Email Campaign Analysis

A Project Report

Submitted in the partial fulfillment of the requirements for

the award of the degree of

Bachelor of Technology

in

Department of Computer Science and Engineering

By

M. Sravani Chowdary (2010030104)

P. Mounika Reddy (2010030485)

K. Amrutha Varshini (2010030009)

under the supervision of

Mr.K. Sreenivasa Rao
Associate Professor, CSE

Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

March 2022

DECLARATION

The Project Report entitled "**Email Campaign Anaylsis**" is a record of bonafide work of, M. Sravani Chowdary (2010030104), Mounika Reddy (201003485),Amrutha Varshini(2010030009)submitted in partial fulfillment for the award of B.Tech in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/ University/ Institute.

M.Sravani Chowdary
2010030104

P. Mounika Reddy
2010030485

K. Amrutha Varshini
2010030009

CERTIFICATE

This is to certify that the Report entitled"**Email Campaign Anaylsis**" is being submitted by M. Sravani Chowdary (2010030104), Mounika Reddy (201003485),Amrutha Varshini(2010030009) submitted in partial fulfillment for the award of B. Tech in CSE to the K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/ University/Institute

Signature of the Supervisor
MR.K. Sreenivasa Rao
Associate Professor,
Department of CSE,
KLH University,
Hyderabad, India.

Signature of the HOD **Signature of the External Examiner**

ACKNOWLEDGEMENT

First and foremost, we thank the lord almighty for all his grace & mercy showered upon us, for completing this project successfully.

We take a grateful opportunity to thank our beloved Founder and Chairman who has given constant encouragement during our course and motivated us to do this project. We are grateful to our Principal **Dr. L. Koteswara Rao** who has been constantly bearing the torch for all the curricular activities undertaken by us.

We pay our grateful acknowledgment & sincere thanks to our Head of the Department **Dr. Chiranjeevi Manike** for her exemplary guidance, monitoring, and constant encouragement throughout the project. We thank **MR.K. Sreenivasa Rao** of our department has supported me throughout this Project holding the position of supervisor.

We wholeheartedly thank all the teaching and non-teaching staff of our department without whom we won't have made this project a reality.

ABSTRACT

Most small to medium business owners are using Gmail-based Email Marketing Strategies for offline targeting of converting their prospective customers into leads so that they stay with them in Business. Nowadays email marketing is considered a cost-effective marketing tool for different organizations to market their products or services. Despite the rise of the latest online communication tools such as social media, email marketing remains one of the most credible and profitable digital marketing tools.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Most 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. The main objective is to create a machine learning model to characterize the mail and track the mail that is ignored; read; acknowledged by the reader.

Email advertising is the act of sending promotional emails to customers in mass quantities. It commonly is to generate income or leads and it can include advertising. Most importantly, email marketing allows businesses to build relationships with leads, new customers, and past customers. It's a way to communicate directly to the customers in their inbox, at a time that is convenient for them. Email is one of the most important marketing channels with the right messaging tone and strategies.

Email campaign effectiveness is a way of analyzing the kind of email campaigns being run by businesses to carry out their marketing and promotional agendas and hence they need to know how well the campaign is working.

The work here characterizes and predicts the emails they are going to be ignored; read; acknowledged based on the various features related to the emails in the dataset and makes recommendations to lower the number of ignored emails.

CHAPTER 2

LITERATURE SURVEY

Literature Survey -01

TITLE: Automating The Collection And Analysis Of Campaign Emails

This Paper Was Published by Independent Work Report Fall in 2018. The Author Of this paper is Maia Hamin. In this Paper the Analysis was done automatically based on the collected or gathered data. The techniques and dataset used are crawler. The Crawler was a technique were crawler's integration with a full data pipeline also offers increased performance with decreased effort. But There was a consequence in this technique that the proposed model integration with MySQL database is time consuming. If there was a small data, then it is suitable but when it comes to large data it is difficult to integrate and it is time consuming. The Paper concludes that this model is suitable for only small data not for the large data.

Literature Survey -02

Title: Email Marketing: A Paradigm Shift To Marketing

This paper was published by International Journal Of Advanced Research in Management and Social Sciences in the year of 2020.The Author Of this paper is prof k. Venugopal, Dr. Vishnu Murthy. Three Datasets are used first is for analysing the cost ,second for analysing the Risk and last for analysing the data. This paper says that it is an effective way to reach a target audience around the world. It has few consequences like email marketing is to issue emails that are likely to be constructed to be spam.

Literature Survey -03

TITLE: Email Marketing Campaign For Selected Company

This paper was published by Tomas Bata University In Zlin. The Author for this paper is Bc. Natalia Vataman. Dataset used there was a company called purcari company dataset. The techniques used are Smith and Taylor offered the simplest Framework. The main aim of this paper was to analyse the company with the name called purcari company and mainly it depicts that it improves communication on with its customers and increase brand awareness and maintains a current customer base. The cost is high, and it is fixed.

CHAPTER 3

PROBLEM STATEMENT

Most 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. The main objective is to create a machine learning model to characterize the mail and track the mail that is ignored; read; acknowledged by the reader.Analyzing the data to track the emails that are been ignored, read, and acknowledged by the reader

CHAPTER 4

OBJECTIVES

The Aim of our project is to analyse the Analyzing data to track the emails that are ignored, read, and acknowledged by the reader. To achieve this we come up with two different ways 1. dashboard creation helps us to analyse the Emails that are being ignored or read by the reader through ML prediction the model is predicting the mail that is being ignored, read, or acknowledged by the reader

CHAPTER 5

HARDWARE AND SOFTWARE REQUIREMENTS

5.1 Hardware requirements:

Hardware requirements for insurance on the internet will be the same for both parties which are as follows:

Processor: Dual Core

RAM: 2 GB

Hard Disk: 320 GB

NIC: For each party

5.2 Software requirements Operating System:

Windows10 Ultimate which supports networking.

Python development toolkit.

Power Bi

Python tool kit 3.7.1 and 3.8.

CHAPTER 6

METHODOLOGY

6.1 XG Boost

Extreme Gradient Boosting is abbreviated as XgBoost. It is a classification and regression technique that uses the Boosting Ensemble Learning Technique. Boosting is an ensemble modeling strategy that aims to create a strong classifier from a collection of weak ones. It's done by stringing together weak models to create a model. To begin, a model is created using the training data. The second model is then created, which attempts to correct the faults in the previous model. This approach is repeated until either the entire training data set is properly predicted, or the maximum number of models has been added. Here is a representation of Boosting.

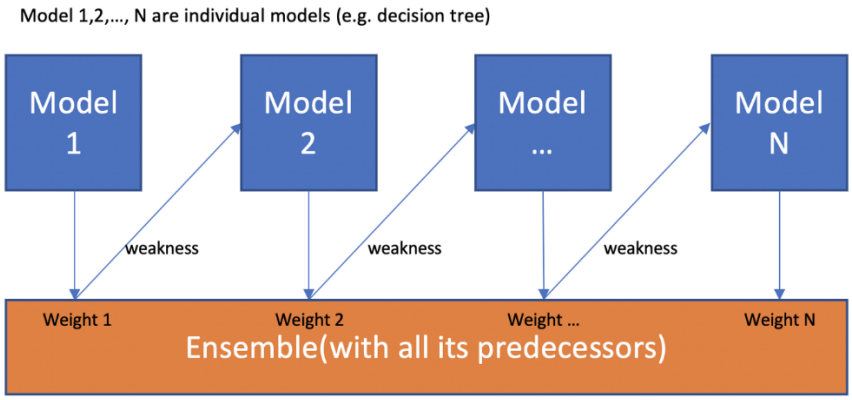
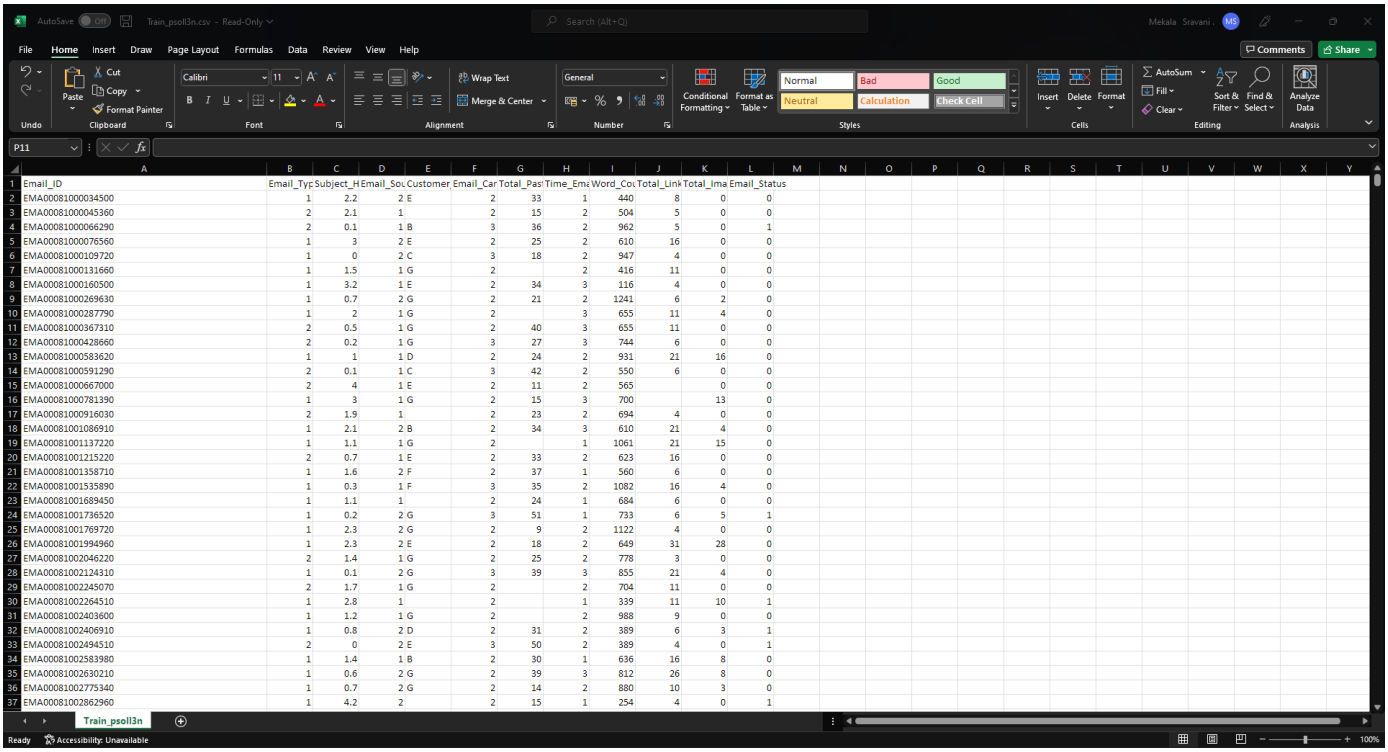


Fig: 6.1 XG BOOST Algorithm

In XGBoost, weights are very significant. All the independent variables are given weights, which are subsequently fed into the decision tree, which predicts outcomes. The weight of factors that the tree predicted incorrectly is increased, and these variables are fed into the second decision tree. Individual classifiers/predictors are then combined to form a more powerful and precise model. It can be used to solve problems including regression, classification, ranking, and user-defined prediction.

6.2 DATASETS



Email_ID	Email_Type	Subject_Hotness_Score	Email_Source	Customer_Location	Email_Campaign_Type	Total_Past_Communications	Time_Email_Sent_Category	Word_Count	Total_Links	Total_Images	Email_Status
EMA00081000034500	1	2.2	2	E	2	33	1	440	8	0	0
EMA00081000045360	2	2.1	1		2	15	2	504	5	0	0
EMA00081000066290	2	0.1	1	B	3	36	2	962	5	0	1
EMA00081000076560	1	3	2	E	2	25	2	610	16	0	0
EMA00081000109720	1	0	2	C	3	18	2	947	4	0	0
EMA00081000131660	1	1.5	1	G	2		2	416	11	0	0
EMA00081000160500	1	3.2	1	E	2	34	3	116	4	0	0
EMA00081000269630	1	0.7	2	G	2	21	2	1241	6	2	0
EMA00081000287790	1	2	1	G	2		3	655	11	4	0
EMA00081000367310	2	0.5	1	G	2	40	3	655	11	0	0
EMA00081000428660	2	0.2	1	G	3	27	3	744	6	0	0
EMA00081000583620	1	1	1	D	2	24	2	931	21	16	0
EMA00081000591290	2	0.1	1	C	3	42	2	550	6	0	0
EMA00081000667000	2	4	1	E	2	11	2	565		0	0
EMA00081000781390	1	3	1	G	2	15	3	700		13	0
EMA00081000916030	2	1.9	1		2	23	2	694	4	0	0
EMA00081001086910	1	2.1	2	B	2	34	3	610	21	4	0
EMA00081001137220	1	1.1	1	G	2		1	1061	21	15	0
EMA00081001215220	2	0.7	1	E	2	33	2	623	16	0	0
EMA00081001358710	1	1.6	2	F	2	37	1	560	6	0	0
EMA00081001535890	1	0.3	1	F	3	35	2	1082	16	4	0
EMA00081001689450	1	1.1	1		2	24	1	684	6	0	0
EMA00081001736520	1	0.2	2	G	3	51	1	733	6	5	1
EMA00081001769720	1	2.3	2	G	2	9	2	1122	4	0	0
EMA00081001994960	1	2.3	2	E	2	18	2	649	31	28	0
EMA00081002046220	2	1.4	1	G	2	25	2	778	3	0	0
EMA00081002124310	1	0.1	2	G	3	39	3	855	21	4	0
EMA00081002245070	2	1.7	1	G	2		2	704	11	0	0
EMA00081002264510	1	2.8	1		2		1	339	11	10	1
EMA00081002403600	1	1.2	1	G	2		2	988	9	0	0
EMA00081002406910	1	0.8	2	D	2	31	2	389	6	3	1
EMA00081002494510	2	0	2	E	3	50	2	389	4	0	1
EMA00081002583980	1	1.4	1	B	2	30	1	636	16	8	0
EMA00081002630210	1	0.6	2	G	2	39	3	812	26	8	0
EMA00081002775340	1	0.7	2	G	2	14	2	880	10	3	0
EMA00081002862960	1	4.2	2		2	15	1	254	4	0	1

Fig: 6.2 Datasets

Our email campaign dataset has 68353 observations and 12 features. Clearly, Email_Status is our target variable.

Our features:

- **Email Id** - It contains the email id's of the customers/individuals
- **Email Type** - There are two categories 1 and 2. We can think of them as marketing emails or important updates, notices like emails regarding the business.
- **Subject Hotness Score** - It is the email's subject's score based on how good and effective the content is.
- **Email Source** - It represents the source of the email like sales and marketing, or important admin mails related to the product.
- **Email Campaign Type** - The campaign type of the email.

- **Total Past Communications** - This column contains the total previous mails from the same source, the number of communications had.
- **Customer Location** - Contains demographical data of the customer, the location where the customer resides.
- **Time Email sent Category** - It has three categories 1,2 and 3; the time of the day when the email was sent, we can think of it as morning, evening and night time slots.
- **Word Count** - The number of words contained in the email.
- **Total links** - Number of links in the email.
- **Total Images** - Number of images in the email.
- **Email Status** - Our target variable which contains whether the mail was ignored, read, acknowledged by the reader.

6.3 FLOW CHART

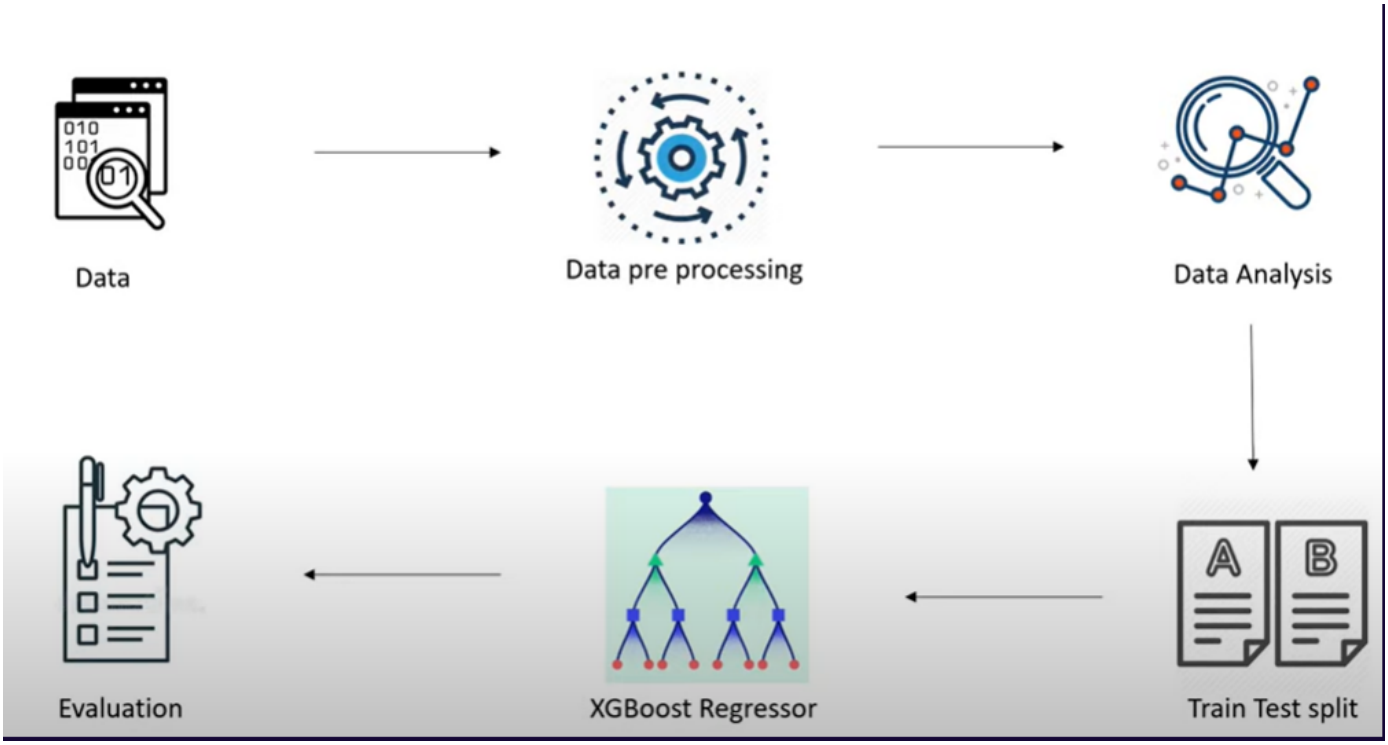


Fig: 6.3 Flow Chart.

6.4 WORK PROGRESS

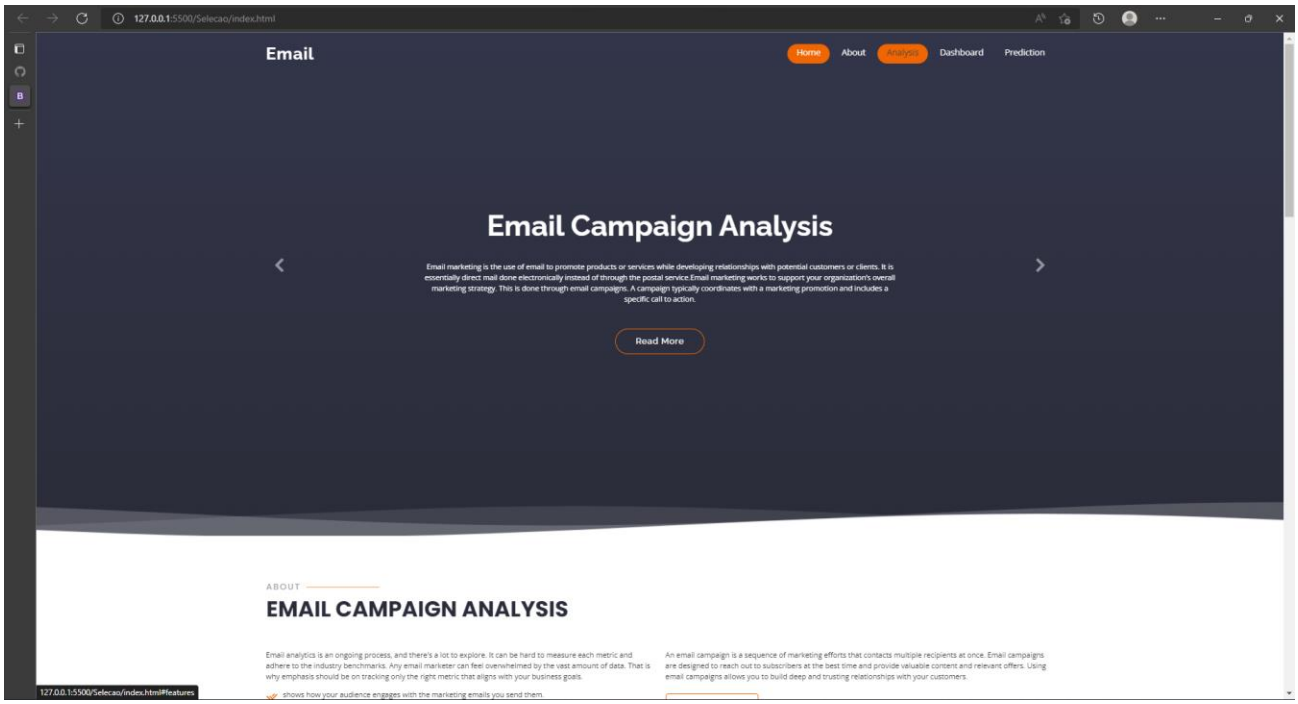


Fig: 6.4 Output Screenshots

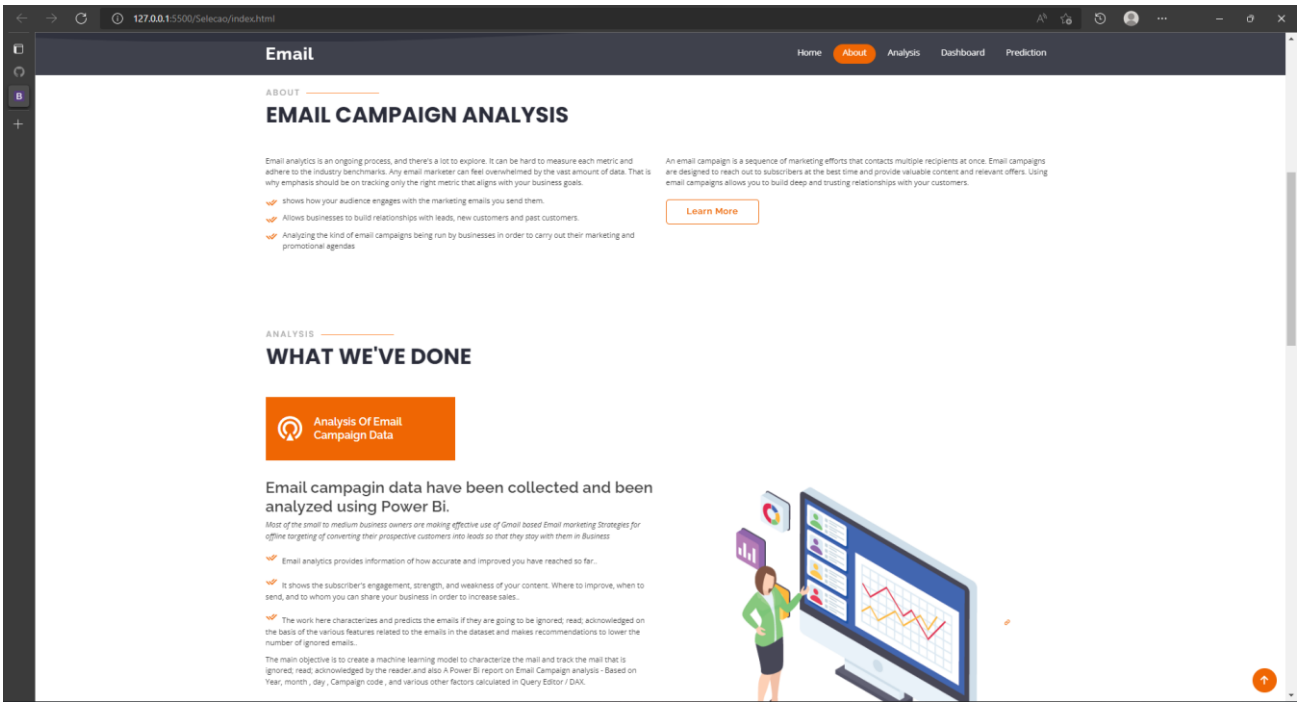


Fig: 6.4.1 Main screen Screenshots

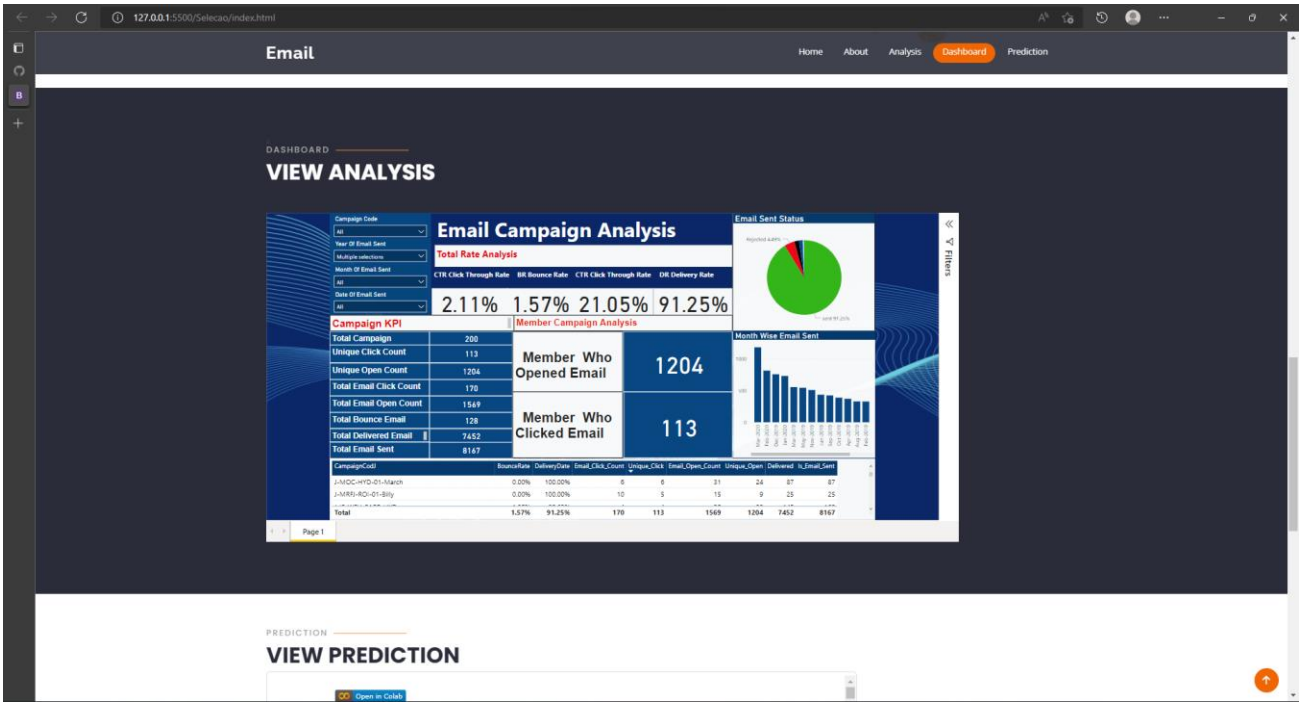


Fig: 6.4.2 Dashboard Screenshot

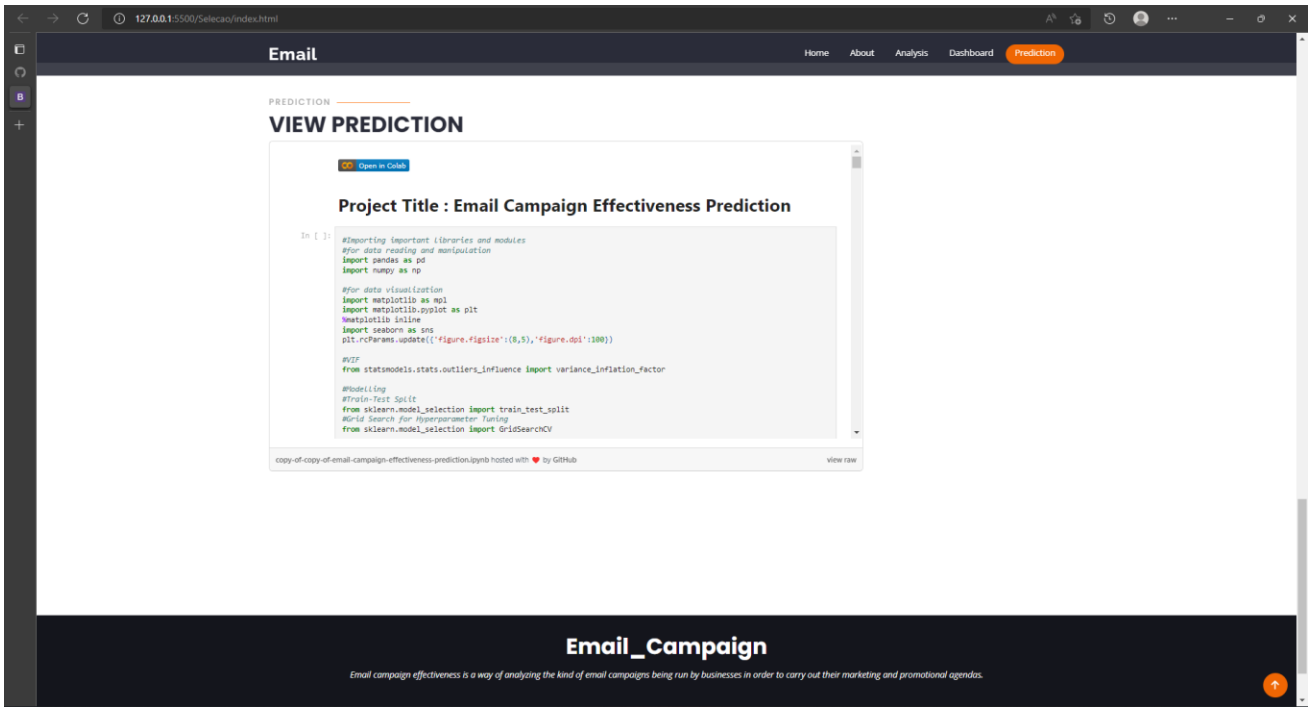


Fig: 6.4.3 Prediction Screenshots

CHAPTER 7

RESULTS

Imbalanced Class Handling techniques such as Under sampling and SMOTE were done after the train-test split only on the training data, to make sure that the model doesn't catch up to the test set at all and it remains unknown which somewhat reduced our results.

It is observed that SMOTE worked considerably better than Random Under sampling, which may have led to the loss of information.

The decision Tree Model is overfitting. It is working great on train data and worse on test data.

XGBoost Algorithm worked in the best way possible with such imbalanced data with outliers with an F1 Score of 0.77 on the test set.

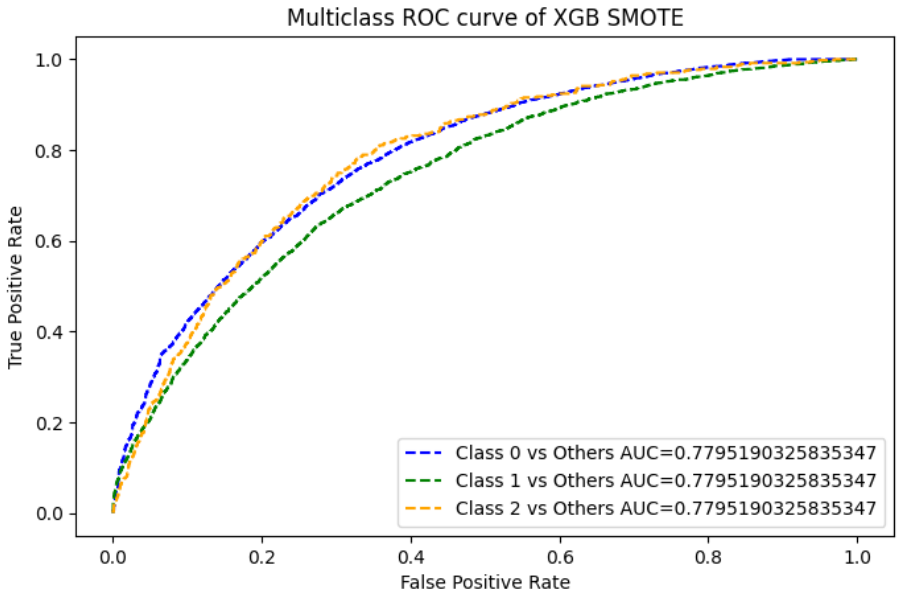


Fig: 7.1 Results Screenshot

CHAPTER 8

CONCLUSION

An email campaign is a sequence of marketing efforts that contacts multiple recipients at once. Email campaigns are designed to reach out to subscribers at the best time and provide valuable content and relevant offers. Using email campaigns allows you to build deep and trusting relationships with your customers. The main objective is to create an interactive dashboard to improve email marketing strategy skills and characterize the mail and track the mail that is ignored; read; or acknowledged by the reader.

CHAPTER 9

FUTURE ENHANCEMENT

we can add an extra feature like which email is doing better and which are not and provide them the reasons why it is been visited low and provide them feedback.

CHAPTER 10

REFERENCES

[1] Dinh, Son, Taher Azeb, Francis Fortin, Djedjiga Mouheb, and Mourad Debbabi. "Spam campaign detection, analysis, and investigation." *Digital Investigation* 12 (2015): S12-S21.

[2] Howes, Calvin, Pablo Saide, Sharon Burton, Hugh Coe, Amie Dobracki, Marta Fenn, Richard Ferrare et al. "Multi-campaign analysis of smoke properties and cloud interactions in the Southeast Atlantic using ORACLES, LASIC, and CLARIFY data with WRF-CAM5." (2021).

[3] Hamin, Maia. "" don’t ignore this:" Automating the Collection and Analysis of Campaign Emails." (2018).

[4] Venugopal, K., Saumendra Das, and D. Vishnu MURTHY. "Email Marketing: A Paradigm Shift to Marketing." International Journal of Advanced Research in Management and Social Sciences, S 1, no. 1 (2012): 101-116.

[5] Vataman, Natalia. "Project Of Email Marketing Campaign for Selected Company." (2016).