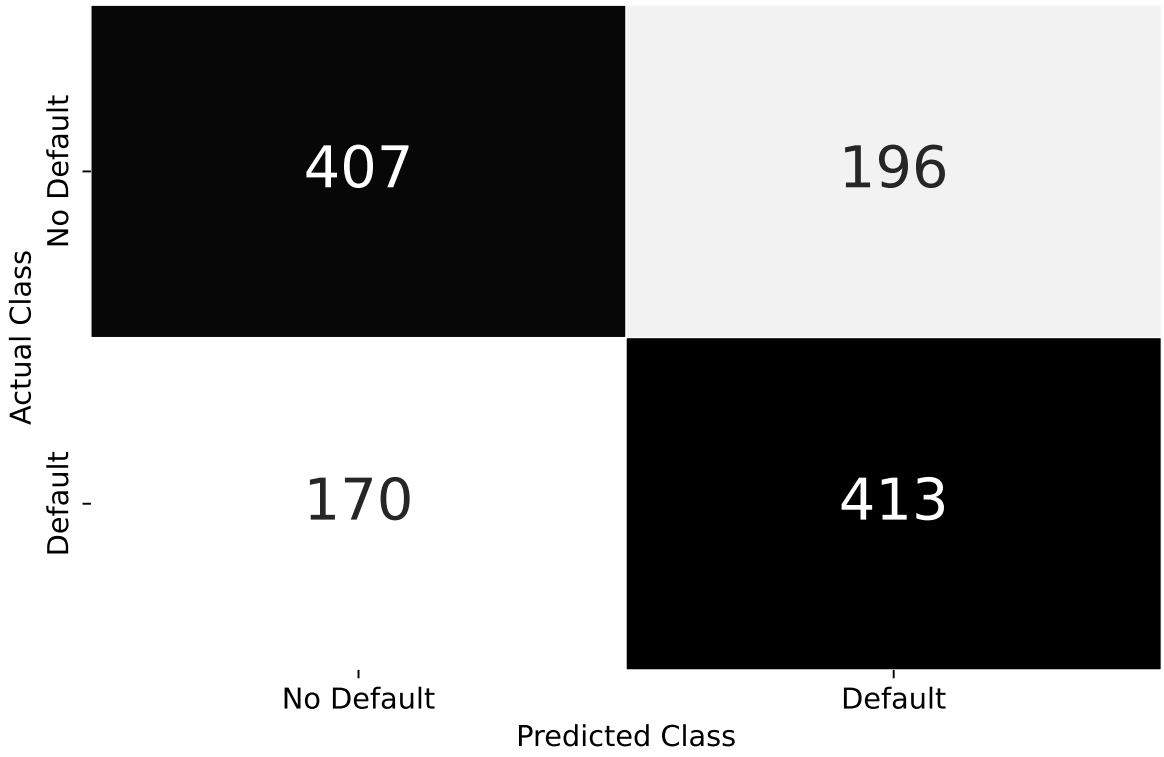

Figure ?: Confusion Matrix for NB Model

**3.3 XGBoost

```
Fitting 5 folds for each of 81 candidates, totalling 405 fits
```

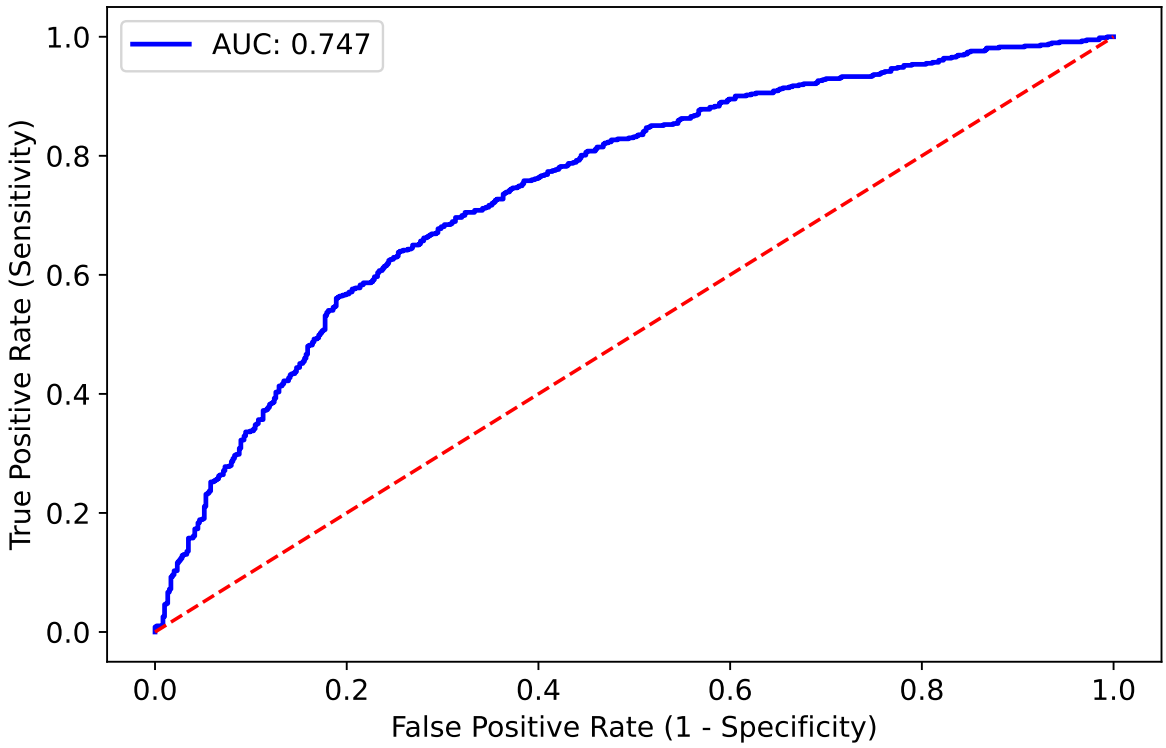

Figure ?: ROC Curve for XGBoost Model

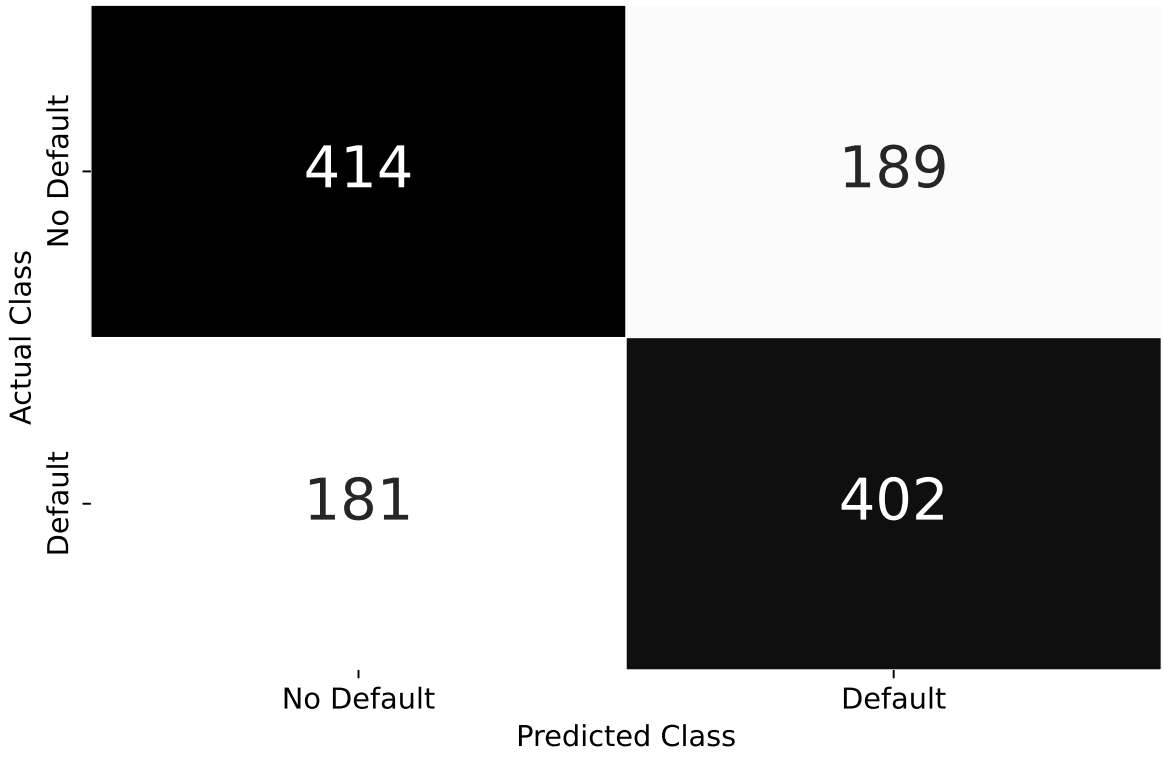

Figure ?: Confusion Matrix for SVM Model

3.4 Model Evaluation and Comparisons

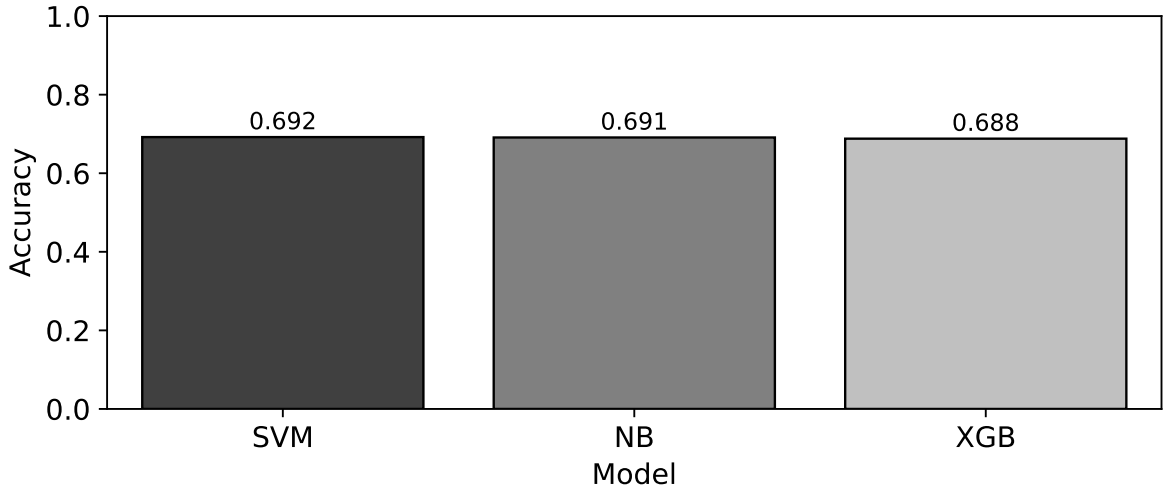

Figure ?: Accuracy for Each Model

Table 3: Performance Metrics for Each Model

Model	Accuracy	Precision	Recall	F1_Score	AUC
SVM	0.692	0.679	0.708	0.694	0.757
NB	0.691	0.678	0.708	0.693	0.753
XGB	0.688	0.68	0.69	0.685	0.747