

modeling\_banking\_customers

Public

1 Branch

0 Tags

Go to file

t

Go to file

About

Add file

Code

MattLeRoi

 Add files via upload

cc4510d · 3 minutes ago

catboost\_i... finishing notes 40 minutes ago

data update readme 11 hours ago

images update readme 10 hours ago

.gitignore Add data\_for\_eval... 3 weeks ago

Bank-Mark... Add files via upload 3 minutes ago

README.md Update README.... 3 minutes ago

index.ipynb Final version usin... 18 minutes ago

README

# Bank Marketing Campaign Analysis

Author: [Matt LeRoi](#)

No description, website, or topics provided.

Readme

Activity

0 stars

0 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

Jupyter Notebook 100.0%



## Overview

---

The banking industry runs on trust - a customer's faith that the money they are giving to the bank will be held safely and that the bank's promises will be kept. The acquisition cost of a customer can be high, but once a relationship is established it can be lasting and valuable.

## Business Understanding

---

A Portuguese banking institution recorded data about the potential customers targeted for a phone marketing outreach campaign. This involved one or a series of calls. A more detailed understanding of the results of these campaigns could show which customers are more likely to sign, or show which details about the interactions are most likely to lead to customers signing.

## Data

---

The bank provided data on ~45,000 customers, including background information on the client (such as age, job, marital status, etc.) and details about their interactions (such as how frequently and how long ago they were contacted). About 12% of the customers signed.

## Modeling

---

Several baseline models were created for this project (LogisticRegression, DecisionTreeClassifier, XGBClassifier, LGBMClassifier, and CatBoostClassifier), and one was selected for further tuning. These were all used as binary classifiers to predict whether a customer would sign or not. 75% of the data was used for training, 10% as a validation set to evaluate and tune the models, and 15% as a final test set for the final tuned model.

## Evaluation

---

The bank would like to know which customers to give special attention to, but also doesn't want to waste its energy on customers unlikely to sign. The Area Under the Curve of the Receiver Operating Characteristic curve (ROC-AUC) is a metric that takes into account both the true positive and false positive rates, and will be used for evaluation and comparison between models. A general sense of how to interpret the result is [published here](#) and reproduced below:

1.0: Perfect classifier

0.9-1.0: Excellent discrimination

0.7-0.9: Acceptable discrimination

0.5-0.7: Poor to inadequate discrimination

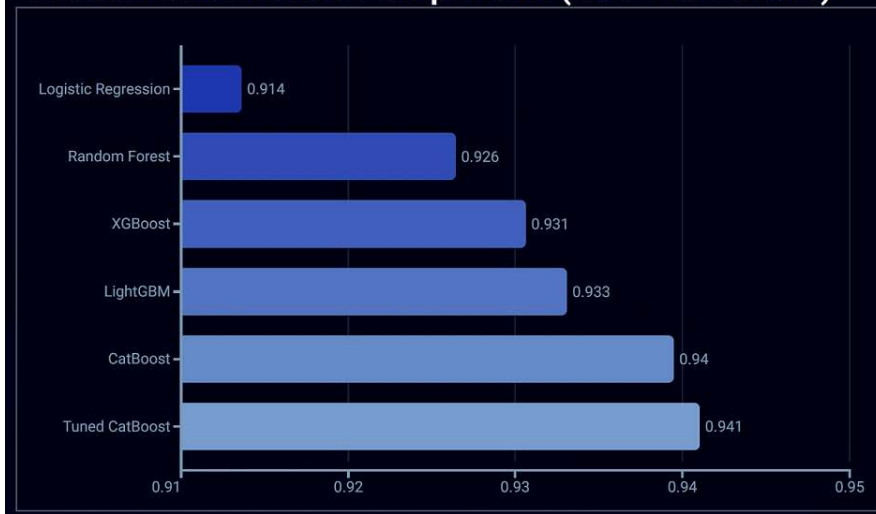
0.5: No discrimination; the model is no better than random chance

## Conclusion

---

The baseline models each performed well, with the CatBoost Classifier performing the best. It was then tuned and improved further, with the results shown below:

## Model Performance Comparison (ROC-AUC score)



The most significant contributors to the model were found to be:

1. Call duration (longer calls were more likely to sign)
2. Previous outcome success (a previously successful outcome was more likely to lead to another successful outcome)
3. Whether the customer had a housing loan (having a housing loan was negatively associated with a customer signing)
4. Contact type (unknown contact types were less likely to sign)
5. Number of contacts during per campaign (more contacts were more likely to sign)

These charts show that a previously successful outcome was more likely to lead to another successful outcome and that having a housing loan was negatively associated with a customer signing.

