THE iSCHOOL
Syracuse University

PREDICTION FOR CLIENT TERM DEPOSIT SUBSCRIPTION

INSTRUCTOR
Ying Lin

TEAM
Adheesh Phadnis, Lakshya Gupta, Vidushi Mishra

CONTEXT

The core business of a financial institution can be broadly classified as lending and borrowing. Lending generates revenue to the bank in the form of interest from customers with some level of default risk involved. A low risk strategy for the banks is to attract public savings into the bank. To achieve this goal, one of the popular ways for the banks is telemarketing. It is an interactive technique of direct marketing via the phone which is widely used by banks to sell long -term deposits. The bank telemarketing data used here is related with direct marketing campaigns of a Portuguese bank institution.

PROBLEM AND OBJECTIVE

The aim of this study is to predict whether a client is going to subscribe a long-term deposit through telemarketing strategy. Further, this study will analyze the characteristics of the clients who are predicted to invest in the long-term deposits. The bank can then utilize this information to allocate resources to focused customers and thus increase their revenue.

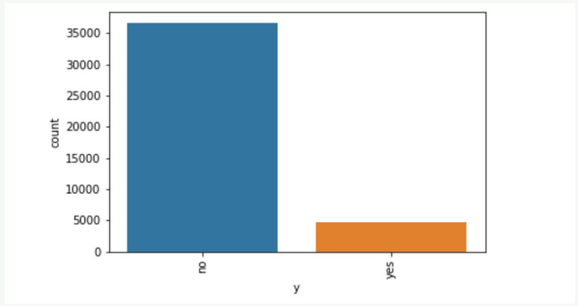
DATA DESCRIPTION

The data contains 41188 instances and 20 features.
Feature names:
Bank client data such as age, type of job, marital status, education, has credit in default, balance, housing
Target variable: Term deposit subscription (yes, no)
Source link:
https://archive.ics.uci.edu/ml/datasets/bank+marketing#

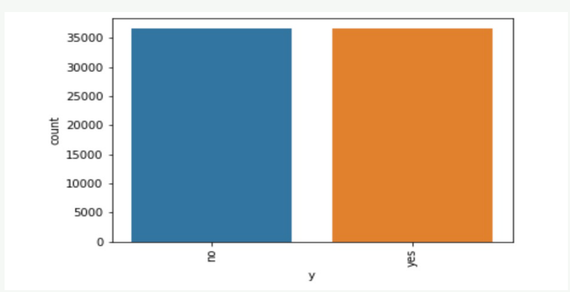
PREPROCESSING

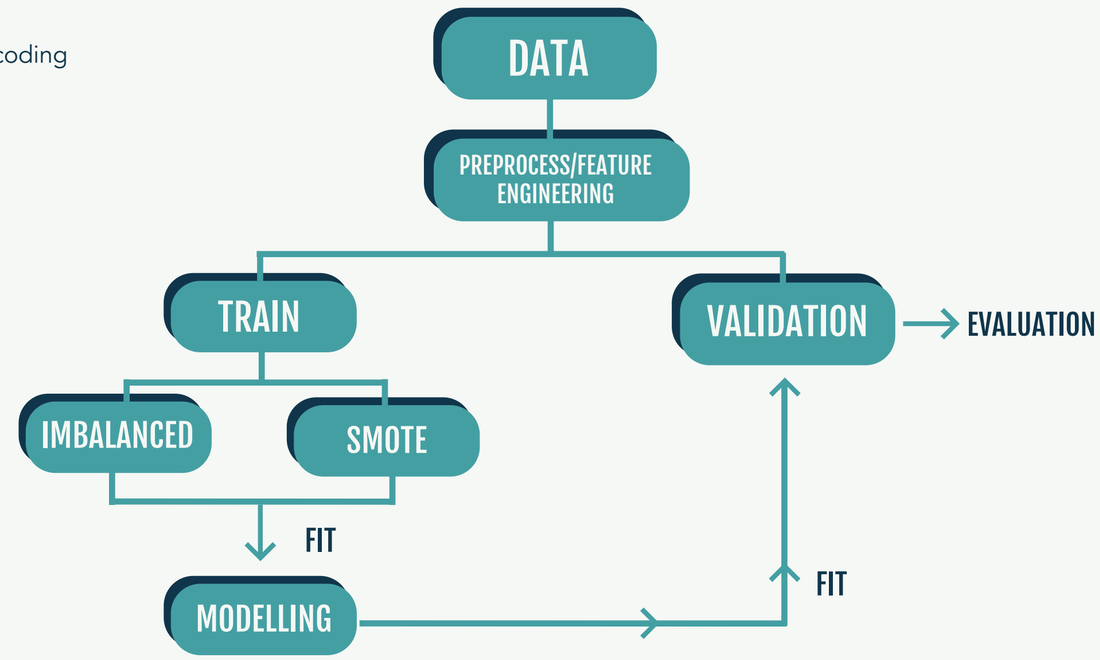
data visualisation, analysis, treating imbalanced data, one hot encoding

IMBALANCED



SMOTE





MODELLING DETAILS

Models run:
1. SVM
2. GBM
3. KNN
4. Random forest
5. ANN
6. Decision tree
7. Logistic Regression

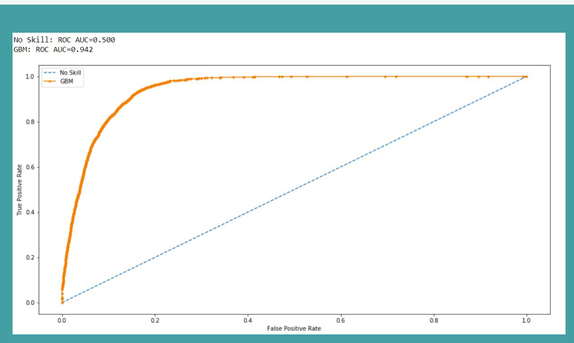
TOP THREE PERFORMERS SPECIFICATIONS

MODEL	HYPERPARAMETERS
KNN	n_neighbors = 15
GBM	max_depth= 5, max_features= 25 ,n_estimators = 500
RANDOM FOREST	max_depth= 8, max_features= 25 ,n_estimators = 500

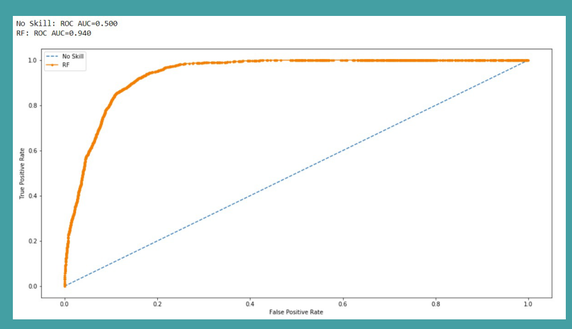
MODEL EVALUATION

METRIC \ MODEL	KNN	GBM	RANDOM FOREST
Data	SMOTE	SMOTE	SMOTE
Accuracy	83%	91%	86%
Balanced Accuracy	86%	81%	87%
Precision	40%	60%	45%
Recall	89%	67%	89%

GBM

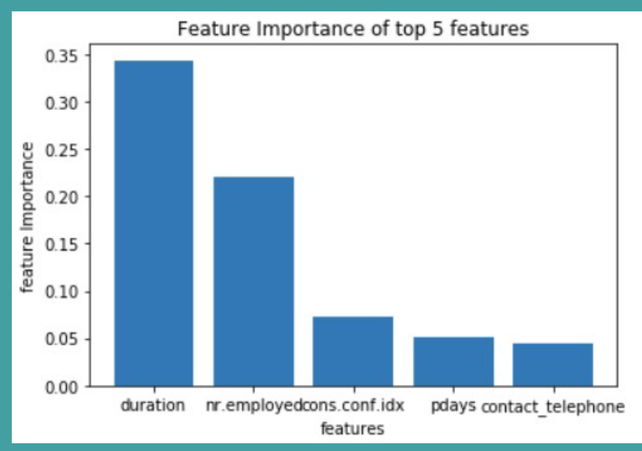


RANDOM FOREST



CONCLUSION

Adhering to the principle of 'No free lunch', extensive modelling experiments using various preprocessing and machine learning algorithms was done. The top models are used to develop a functional web app with allows the users to learn key parameters of the models and evaluation metrics.



APPLICATION

This model can be further used to analyse the characteristics of customers who have higher likelihood of subscribing to the term deposits.Duration euribor and nr. employed are the top three features that contribute to the predictive model. It might allow the bank to allocate the resources judiciously to obtain maximum returns from the investments in marketing campaign.