# Credit Rating using Random Forest Report


**Shivaram Andiappan Selvaraj**

28.01.2018

Graduate Student, NUS

## Problem Description

Bank ABC wants to predict app score for current credit card customers. The app score will denote a customer's credit worthiness and help the bank in reducing credit default risk.

The target variable denoted as Bad_flag_worst6.

If value is 0 – it means customer has good credit history

If value is 1 – it means customer has falls into 30 DPD + bucket

## Methodology


Crisp DM methodology is followed to build and evaluate the model.

## Data Preparation

1. Data Imputation
2. Data Balancing
3. Feature Extraction

## Data Imputation

1. Cashlimit and cur_balance_amt are imputed using zeros.
2. Cur_balance_amt with negative values are replaced with 1.
3. Sed diam nonummy nibh euismod

## Data Balancing

### Synthetic Minority Over-sampling Technique

### Advantages

- Mitigates the problem of overfitting caused by random oversampling as synthetic examples are generated rather than replication of instances
- No loss of useful information

**Before SMOTE**

Total Observations = 34,111

Fraudulent  Observations = 1464

Non Fraudulent Observations = 32647

Event Rate = 4 %

**After SMOTE**

Total Observations = 40104

Fraudulent  Observations = 7457

Non Fraudulent Observations = 32647

Event Rate = 18.5 %

## Feature Extraction

Payment_history_avg_dpd_0_29_bucket, total_diff_lastpaymt_opened_dt, min_months_last_30_plus, utilisation_trend, Ratio_currbalance_creditlimit, mean_diff_lastpaymt_opened_dt, mean_diff_open_enquiry_dt, payment_history_mean_length, max_freq_enquiry, perc_unsecured_others.

The above columns are extracted using the formula given in the questionnaire.

## Insights

| Rule Number | Rule | Cover | Output |
|---|---|---|---|
| 1 | feature_1=Golf Card,Insignia | 2657 (6%) | 1 |
| 2 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh<br><br>dpd_29=12,13,21,30,33 | 794 (2%) | 1 |
| 3 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh feature_51=,American Express Banking Corporation | 942 (2%) | 1 |

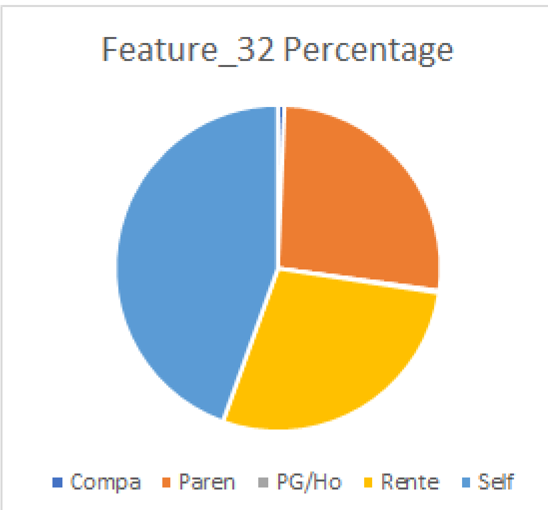
| | | | |
|---|---|---|---|
| | e_total_diff_lastpaymt_opened_dt>=7173<br><br>feature_36=,Government Organisation,MNC,Private Ltd. Co.,Proprietorship | | |
| 4 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh<br><br>dpd_29=?,1,10,11,14,15,16,17,18,19,2,20,22,23,24,25,26,27,28,29,3,31,32,37,38,39,4,40,42,45,5,6,69,7,8,9<br>feature_27=<br><br>feature_51=,American Express Banking Corporation<br><br>e_total_diff_lastpaymt_opened_dt>=7173<br>feature_36=Public Ltd Co. | 398 (1%) | 0 |
| 5 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh<br><br>dpd_29=?,1,10,11,14,15,16,17,18,19,2,20,22,23,24,25,26,27,28,29,3,31,32,37,38,39,4,40,42,45,5,6,69,7,8,9<br>feature_27=<br><br>feature_51=,American | 1786 (4%) | 0 |

| | | | |
|---|---|---|---|
| | Express Banking Corporation<br><br>e_total_diff_lastpaymt_opened_dt< 7173 | | |
| 6 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh<br><br>dpd_29=?,1,10,11,14,15,16,17,18,19,2,20,22,23,24,25,26,27,28,29,3,31,32,37,38,39,4,40,42,45,5,6,69,7,8,9<br>feature_27=<br>feature_51=AXIS Bank,Citibank N.A.,HDFC Bank,HSBC Bank,ICICI Bank,IndusInd Bank Ltd.,Kotak Mahindra Bank Ltd.,Standard Chartered Bank,State Bank of India | 3364 (8%) | 0 |
| 7 | feature_1=Platinum Cricke,Platinum Deligh,Platinum Maxima,RBL Bank Fun+,Titanium Deligh<br><br>dpd_29=?,1,10,11,14,15,16,17,18,19,2,20,22,23,24,25,26,27,28,29,3,31,32,37,38,39,4,40,42,45,5,6,69,7,8,9<br><br>feature_27=Architect,CA,Diploma,Doctor,Engineer,Graduate,MBA/MMS,Others,Post-Graduate,Professional | 32405 (77%) | 0 |

| | | | |
|---|---|---|---|
| | | | |

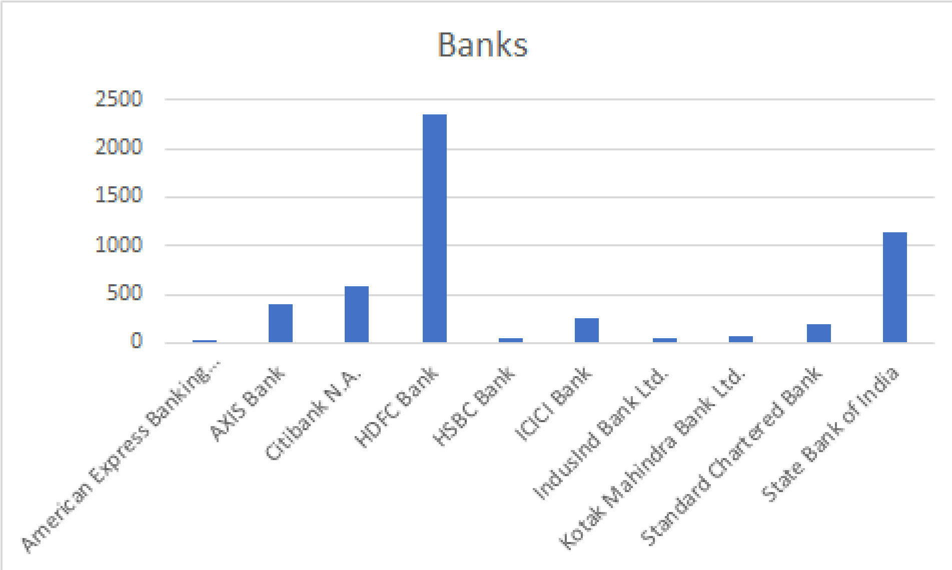
## Feature 32

If a customer is not a credit defaulter then he is likely to stay with Companion 44%, Parents 27%, Rent 28%.



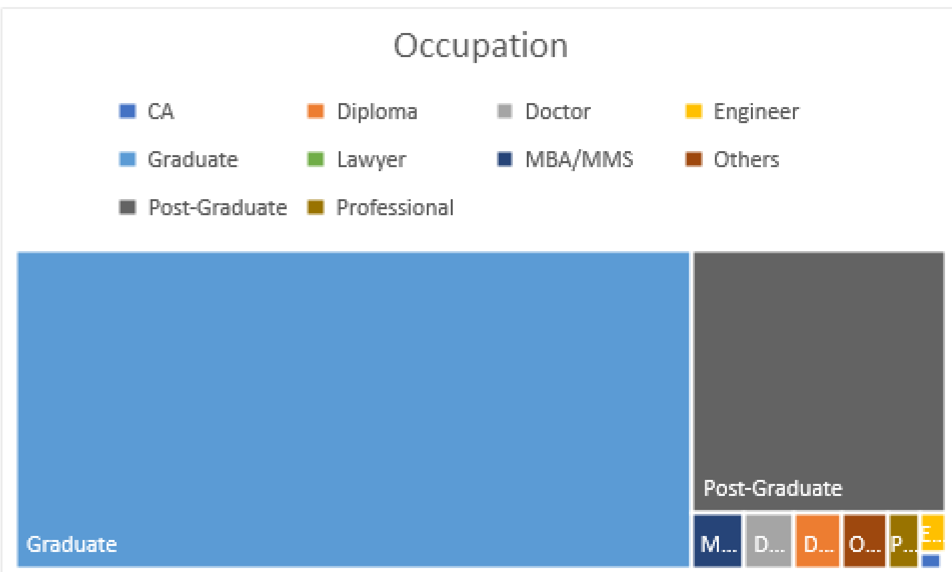

## Customers Bank

Customers having account in HDFC and State Bank of India are safe(loans can be given) and less likely to default.



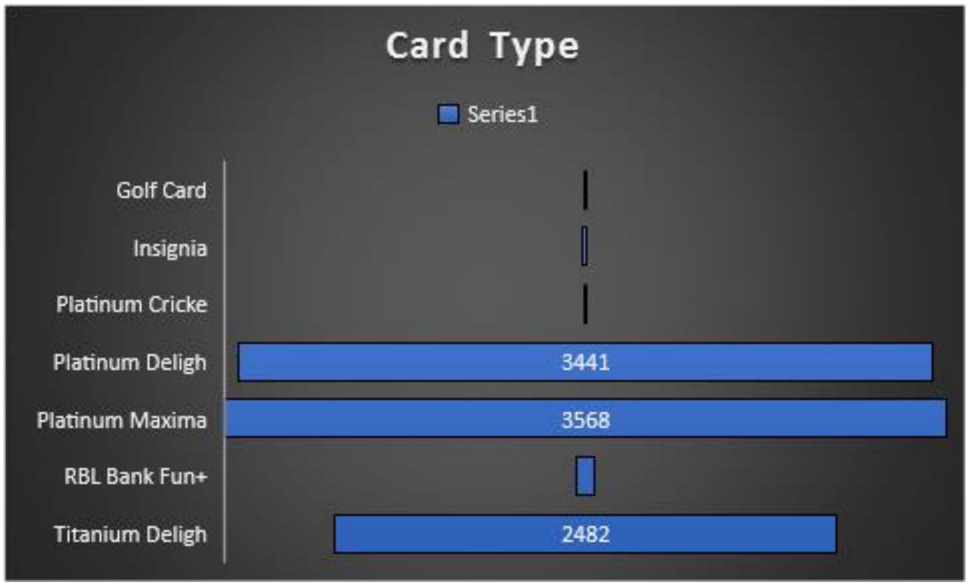

## Occupation

Graduates and Post Graduates are the majority who apply for loan.



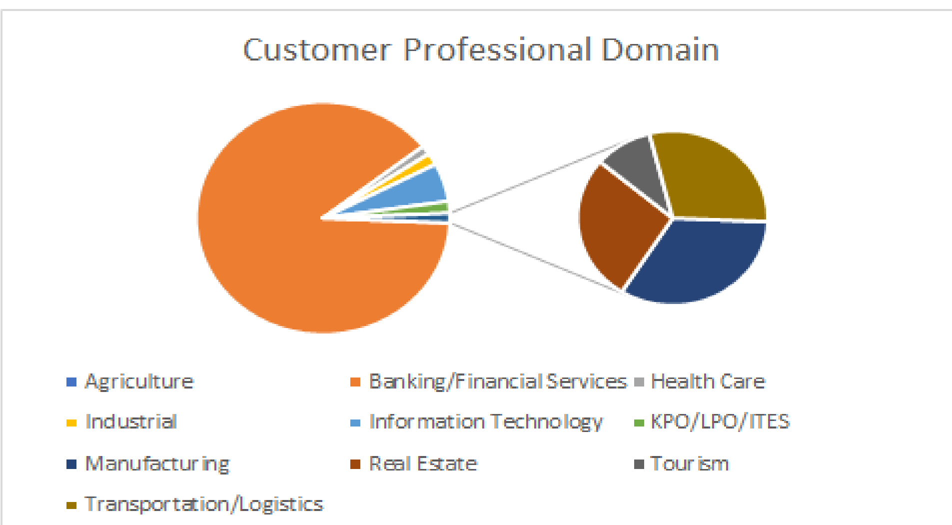

## Card Type

Platinum Maxima, Platinum Deligh and Titanium Deligh card holders are reliable borrowers.



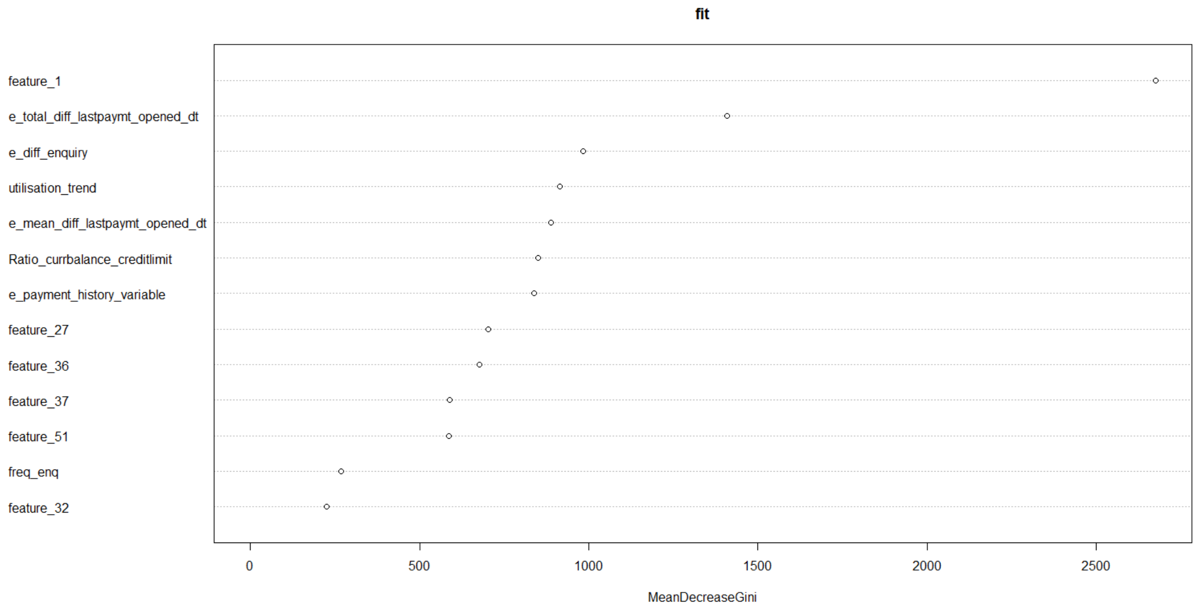

## Customer Professional Domain

Customer from Banking are topping the charts in applying loan and followed by Information Technology.




## RESULTS

## Variable Importance



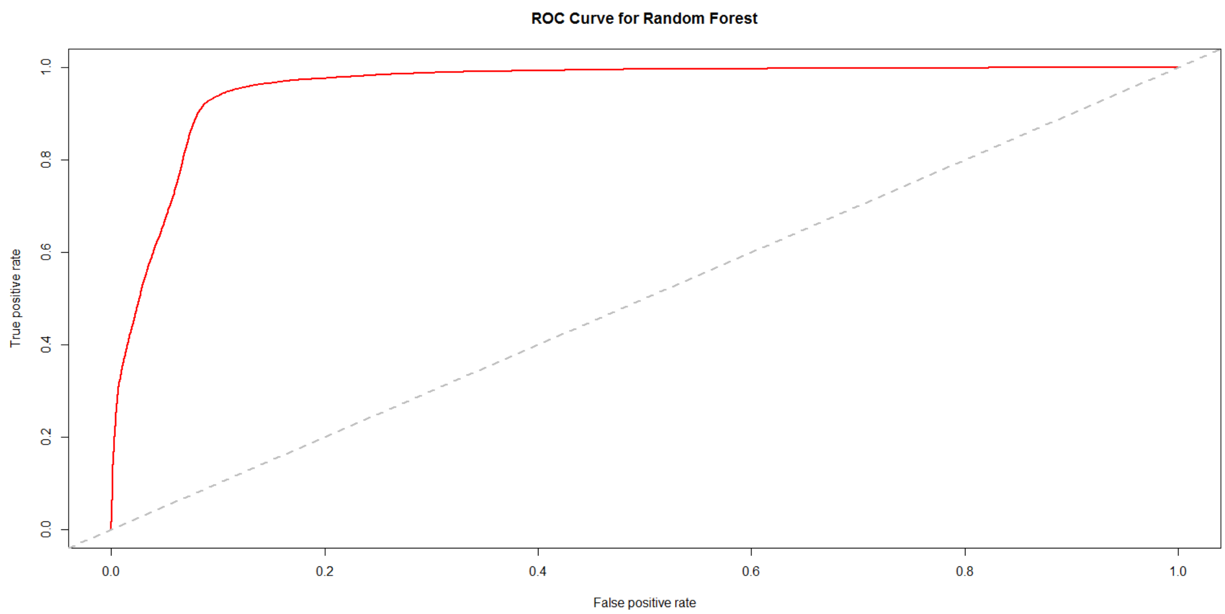

## Confusion Matrix

```
predicted    0    1
        0 9603  454
        1  166    8
```

## Accuracy

```
> accuracy
[1] 0.9393999
```

## ROC




## CONCLUSION

The Model is able to capture major part of the borrowers who could repay the loan but performs poorly in capturing the defaulters. So better to give loans to borrowers who comes under True positive category.

## REFERENCES

1. https://cran.r-project.org/doc/contrib/Sharma-CreditScoring.pdf
2. https://www.investopedia.com/ask/answers/102814/what-difference-between-available-credit-and-credit-limit.asp
3. https://www.investopedia.com/articles/pf/10/credit-score-factors.asp?ad=dirN&qo=serpSearchTopBox&qsrc=1&o=40186
4. http://www.listendata.com/2014/11/random-forest-with-r.html
5. https://www.analyticsvidhya.com/blog/2017/03/imbalanced-classification-problem/
6. https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/