

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

| Team Member's Name, Email and Contribution: | |
|--|-----------------------------|
| 1. Rishekh Dubey | 91221rishekhdubey@gmail.com |
| | |
| Please paste the GitHub Repo link. | |
| Github Link:- https://github.com/Rishekh/Email-Campaign/blob/main/project.ipynb Drive link:- https://drive.google.com/drive/folders/1h0jSzKp75V9uEMtNWq13Pq726fbAOM1u?usp=sharing | |
| Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words) | |

Problem Statement:

Email marketing is the act of sending a commercial message, typically to a group of people, using email. In its broadest sense, every email sent to a potential or current customer could be considered email marketing. It involves using email to send advertisements, request business, or solicit sales or donations.

Most of the small to medium business owners are making effective use of Gmail-based Email marketing Strategies for offline targeting of converting their prospective customers into leads so that they stay with them in Business.

Objective:

We are required to build a model to characterize the mail and track the mail that is ignored; read; acknowledged by the reader.

Data Set Used:

data_email_campaign.CSV

Scientific Computing tool & Data manipulation:

1. NumPy
2. Pandas
3. Sklearn
4. Warnings

Visualization libraries used:

1. Matplotlib
2. Seaborn

Contribution Role:

1. Rishekh Dubey

★ data_email_campaign.CSV

Cleaning & transforming each feature in data_email_campaign.

- ★ Null value treatment
- ★ Removing outliers
- ★ IQR based filtering
- ★ Customer's Locations
- ★ Email Campaign type for every location
- ★ Email Hotness Score
- ★ Email Status based on previous communication
- ★ Heatmap Correlation Matrix
- ★ Standardization
- ★ Train and Test Analysis
- ★ Random Forest Classifier
- ★ Hyperparameter tuning
- ★ Grid Search CV
- ★ Accuracy check on the train and test data set

- ★ Confusion Matrix
- ★ Roc_Auc Score
- ★ Feature importance
- ★ SHAP

Conclusion:

We have a dataset of Email send to every customer over the period.

We have reviewed the dataset and tried to build the model to predict the build the model to track the email if they are getting ignored, read or acknowledged by the reader. with the available dataset.

I have plotted multiple bar plots so that we can compare the data between different variables.

So in order to achieve our target to predict the stock closing prices, we have used random forest classifier.

After applying a random forest classifier and plotting the SHAP we can see that the negative value of shap shows the Emails that have been ignored.

The positive value of shap shows the variable which is responsible for reading and acknowledging the Email.

AUC value for training data- 0.7708421124964002

AUC value for test data- 0.7646248070389722

As we know higher the AUC, the better the model. It proves that our model is predicting values efficiently.

