Impact of Machine Learning in Credit Scoring

Johannes Tebalo Kokozela (714188), Supervisor(s): Prof. Turgay Celik & Dr. Terence Van Zyl
School Of Computer Science & Applied Mathematics

1. Abstract

Our research seeks to answer the following question: are machine learning (ML) algorithms better predictors than statistical algorithms in credit scoring when using a huge a dataset? We hypothesised that ML algorithms perform better than traditional statistical algorithms, When the performance measures are Receiver Operating Characteristic's (ROC), Area Under the Curve (AUC), accuracy, precision, recall and F_1-Score. We compared four machine learning algorithms namely Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest Neighbour (k-NN) and Naïve Bayes (NB) with traditional statistical methods for credit scoring namely Logistic Regression (LR) and Linear Discriminant Analysis (LDA).

3. Results

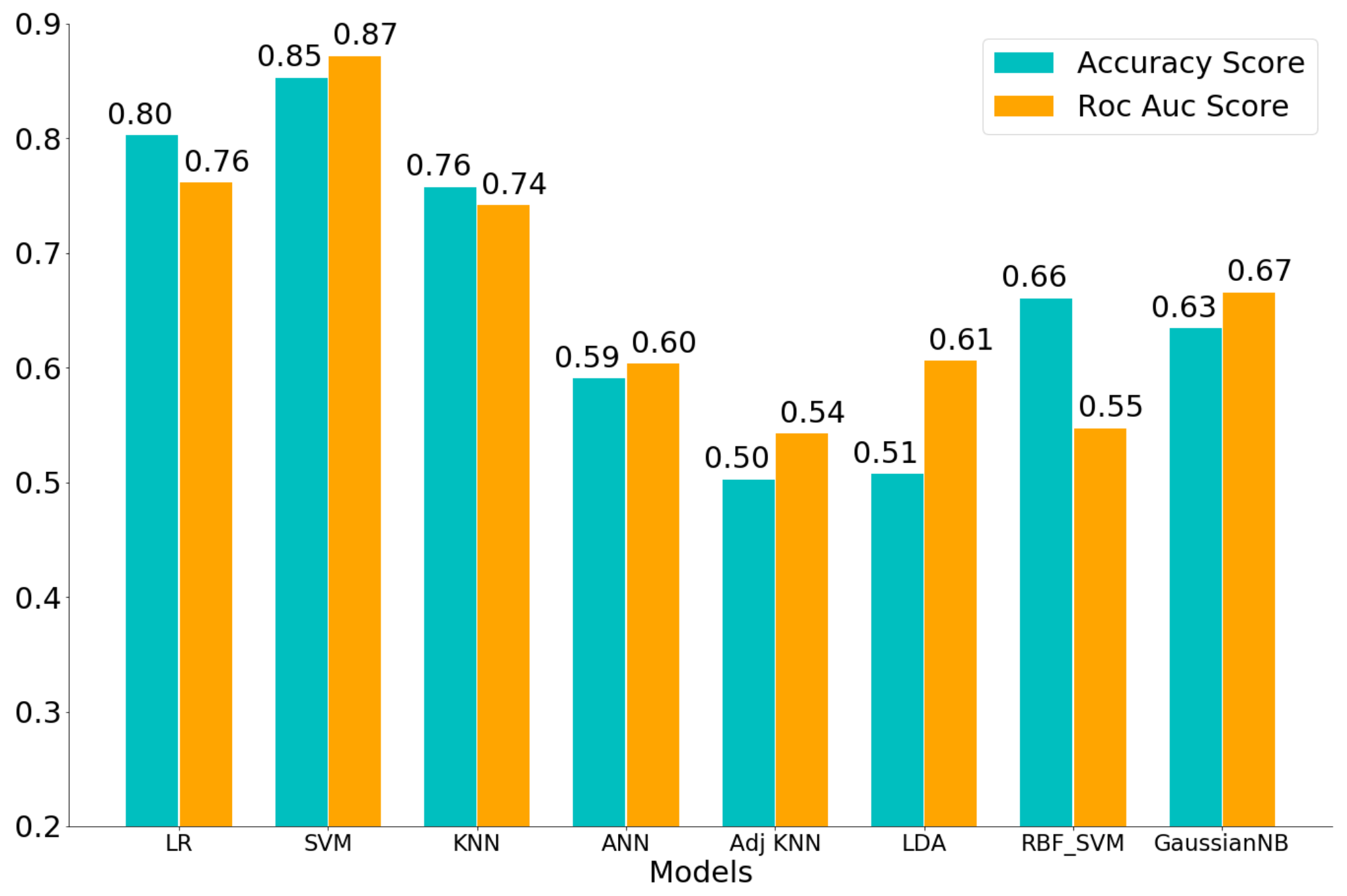


Figure 2: Accuracy and AUC plots in histogram

Figure 2 shows that linear kernel SVM has the highest Accuracy and AUC, followed by the LR.

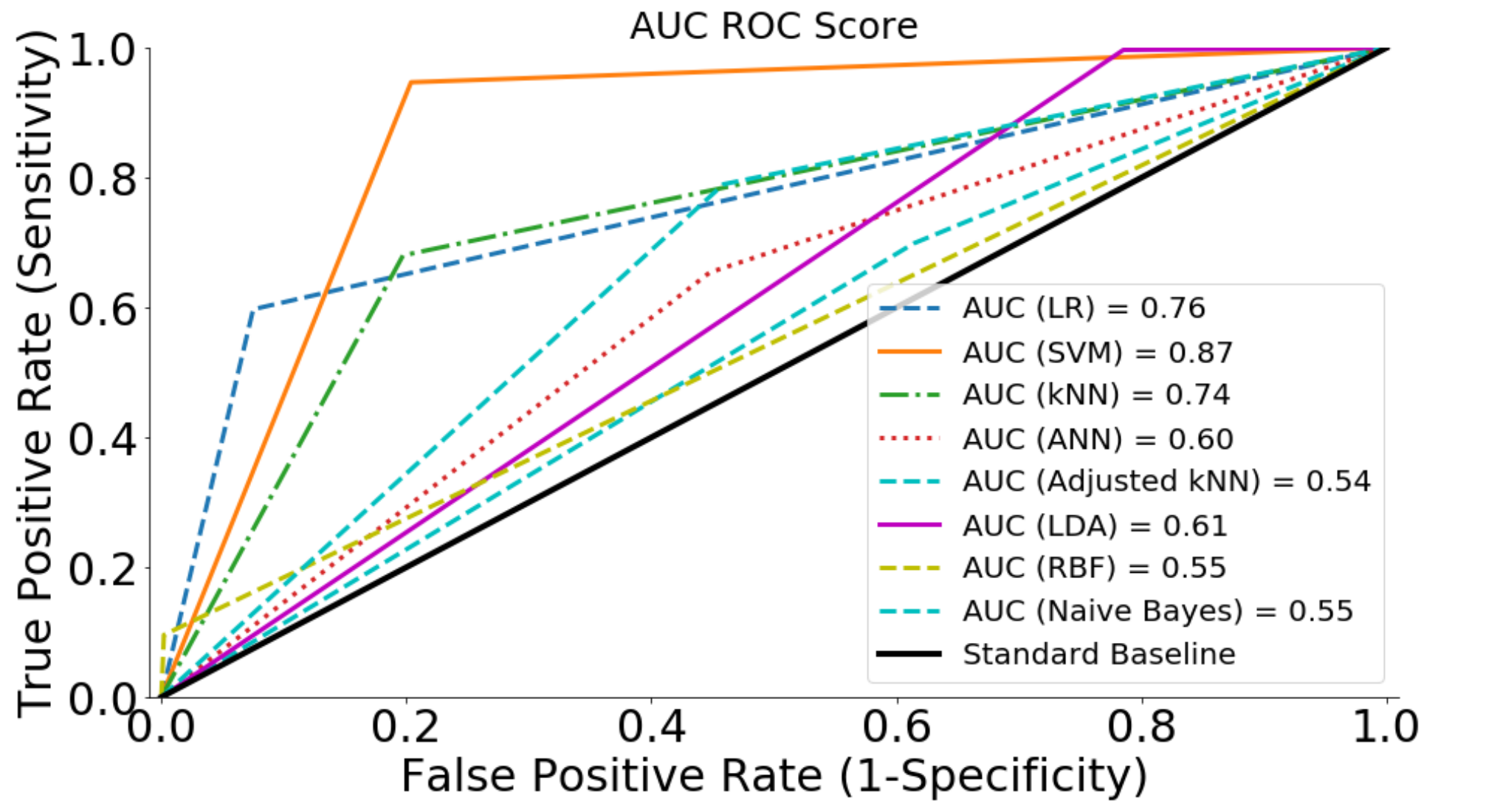


Figure 3: Accuracy and AUC plots in histogram

Table 1: Performance in %

Algorithms	Accuracy	AUC	F_1-Score
LDA	51	61	60
LR	**80**	**76**	69
SVM	**85**	**87**	**82**
k-NN	**76**	**74**	68
ANN	59	60	54
Adj k-NN	50	54	51
NB	63	67	61
RBF-SVM	66	55	**17**

$$F_1\text{−Score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \qquad (1)$$

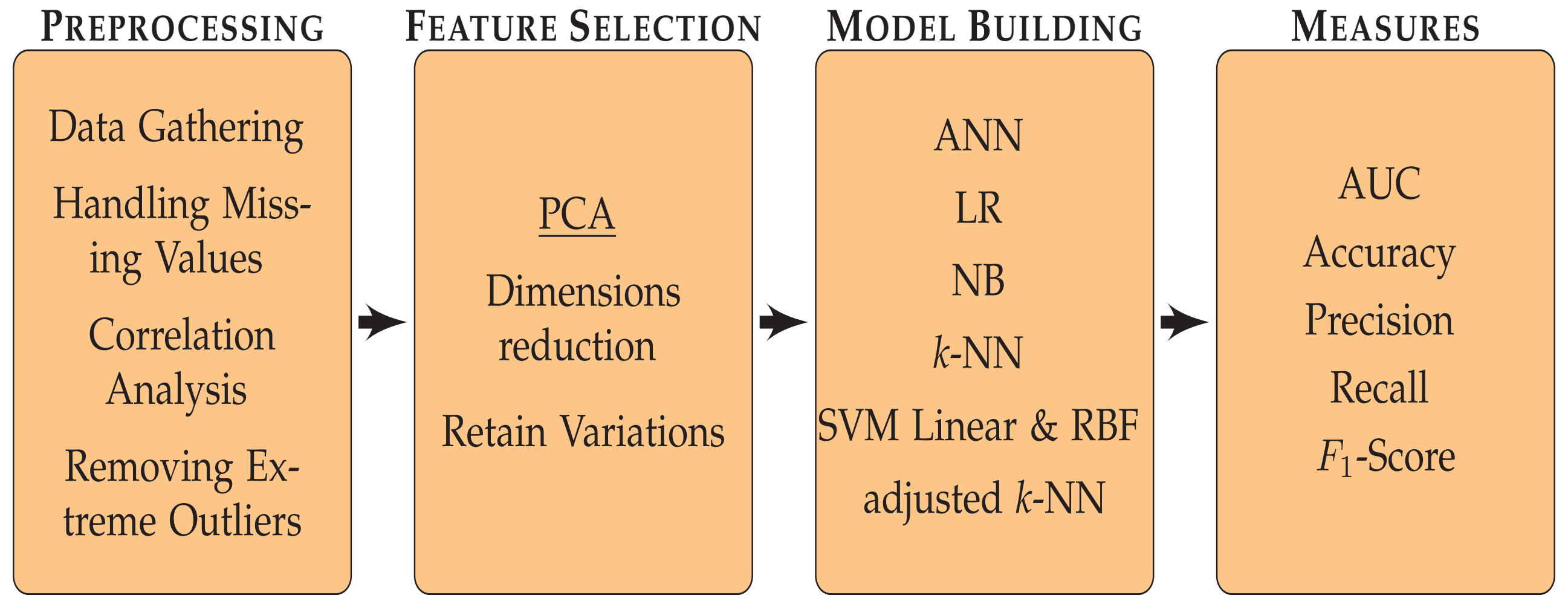
About the figures

- Table 1 shows that linear SVM has a better F_1-Score again followed by LR and standard k-NN.
- Figure 3 shows the ROC-AUC curve, From the plot we observe that none of the algorithms are randomly guessing the results.

5. Future Research

There are several lines of investigations. Firstly, this research did not focus on feature engineering algorithms for credit data. Secondly, we could apply a different kernel for LDA, and look at the polynomial kernel for SVM.
Thirdly, future research could also use ensemble of models that is adaptive boosting and bagging. Finally we could have looked at ways to improve the weak algorithms, this includes finding a better distance matrix, k value for the adjusted k-NN, increasing the number of iterations and changing the architecture of the ANN.

2. Methodology & Materials



Preprocessing: Data Gathering; Handling Missing Values; Correlation Analysis; Removing Extreme Outliers

Feature Selection: PCA; Dimensions reduction; Retain Variations

Model Building: ANN; LR; NB; k-NN; SVM Linear & RBF; adjusted k-NN

Measures: AUC; Accuracy; Precision; Recall; F_1-Score

$n = 200,000$ and $\boldsymbol{X} = 35$ from the Amalgamated Banks Of South Africa (ABSA) was obtained. Before the data could be used for modelling, preprocessing steps were taken, which included correlation analysis through plots and handling of missing values and removal of extreme outliers.

We used Principal Component Analysis for dimensionality reduction. Figure 1 shows the inseparability of the policy data.

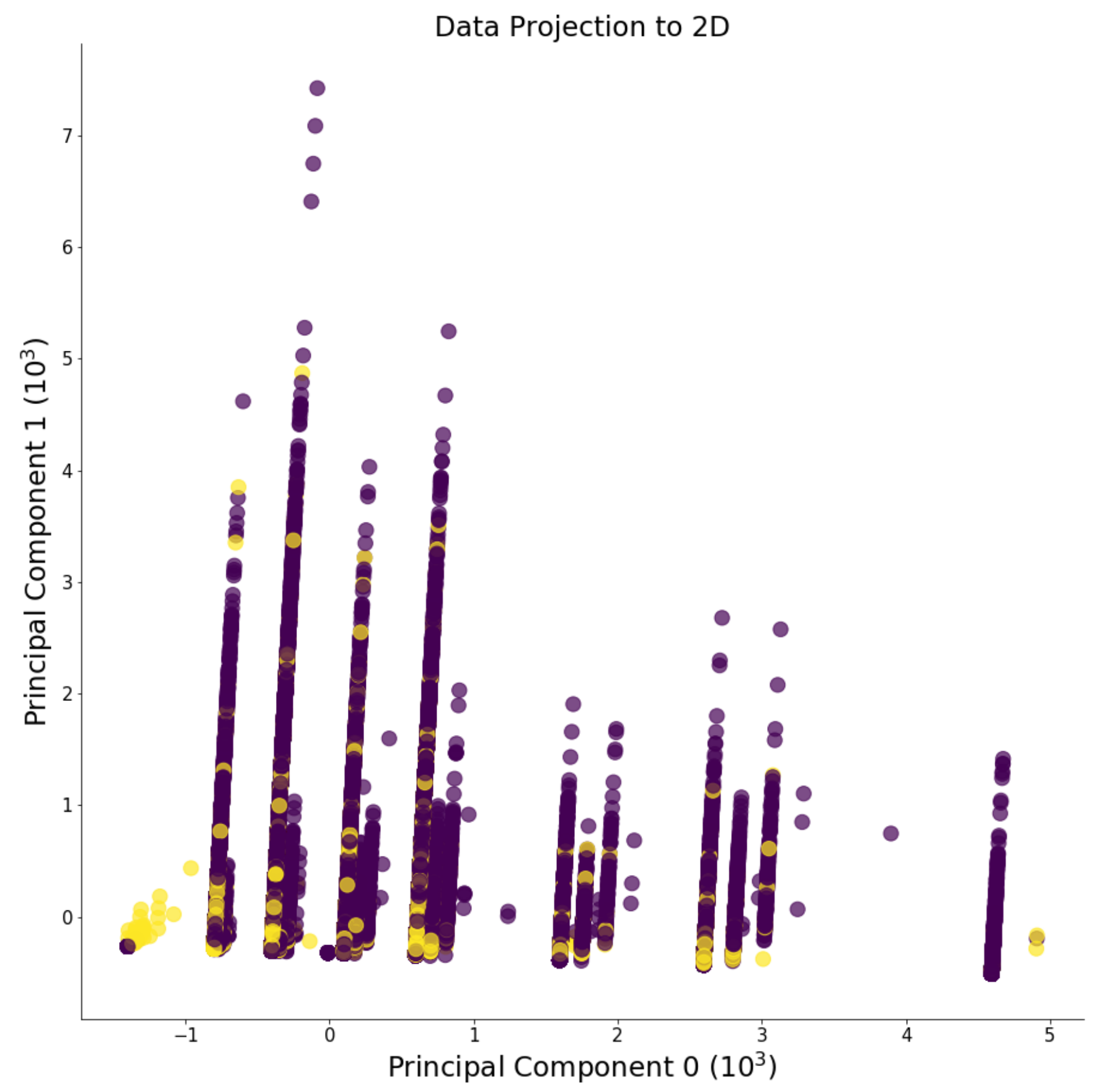


Figure 1: Inseparability Of Credit Scoring Data

4. Conclusion

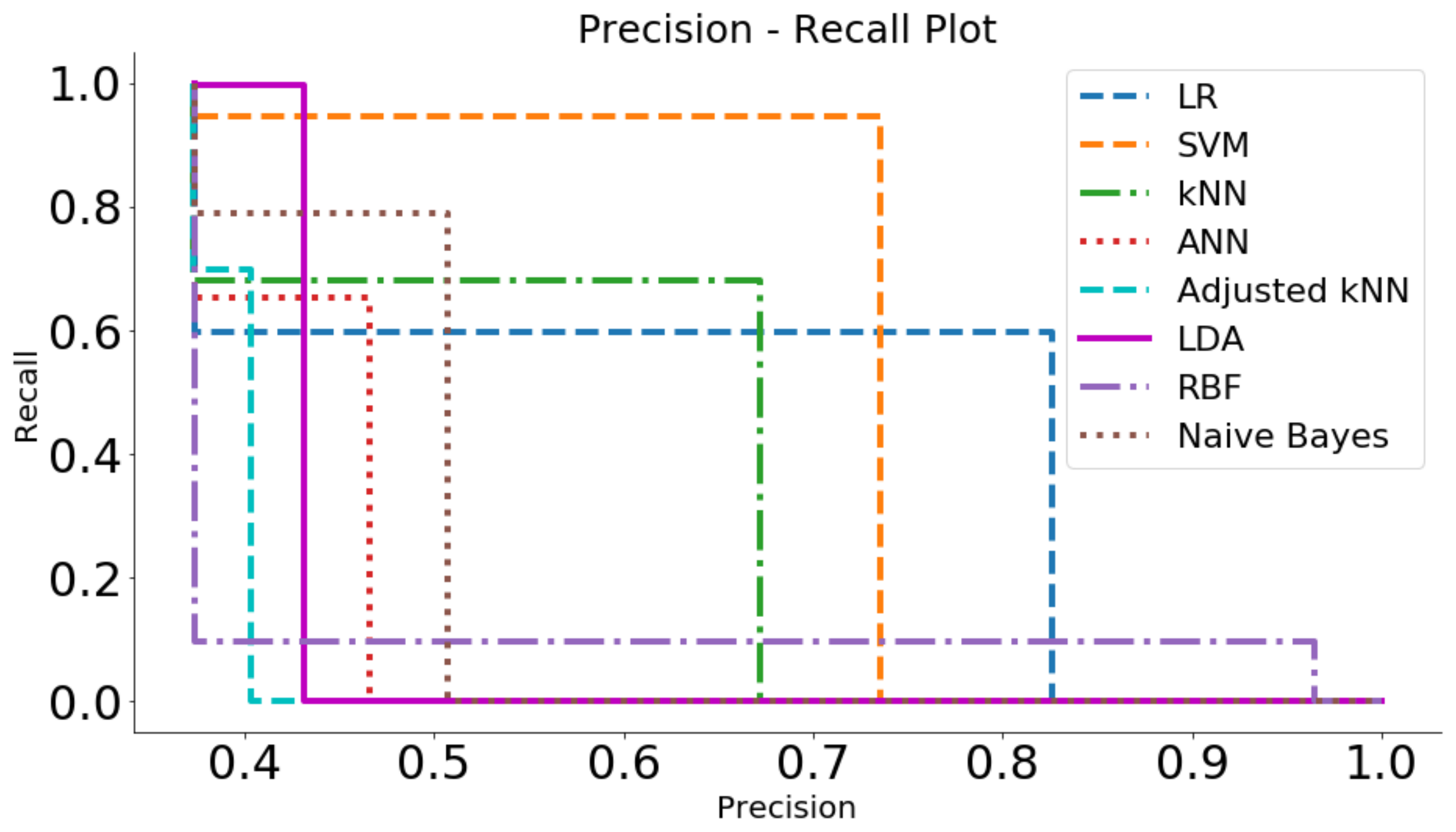


Figure 4: Precision & Recall plot

- Figure 4 shows that statistical models performed better than the ML models on average, thus they attained a higher average AUC.
- The variation within the ML algorithms are very high in comparison to the one for statistical algorithms.
- On average machine learning algorithms were more accurate than statistical models with $0.67(0.11)$ against $0.65(0.15)$.
- Statistical Models attained a better F_1-Score of $0.64(0.05)$ compared to $0.56(0.2)$ for ML

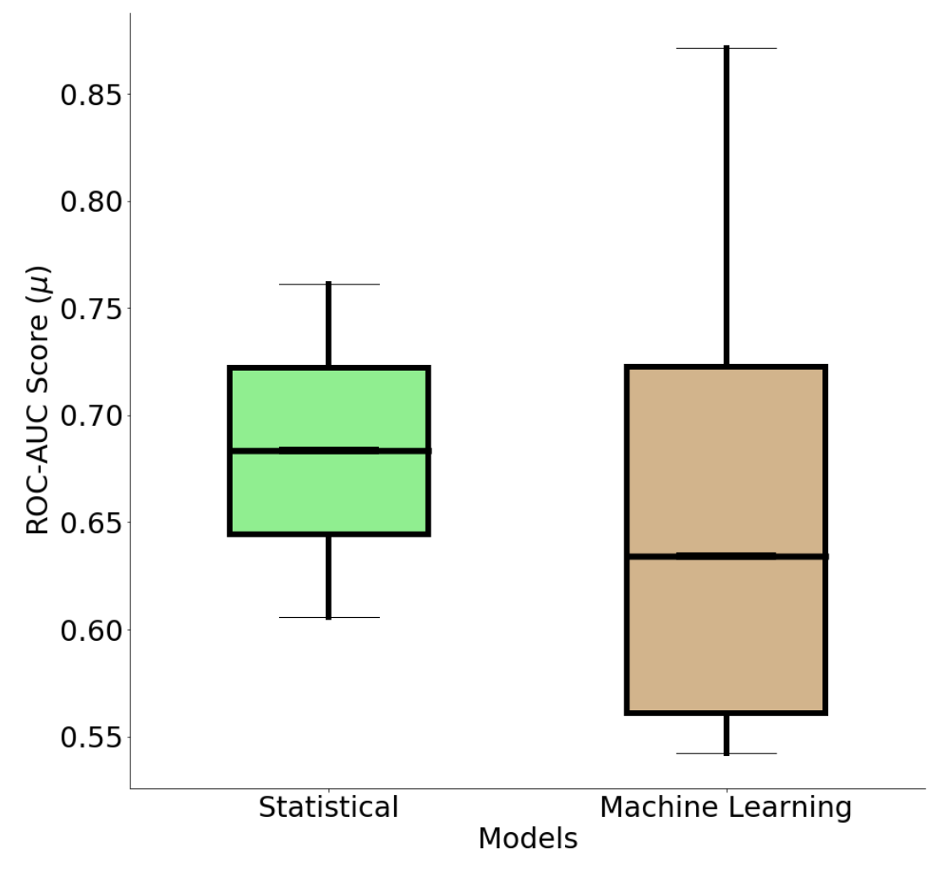


Figure 5: Box and Whisker plot for AUC

6. References

[1] Tony Bellotti and Jonathan Crook. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36(2):3302–3308, 2009.

[2] David West. Neural network credit scoring models. *Computers & Operations Research*, 27(11):1131–1152, 2000.