Credit Card Lead Predictions

Pushkar Chougule

Content

Business Problem and Background

- Private bank existing customer base
- Product offerings : Checking / Savings accounts, credit cards, loans, investments
- Identify Potential customers eligible for Credit Card Products and offer to apply
- Revenue Growth with prospective new Credit Card accounts

Data Explanation and Preparation

ID	Gender	Age	Region_Code	Occupation	Channel_Code	Vintage	Credit_Product	Avg_Account_Balance	Is_Active	Is_Lead
NNVBBKZB	Female	73	RG268	Other	X3	43	No	1045696	No	0
IDD62UNG	Female	30	RG277	Salaried	X1	32	No	581988	No	0
HD3DSEMC	Female	56	RG268	Self_Employed	X3	26	No	1484315	Yes	0
BF3NC7KV	Male	34	RG270	Salaried	X1	19	No	470454	No	0
TEASRWXV	Female	30	RG282	Salaried	X1	33	No	886787	No	0

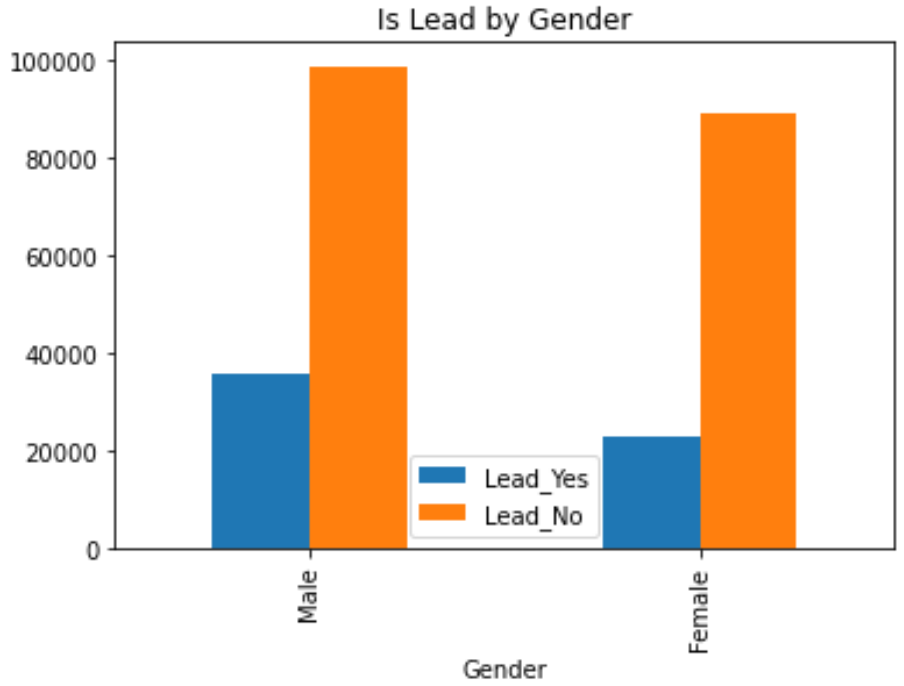
Target Feature

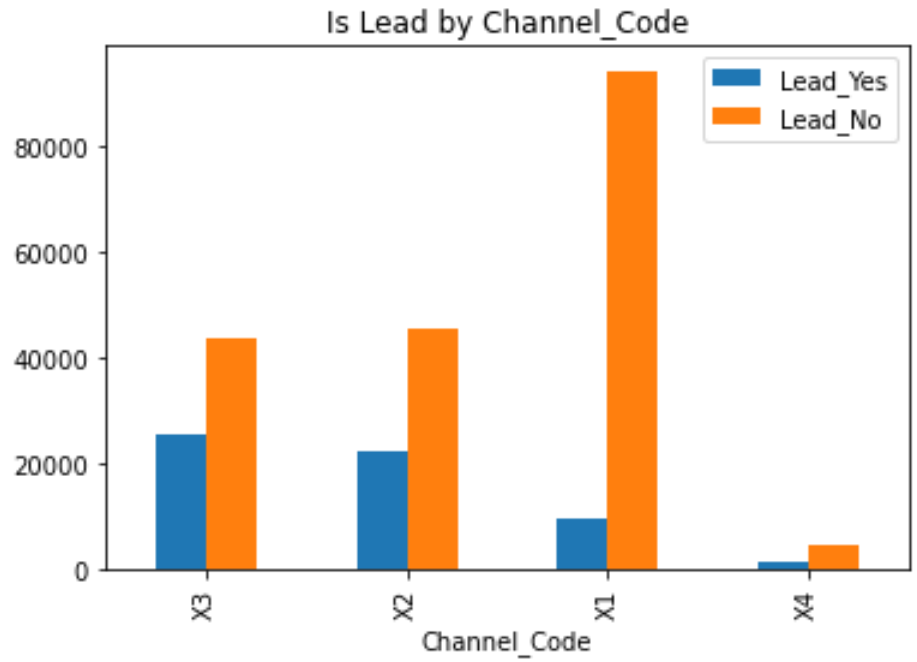
Features with NULL Values

ID	False
Gender	False
Age	False
Region_Code	False
Occupation	False
Channel_Code	False
Vintage	False
Credit_Product	True
Avg_Account_Balance	False
Is_Active	False

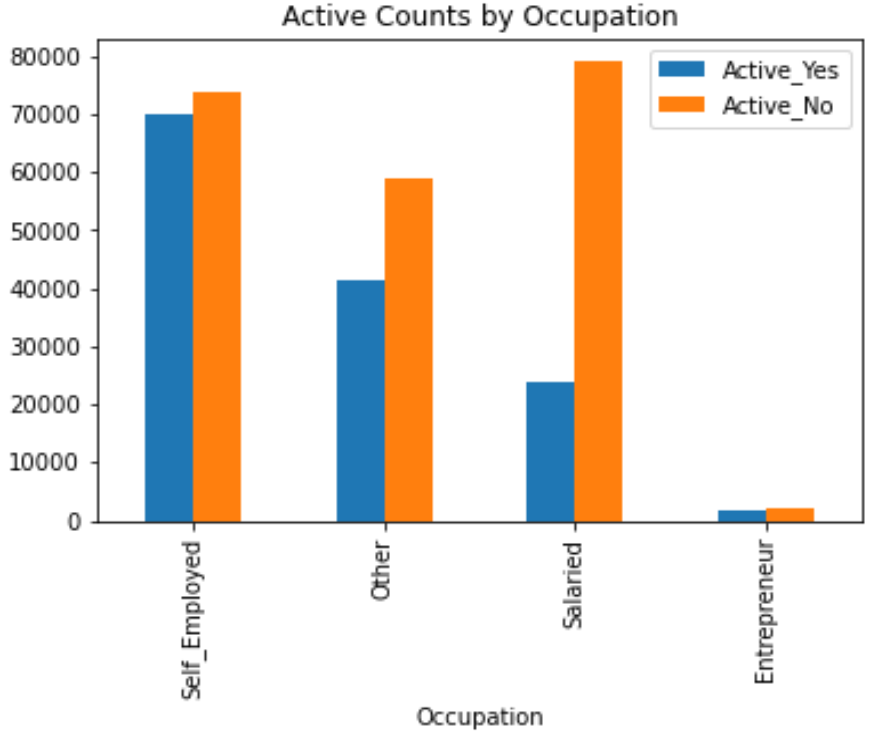
ID	0
Gender	0
Age	0
Region_Code	0
Occupation	0
Channel_Code	0
Vintage	0
Credit_Product	29325
Avg_Account_Balance	0
Is_Active	0

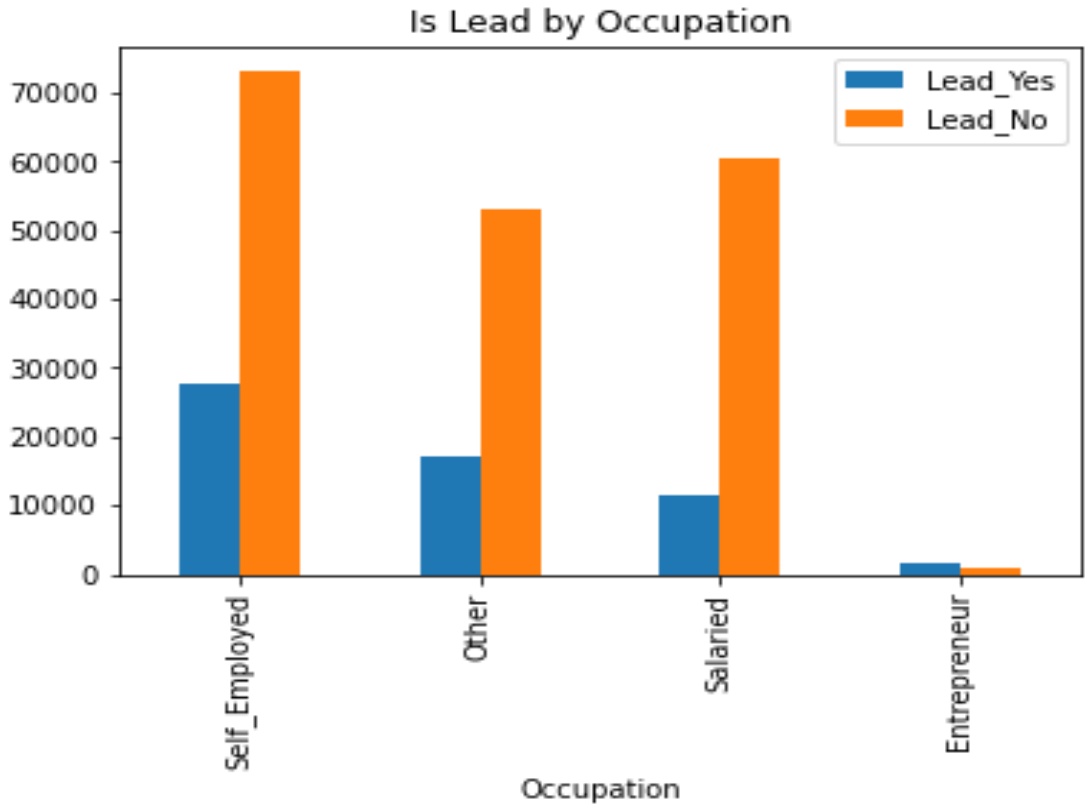
Features Analysis and Methods

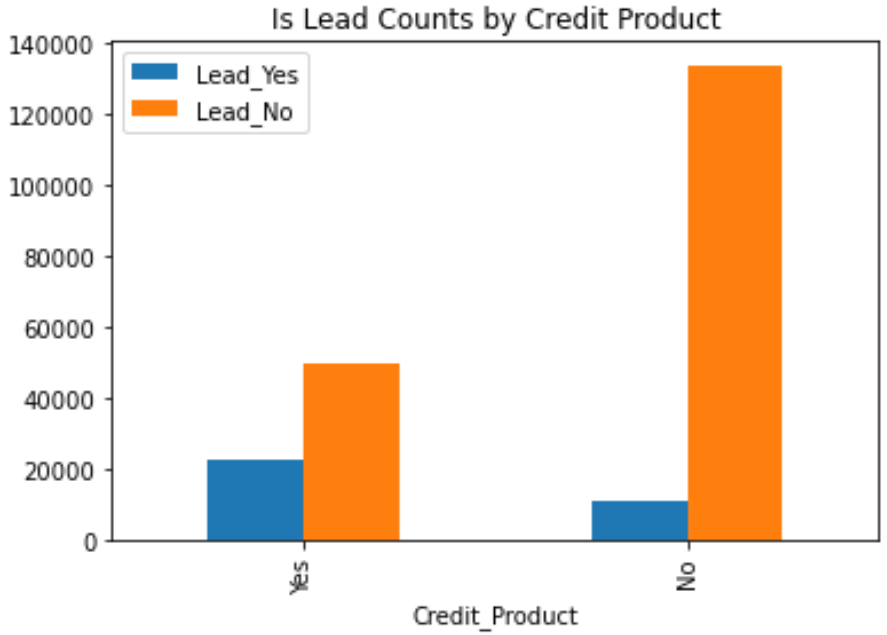




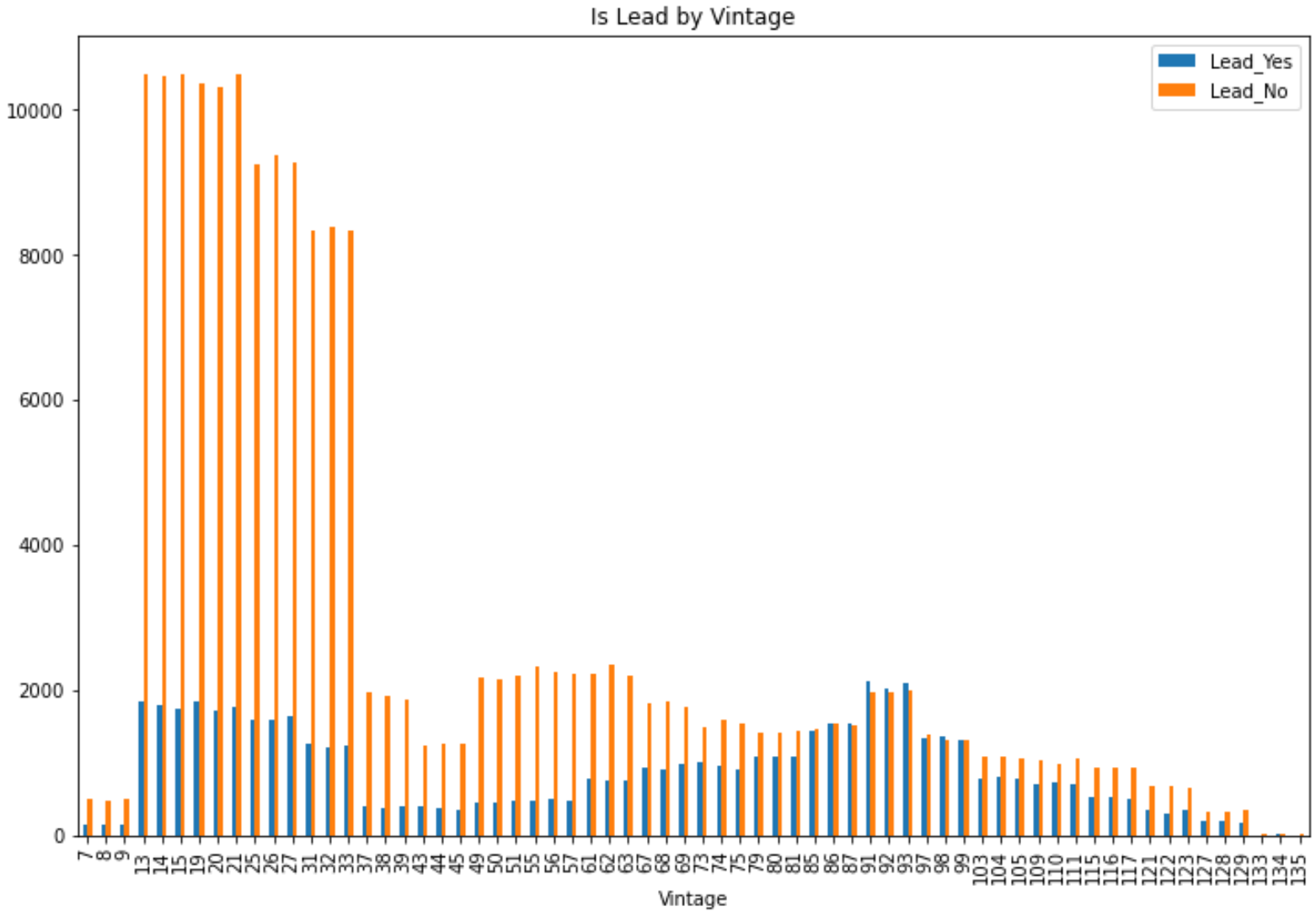
Features Analysis and Methods (continued)

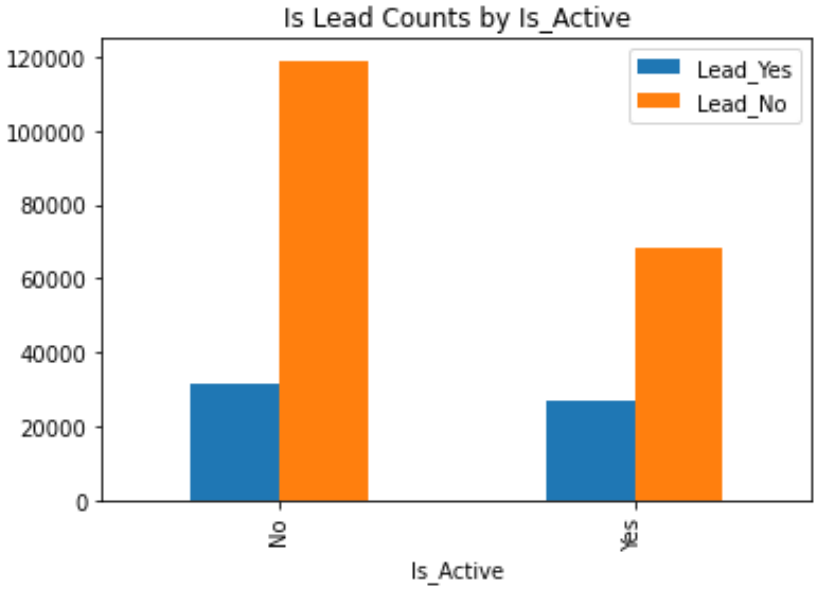


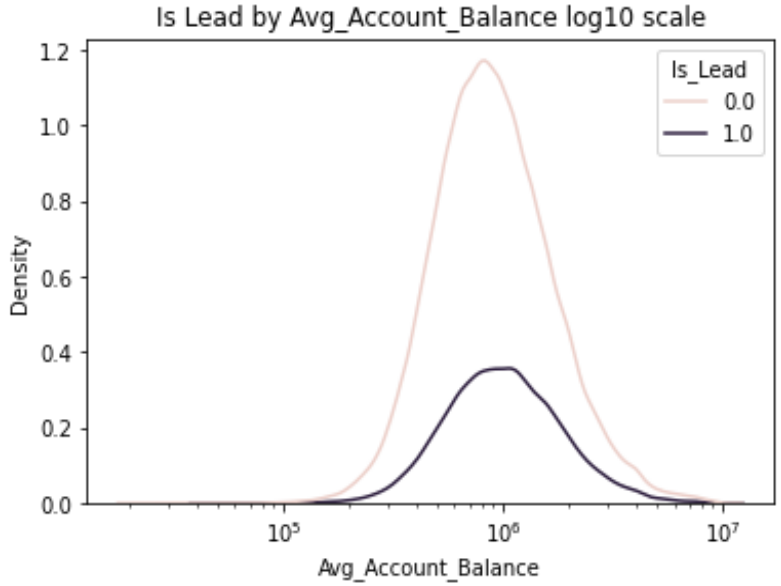




Features Analysis and Methods (continued)







Machine Learning Models and Evaluation

- Logistic Regression
- Random Forest Classifier
- Gaussian Naïve Bayes
- Gradient Boost classifier
- ADA Boost classifier
- XG Boost classifier

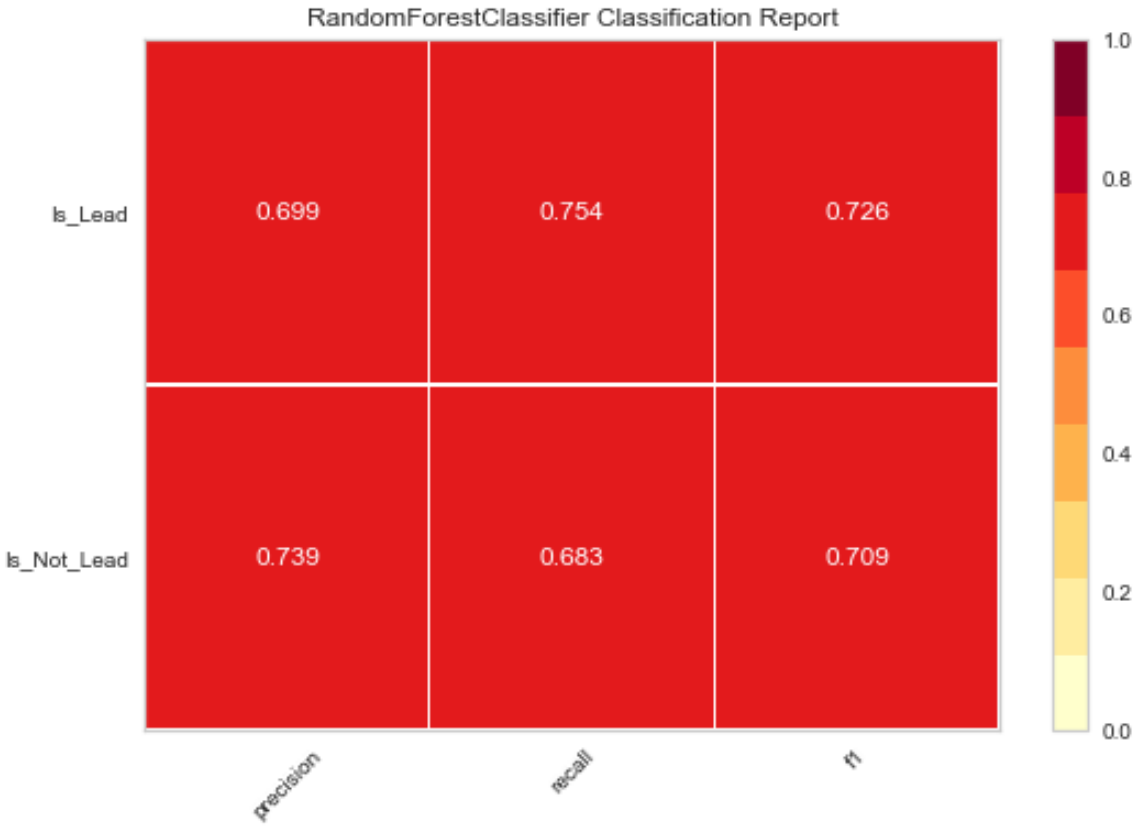
Training results

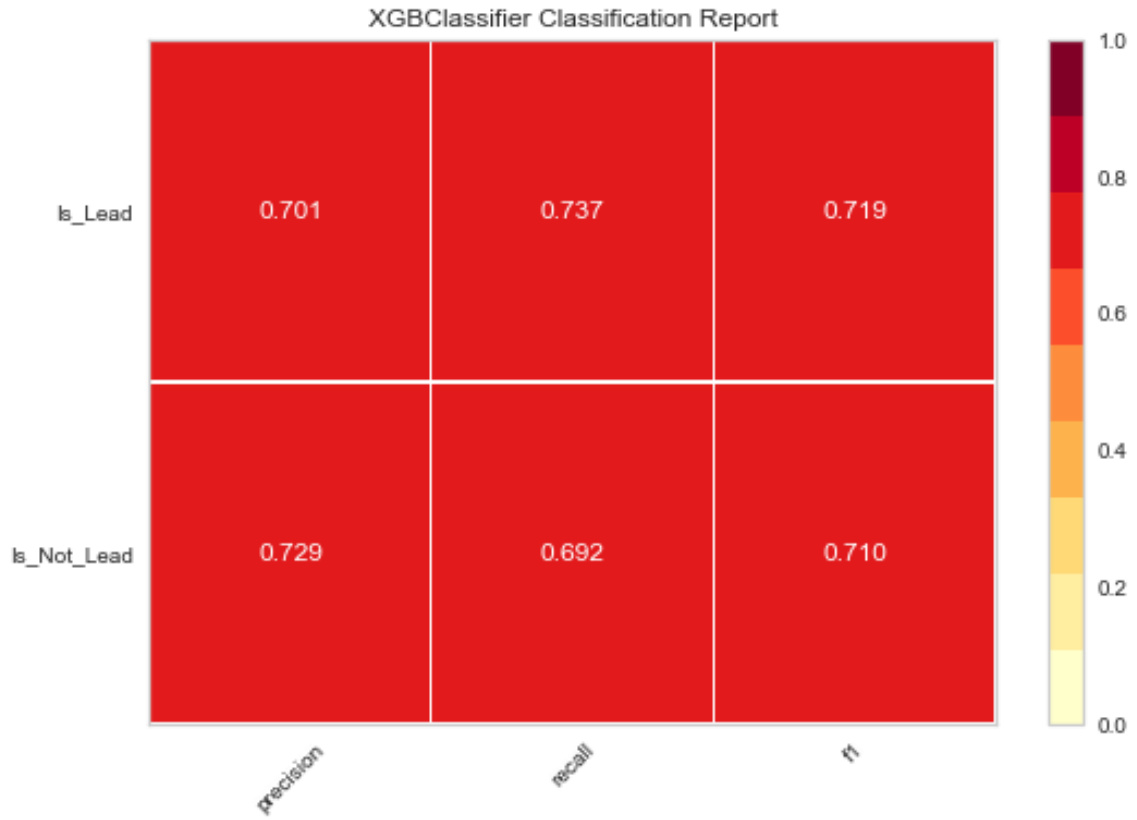
```
LogisticRegression - 65.9%
RandomForestClassifier - 74.2%
GradientBoostingClassifier - 71.6%
XGBClassifier - 74.6%
```

Evaluation results

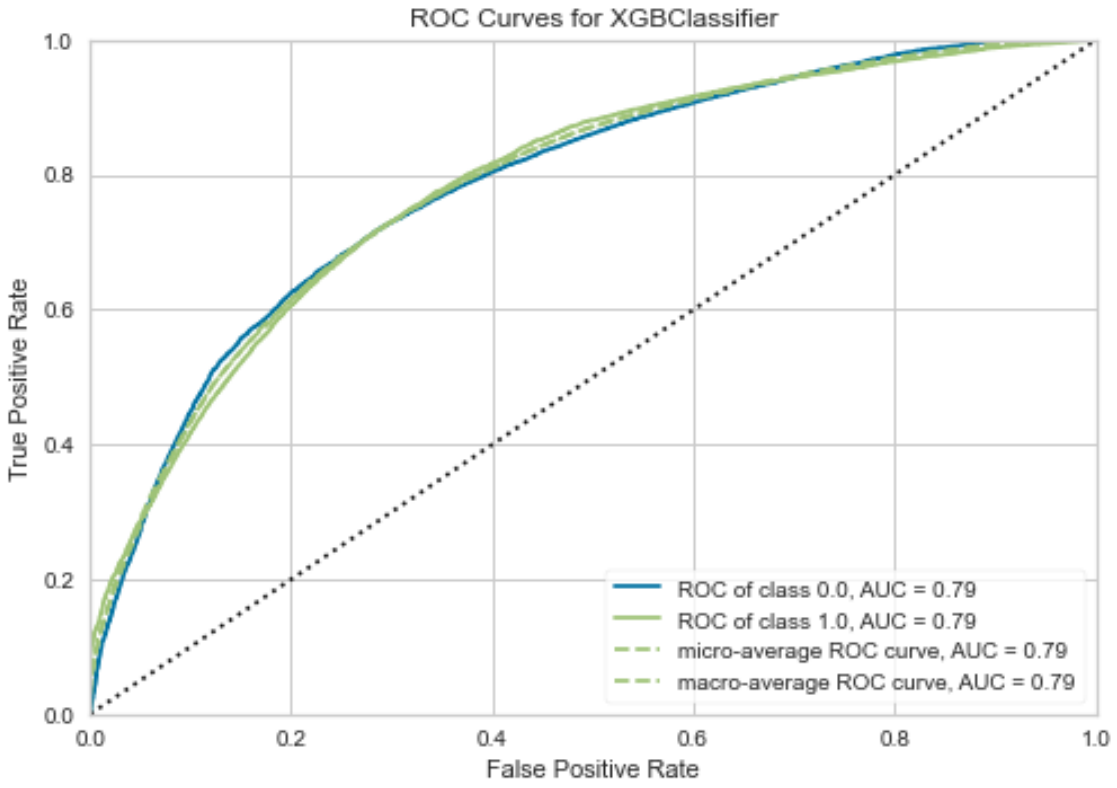
```
LogisticRegression - 65.9%
RandomForestClassifier - 72.0%
GradientBoostingClassifier - 71.5%
XGBClassifier - 71.7%
```

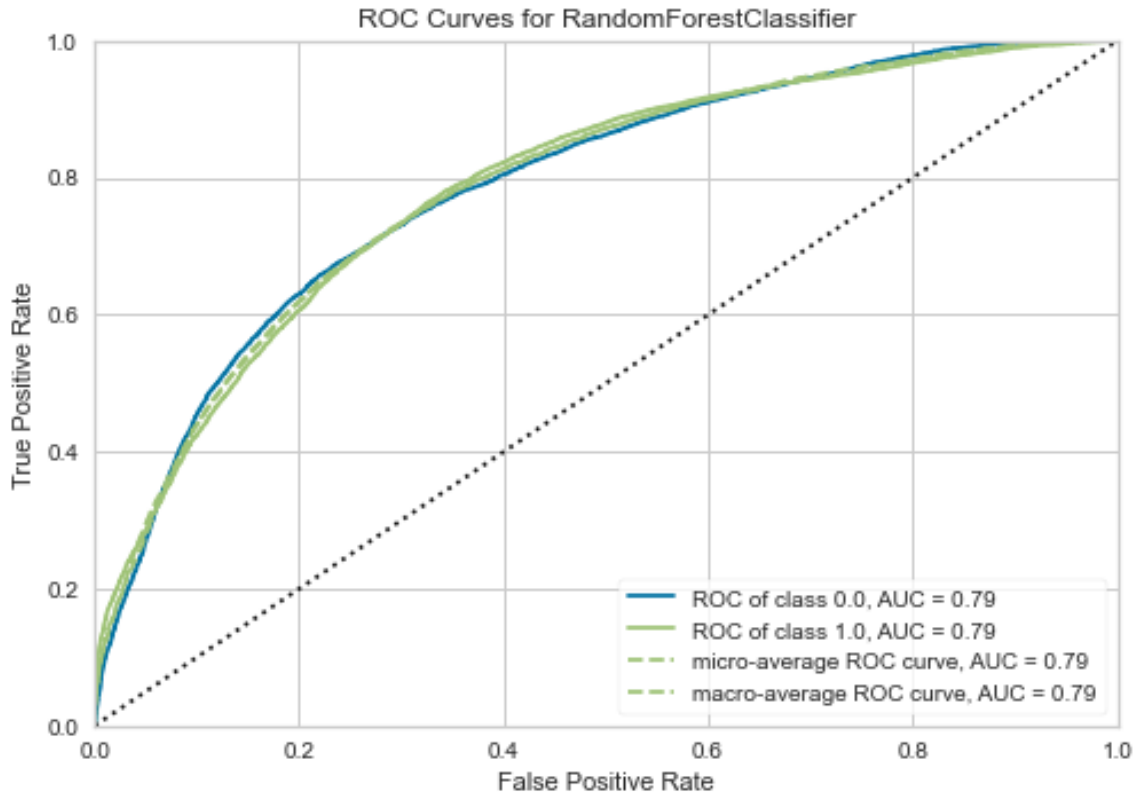
Machine Learning Models and Evaluation
(Classification reports)





Machine Learning Models and Evaluation (ROC curves)





Conclusion

- Certain Age groups showed up as potential candidates along with specific communication channels which were highly effective
- Having existing credit products with the bank was a positive influencer for potential leads
- Self employed customers came up as high potential leads as compared to entrepreneurs
- Vintage turned out to be a handy factor as well along with average account balance
- Currently Active Customers stand a good chance of being potential leads
- Following the potential prediction leads from the designed system could help expand the credit card customer base and thereby add to the revenue of the bank

Assumptions / Limitations

Assumption: with mostly categorical features – all details captured correctly and no outliers present

Limitation : Imbalanced data due to lower size of training data having potential positive leads as compared to non-leads

Future Usage and Recommendations

The modeling approach can be reused for other product offerings like loans or investment products by capturing relevant additional features

Recommend for future improvements, capture and share credit scores and annual income as factors

References

- Kaggle Dataset. https://www.kaggle.com/sajidhussain3/jobathon-may-2021-credit-card-lead-prediction

- Gupta, Aman (2021) https://medium.com/geekculture/xgboost-versus-random-forest-898e42870f30

- Difference Between Random Forest vs XGBoost https://www.educba.com/random-forest-vs-xgboost/

THANK YOU