Predicting the Outcome of 10Alytic Bank Direct Marketing Campaign

CHIKE CHUKWUMAH

HTTPS://GITHUB.COM/CMCHUKWUMAH/BANK-MARKETING-CAMPAIGN.GIT

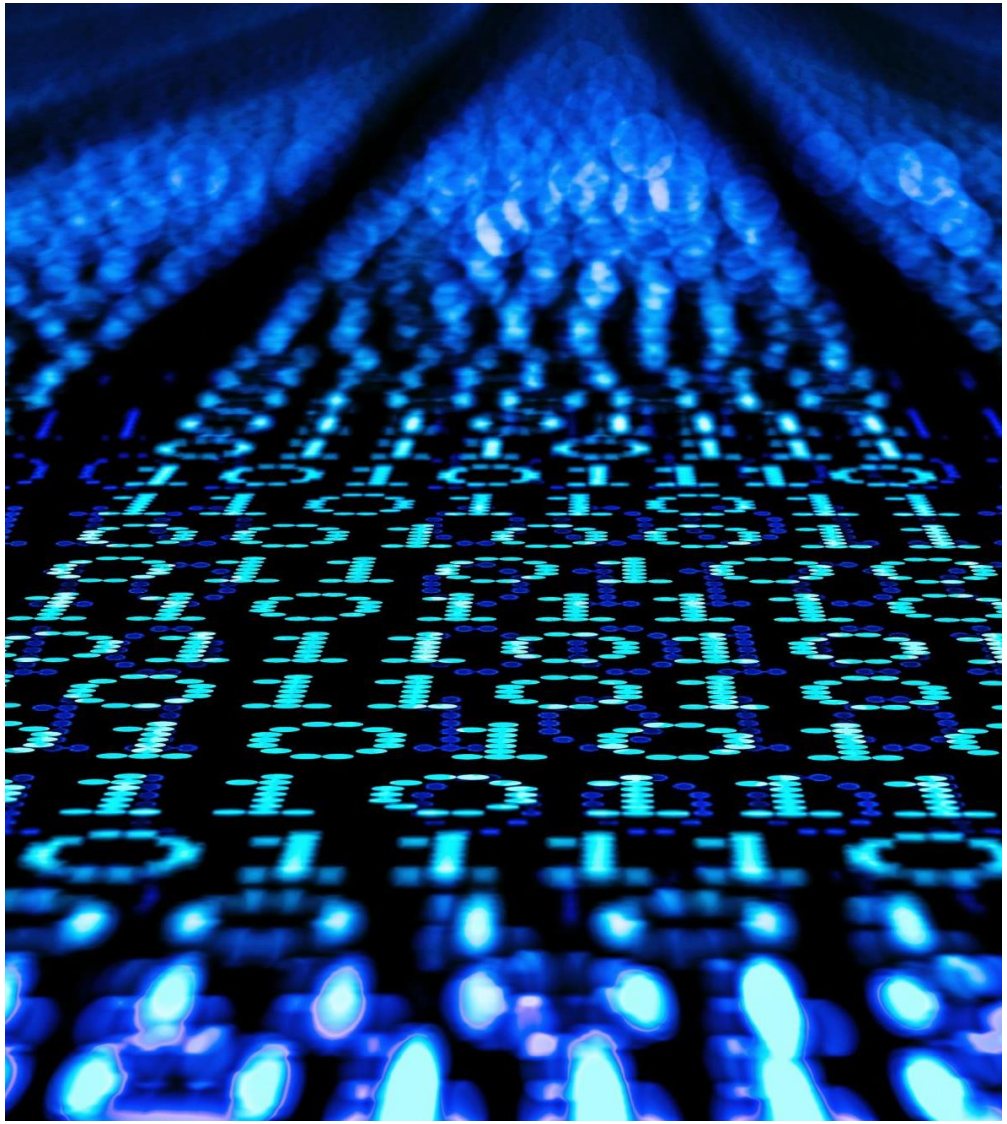
INTRODUCTION

10Alytics, a Banking Institution launched a new product (Term Deposit) and have decided to engage in a direct marketing campaign to access if its customers would subscribe to the product or not.

Problem Statement - To predict if the Bank Customers contacted through the marketing campaign will subscribe to the new product (term deposit)

Data Description

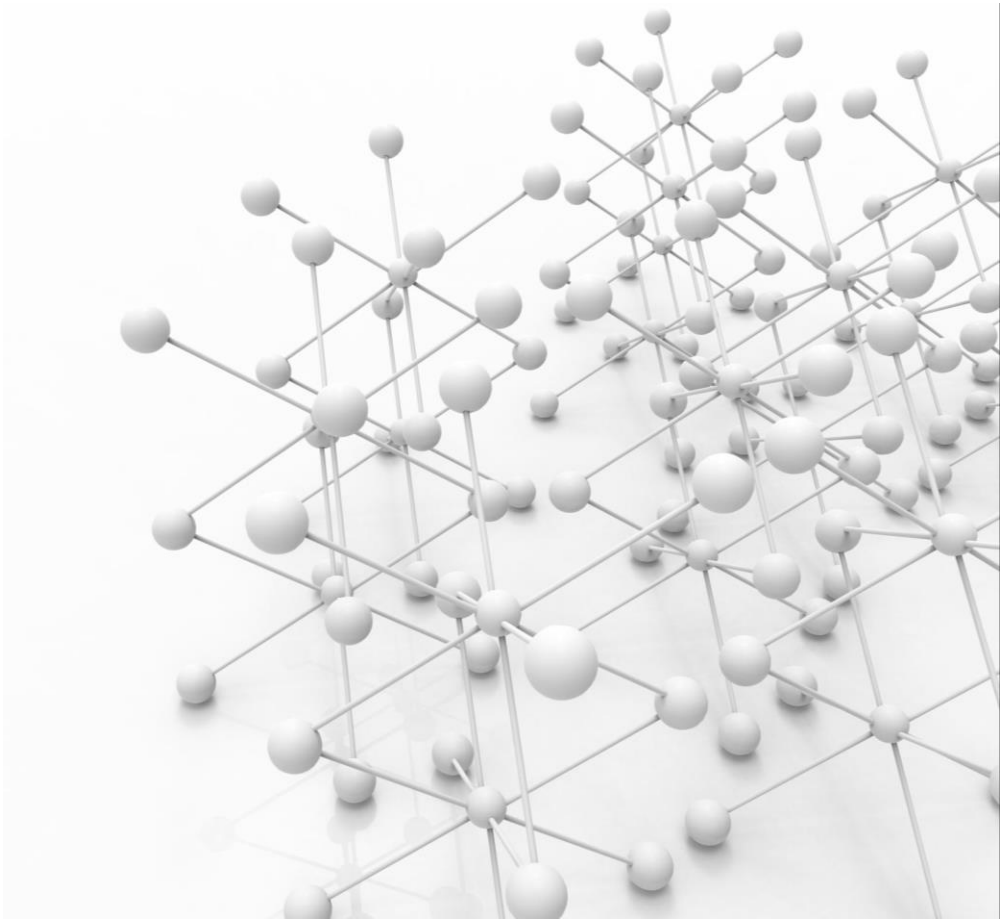
- The data consisted of 4521 rows and 17 columns
- Our dataset also consisted of both categorical and numerical features with 16 independent features, out of these only half of them are important.
- Data was cleaned, manipulated and visualized.



Data Modelling, Testing and Evaluation

- Data was modelled, trained and tested using the following methods

1. Logistic Regression
2. K-neighbor Classifier
3. Decision Tree Classifier
4. Random Forest Classifier



Logistic Regression Classifier

For LogisticRegression, Accuracy score is 0.8901989683124539

	precision	recall	f1-score	support
0	0.98	0.90	0.94	1310
1	0.18	0.62	0.28	47
accuracy			0.89	1357
macro avg	0.58	0.76	0.61	1357
weighted avg	0.96	0.89	0.92	1357

K-Neighbor Classifier

For KNeighborsClassifier, Accuracy score is 0.8769344141488578

	precision	recall	f1-score	support
0	0.97	0.90	0.93	1288
1	0.19	0.45	0.27	69
accuracy			0.88	1357
macro avg	0.58	0.67	0.60	1357
weighted avg	0.93	0.88	0.90	1357

Decision Tree Classifier

For DecisionTreeClassifier, Accuracy score is 0.871039056742815

	precision	recall	f1-score	support
0	0.93	0.93	0.93	1198
1	0.45	0.45	0.45	159
accuracy			0.87	1357
macro avg	0.69	0.69	0.69	1357
weighted avg	0.87	0.87	0.87	1357

Random Forest Classifier

For RandomForestClassifier, Accuracy score is 0.8931466470154753

	precision	recall	f1-score	support
0	0.99	0.90	0.94	1314
1	0.18	0.67	0.29	43
accuracy			0.89	1357
macro avg	0.58	0.79	0.61	1357
weighted avg	0.96	0.89	0.92	1357

Conclusion

- Our dataset consist of categorical and numerical features. We have 16 independent features, out of these only half of them are important.
- Accuracies of all models are about 87 - 89%
- The high accuracies could be associated with the correlation between the input features.
- The Random Forest model should be deployed for production as it the model having the best accuracy as well as Precision and recall.

THANK YOU

HTTPS://GITHUB.COM/CMCHUKWUMAH
/BANK-MARKETING-CAMPAIGN.GIT