jML FOR ADVERTISING IN DIFFERENT ZONES TO TARGET CUSTOMERS

Group Members:

1. Aman Khandelwal(U10116FCS007)
2. Chinju Mary George(U101116FCS025)
3. Gaurav Mundhra(U101116FCS037)
4. Shashwat M Shah(U101116FCS112)
5. Sherry Sharma(U101116FCS287)
6. Abhijit Singh(U101116FCS002)

PROJECT DESCRIPTION:

Problem:

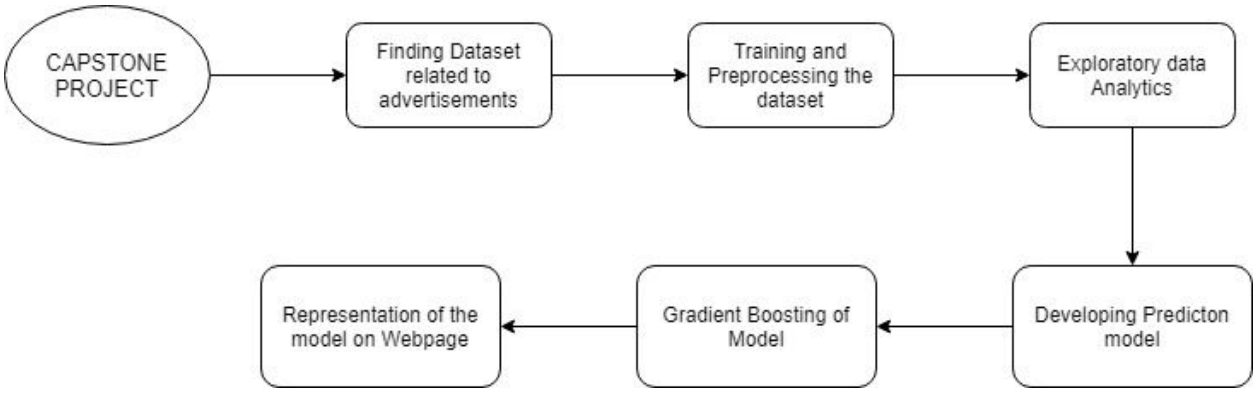
Using Machine learning to target customers of different zone with the help of advertisements.

Solution:

- A model to make predictions for the probabilities of a person clicking on the ad(or not) using click through rate probability to analyze the dataset based on some significant features.
- These significant features were
    - Weekday
    - Hour in a day
    - Banner position
    - Site category
    - Device type
    - App category

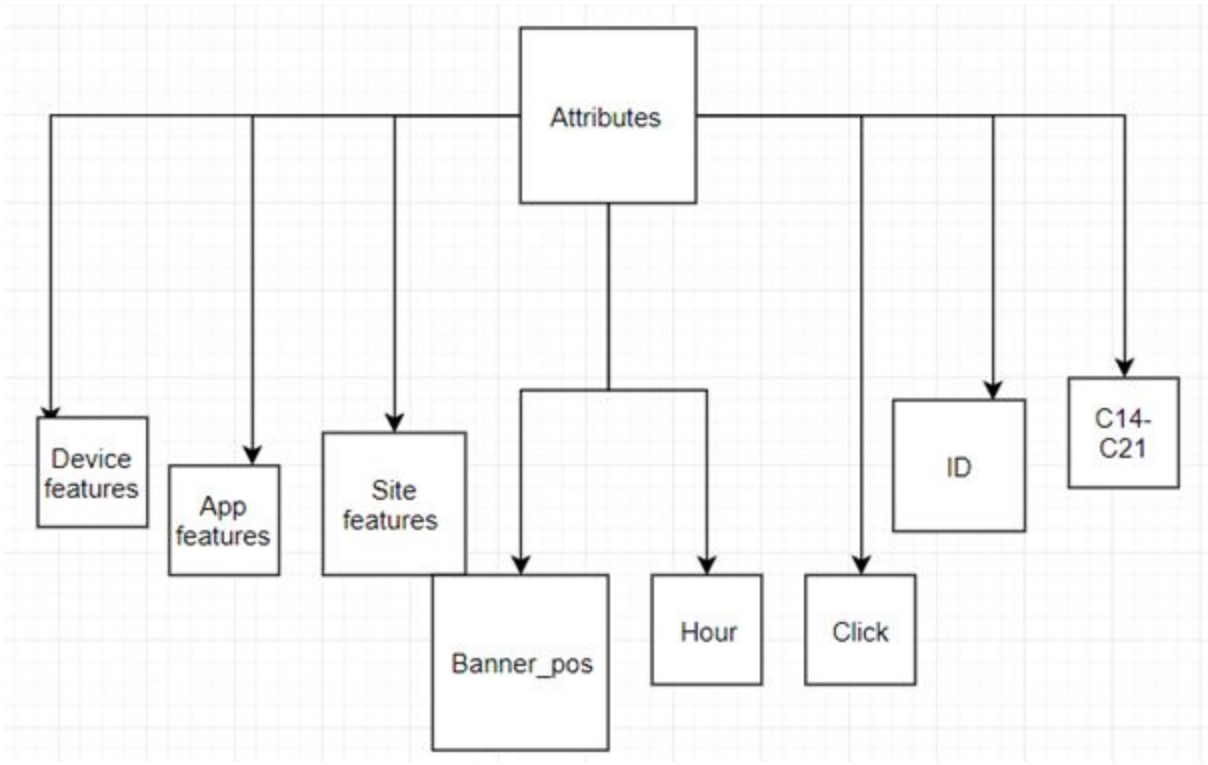
The machine learning algorithm we used is Logistic Regression.

WORKFLOW OF OUR PROJECT



All Achieved and Working Functionalities :

We collected advertisment dataset of the business organisation. Analyze the whole dataset with the help of python libraries(Numpy and Pandas). We analyze some basic basic features like device_features, app_features, site_feature, banner-pos, hour, no_of_clicks, id and try to find out some relationships among them through graphical representation.



After analysis of dataset we develop the prediction model as follows:
1. **Data Preparation Stage** : In this stage we take the 10% random data from the dataset to reduce computation and complexity.
2. **One hot encoding** : In this we provide one hot encoding feature on site-category, app_category, banner_pos for providing assistance to our model and for reading the data easily.
3.**Feature Selection** : To reduce the dimensional space occupied and to deal with overfitting,cross validation and regularization is used to obtain a trade-off between number of features and F1 score.

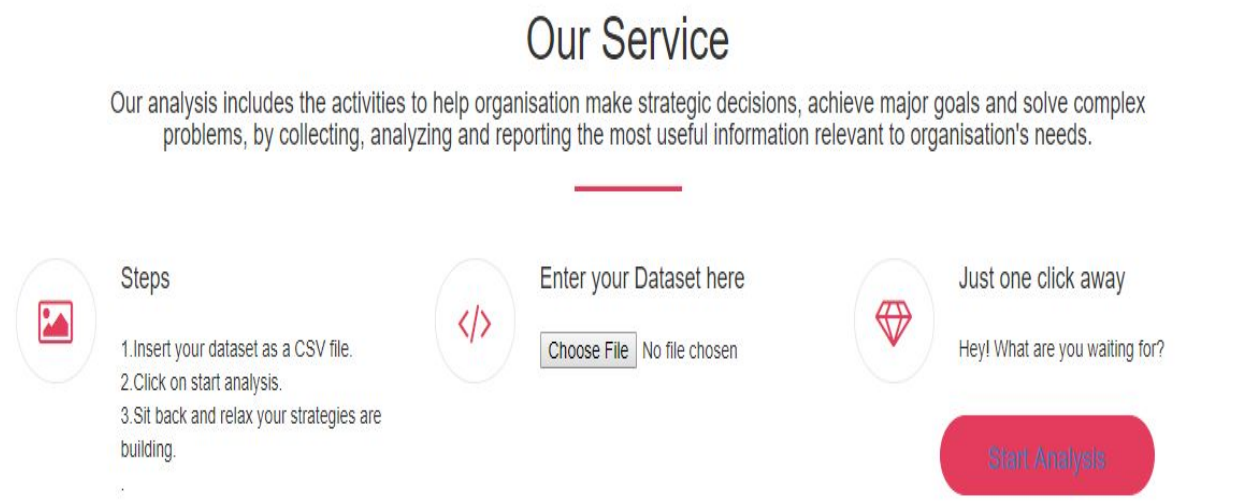
Model Training:
After prediction model stage we applied some machine learning algorithms. Our model is based on Logistic Regression and L1 Regularization with balanced weights.

Further we evaluate the metrics using Gradient Boosting approach. In which we use the xgboost module and classification reports from sci-kit learn to assess the accuracy of the models. The model correctly predicts the non-clicks cases 83% of the time on aggregate and it correctly predicts the clicks 68 times on 100. This varies as samples keep getting randomly

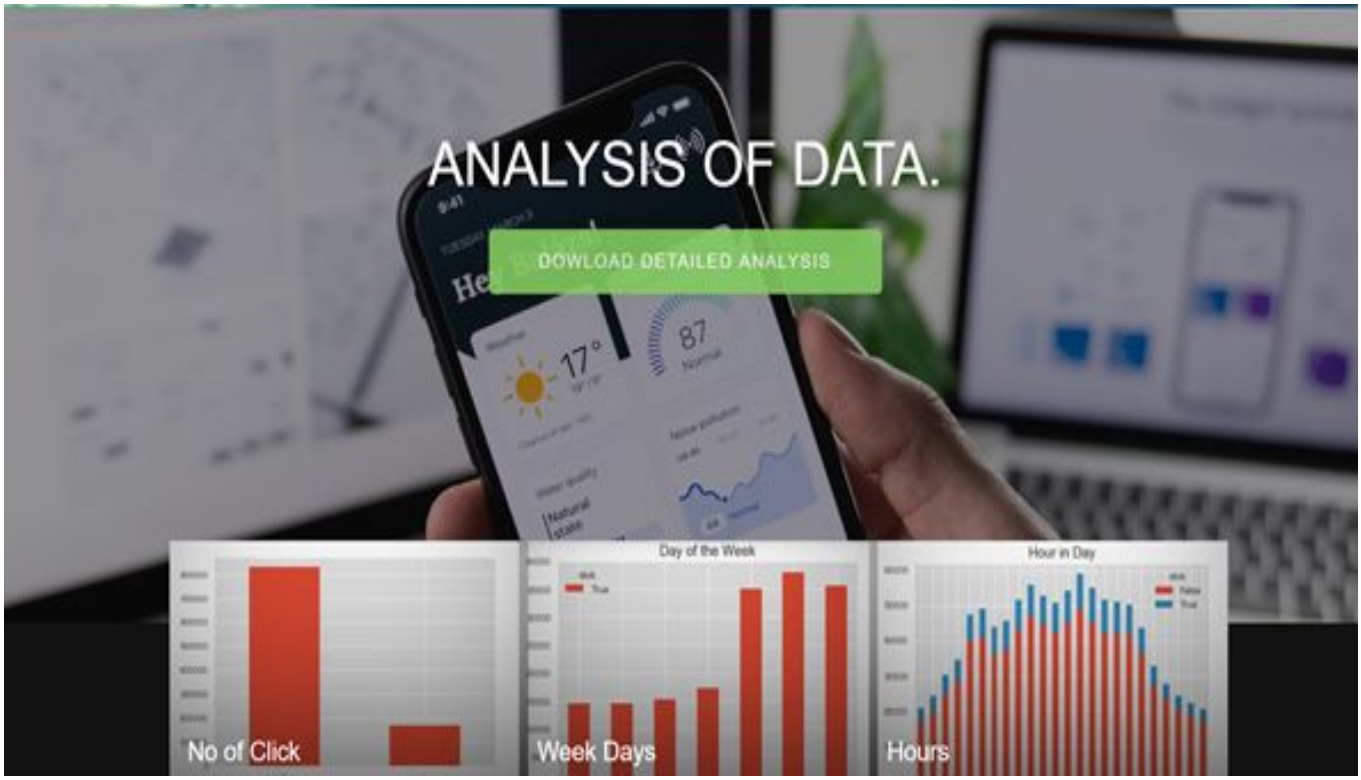
drawn as fractions. The model runs with 83% accuracy. The confusion matrix records the numbers of true positives/negatives and false.

	precision	recall	f1-score	support
False	0.83	1.00	0.91	248673
True	0.68	0.00	0.00	51328
avg / total	0.80	0.83	0.75	300001

We have created a web application in which the user have to submit its dataset(in the proposed order).Then our model work in the backend and analyze the data and displaying its output in the form of graphs in the next action page. We also provide user to download its complete analysis of the dataset in the word document.



CAPSTONE PROJECT REPORT



Title of the project
Description of the project
Name, Roll Number, Section, email and Phone No. of team members
Project Plan and Design
Roles and Responsibilities
Snapshots of project execution
Snapshots of Software development tools.
Learning in the project (including Tools and technologies)
Challenges during the project
Can your project be extended to Capstone 2? If yes, then mention what features can be added.

4. Further Perspective and Implementation

As far as now we have successfully implemented all the modules we planned to complete which is -

1.Exploratory Data Analysis

2.Model development

3.Evaluating the results

As the results are appropriate and end user will have enough evaluated results to have a strategy for their advertisement campaign.

We don’t think there is much more things to explore particularly for this project as the implementation totally depends on type of data you are processing.

CAPSTONE PROJECT REPORT