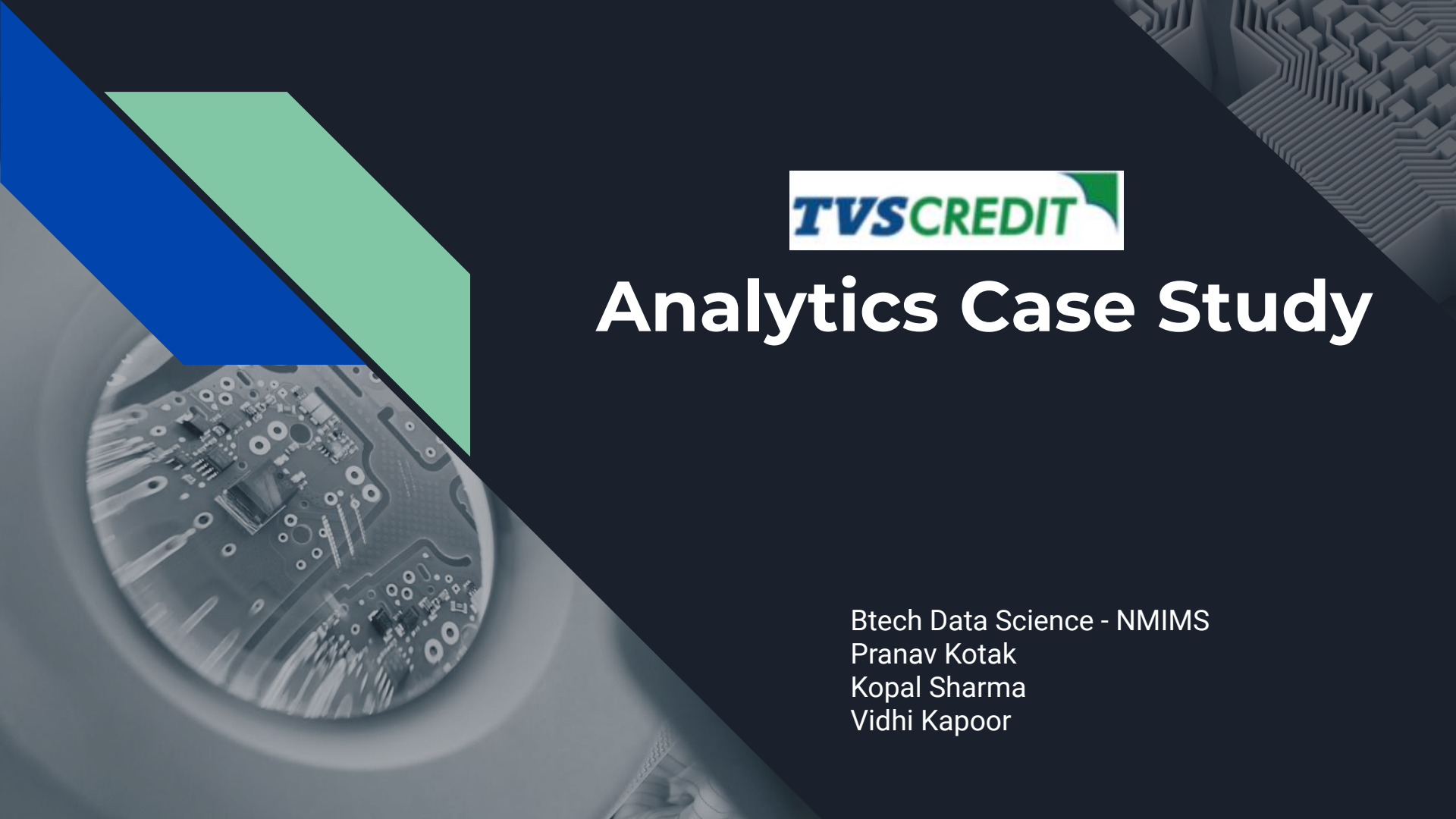


Analytics Case Study

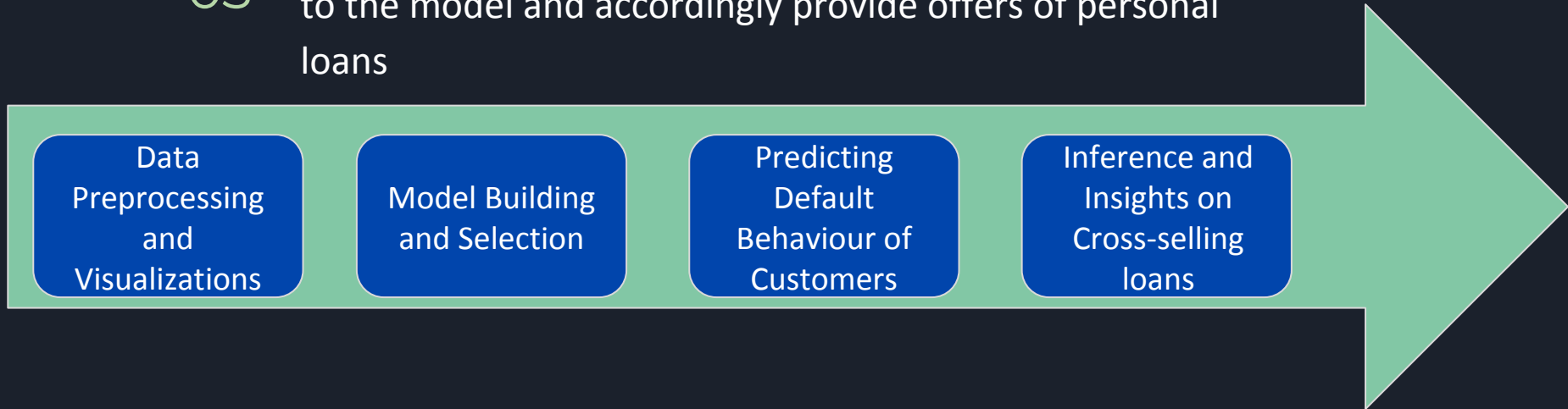
Btech Data Science - NMIMS
Pranav Kotak
Kopal Sharma
Vidhi Kapoor

Understanding the Problem and our Approach

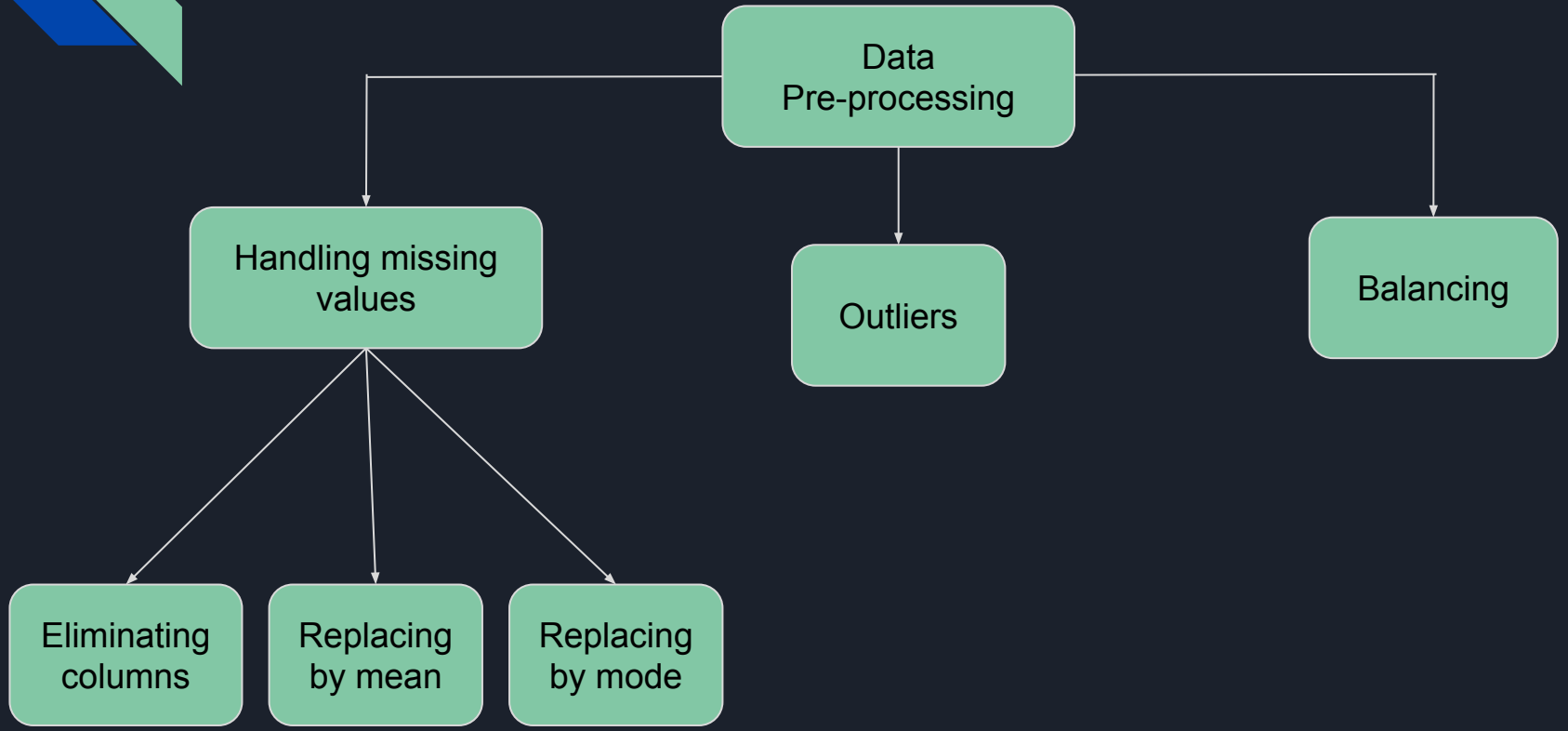
01 To determine the credit worthiness of the customer from the customer two wheeler base

02 To build a risk model which will help in determination of the customers which will be risky or non-risky

03 To find out which features make the most significant impact to the model and accordingly provide offers of personal loans

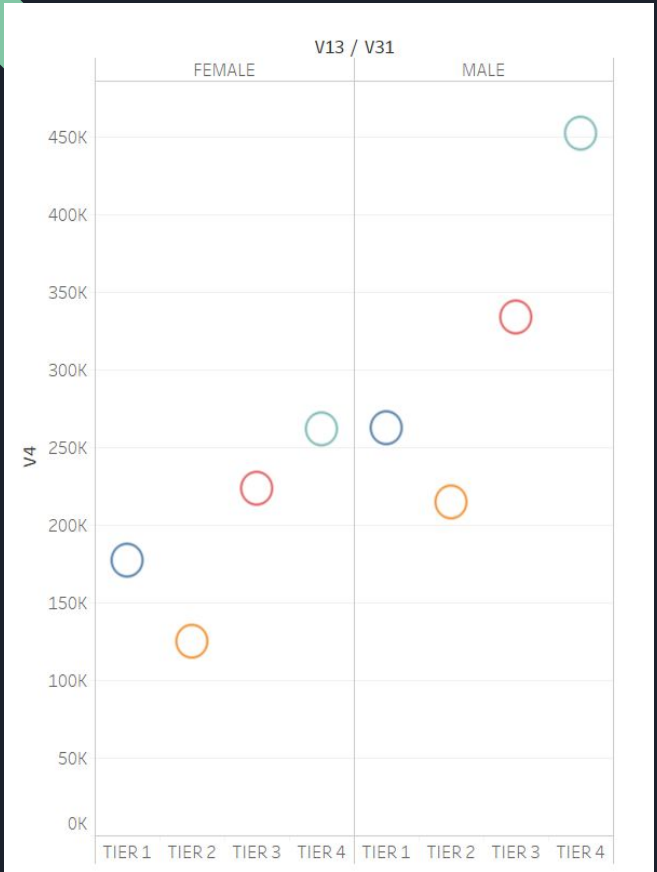


Data Pre-Processing

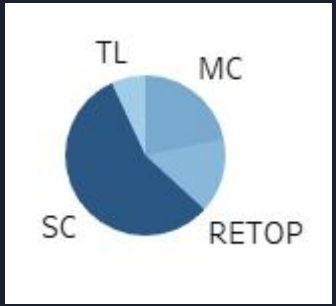


Visualizations

Tier wise male female MOB distribution



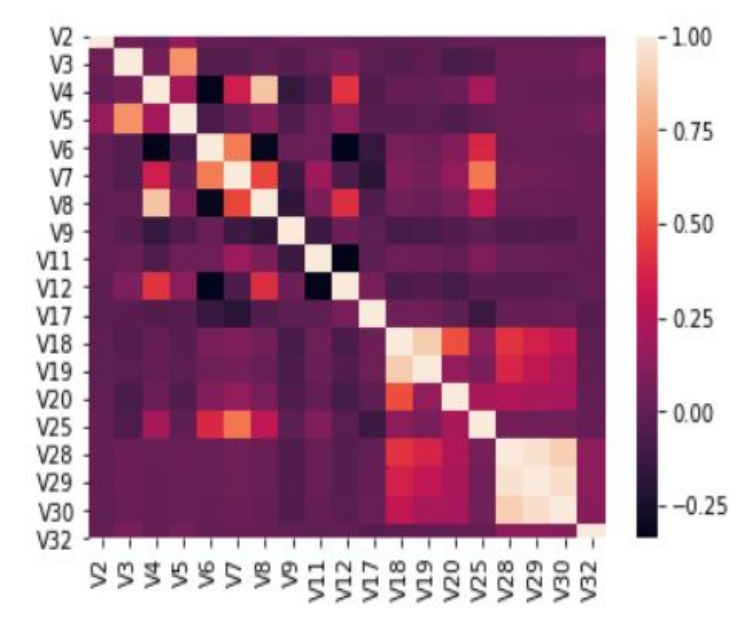
Maximum amount sanctioned for two wheeler typewise



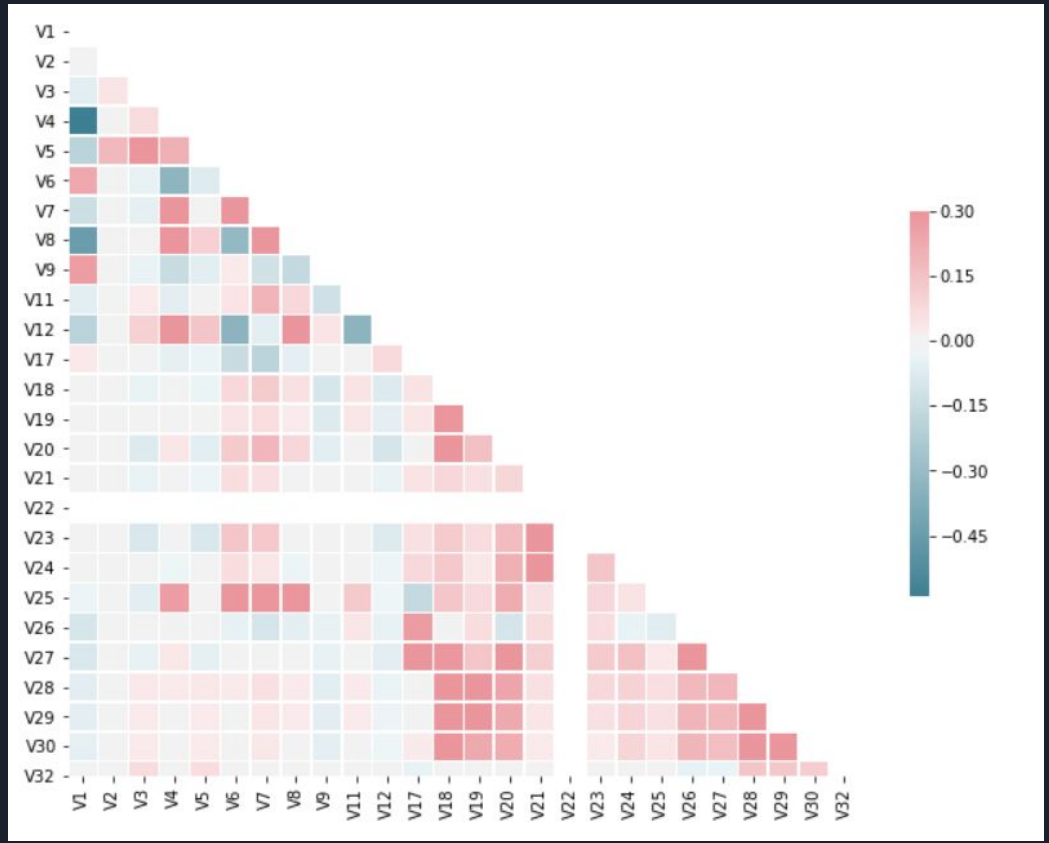
Loan amount employment type wise

Visualizations

Correlation matrix



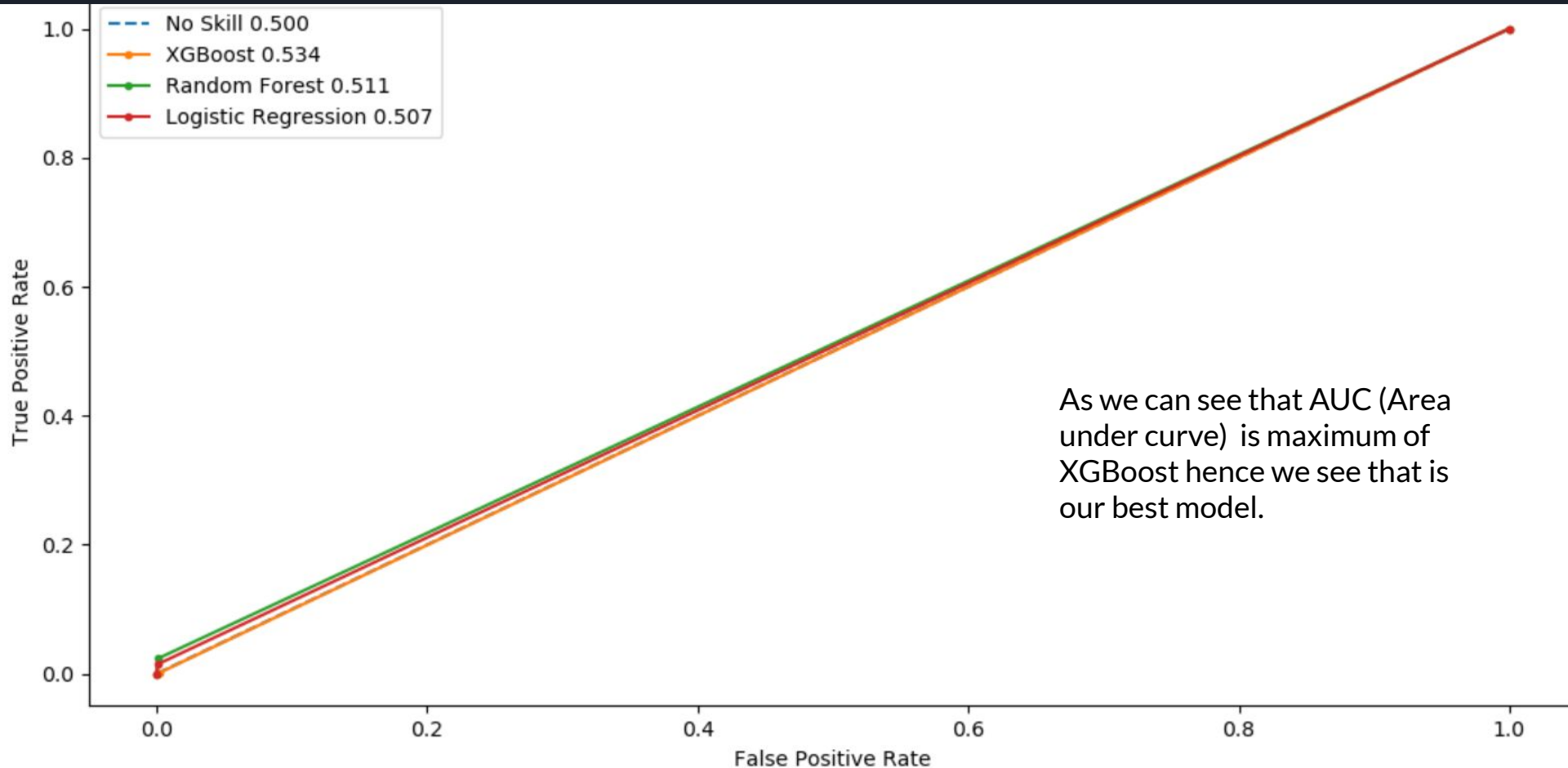
Heat Map



Comparison of different Algorithms

Algorithm Used	Accuracy
Logistic Regression using glm	78.52%
Logistic Regression using Lasso	75.14%
Random Forest Classifier	98.17%
XG Boost	98.63%

Graphical representation of different Algorithms and their AUC



As we can see that AUC (Area under curve) is maximum of XGBoost hence we see that is our best model.

Best Model Performance

It is visible that XGBoost Model has the best accuracy. XGBoost has an accuracy of 98.63% and an AUC of 0.534. Hence it can be used to make further predictions.

The XGBoost library implements the gradient boosting decision tree algorithm.This algorithm goes by lots of different names such as gradient boosting, multiple additive regression trees, stochastic gradient boosting or gradient boosting machines.

Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made.