**A COMPARATIVE ANALYSIS BETWEEN ZOMATO AND SWIGGY**

BY

PULAK JAIN- 20BRS1126

S ABENAV RAM - 20BRS1137

JEEVAN PRAKASH H A- 20BRS1259

A project report submitted to

**Dr. ASNATH VICTY PHAMILA Y**

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

in partial fulfilment of the requirements for the course of

**CSE3505 FOUNDATIONS OF DATA ANALYTICS**

**B.TECH COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE AND ROBOTICS**



**Vandalur – Kelambakkam Road Chennai – 600127**

# ABSTRACT

The project consists of analyzing and comparing data from Zomato and Swiggy to help customers choose the best app for ordering food. Based on this, we will compare different elements of the both the companies and let the customer choose the best in different scenarios. The data must first be cleaned, certain values must be imputed (if necessary), and the data must be pre-processed in order to have better comparability. This will assist the organization in analyzing the variables influencing the value of its business. The dataset(s) under consideration for the sandbox contain a wide range of features, such as cuisine type, table reservations, average cost for two people, well-known culinary items, and much more. All of these characteristics can assist in predicting a better visualization, assisting customers in selecting the pricing, food quality, and culinary style.

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **SERIAL**  **NO.** |  | **TITLE** | **PAGE**  **NO.** |
|  |  | ABSTRACT | 3 |
|  |  |  |  |
| 1 |  | INTRODUCTION | 6 |
|  | 1.1 | OBJECTIVES AND GOALS |  |
|  | 1.2 | APPLICATIONS |  |
|  | 1.3 | FEATURES |  |
| 2 | 2.1 | METHODOLOGY | 7 |
| 3 | 3.1 | SOFTWARE –CODING AND ANALYSIS | 8 |
|  |  | (SNAPSHOTS OF CODING AND RESULTS) |  |
| 4 |  | CONCLUSION AND FUTURE WORK | 23 |
|  | 4.1 | RESULT, CONCLUSION  AND INFERENCE | 22 |
|  | 4.2 | FUTURE WORK |  |

5 REFERENCES 24

# 1. INTRODUCTION

* 1. **OBJECTIVES AND GOALS**
     + Create a data analytics tool which will help in analyzing the food delivering platforms.
     + Provide insight on the food delivering platforms.
     + Being able to provide the best app for the particular situations.
     + To able to suggest restaurants based on the user’s order amount and cuisine choice.
     + To reduce the time for user to order the food.
     + Providing relevant information to the questions by the user .

# APPLICATIONS

* Choosing the food delivering app
* Choosing the restaurant

# FEATURES

* Plots to understand the difference between each app.
* Statistics to differentiate between restaurants
* Ratings of the cuisines in various restaurants.
* Suggest restaurants based on their delivering time

# 2. LITREATURE SURVEY

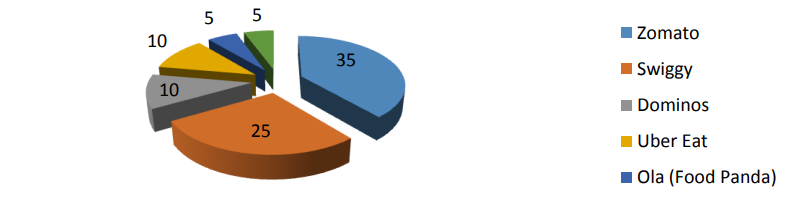
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Title | Journal/year of publication | Data set used | Methodologies used | Metrics used | Interpretation of Results |
| 1. | Sentiment analysis of Customer’s comments using Random Forest Classifier model. | 7th International Scholar’s Conference (ISC7), 2019 | Zomato Data Set taken from Kaggle. The dataset lists all the restaurants in Bangalore and the ratings, votes and other data about the restaurants. | Random Forest | Accuracy  Positive 92% Negative 93%  Neutral 93% | Proposed the system of finding the overall average customer’s view about a restaurant on the basis of the prediction. It can be interpreted that the Random Forest algorithm used in the paper is quite accurate in predicting the overall sentiment towards a customer. |
| 2. | Restaurants Rating Prediction using Machine Learning Algorithms | International Journal of Computer Applications Technology and Research Volume 8, 2019 | Zomato Data Set taken from Kaggle. The dataset lists all the restaurants in Bangalore and the ratings, votes and other data about the restaurants. | Decision Tree  Random Forest  Linear-Regression  KNN algorithm | Accuracy  Linear Regression 30%  KNN 44% Decision Tree 69%  Random Forest 81% | In the Proposed paper, they predict the aggregate rating on the basis of overall textual rating as well as other metrics such as price, votes and etc.  The Random Forest model provides the best accuracy almost double than linear regression and KNN, thus highlighting the application and working of the Random Forest. |
| 3. | Location based analysis on Zomato Dataset | SSRN, 2021 | Zomato Data Set taken from Kaggle. The dataset lists all the restaurants in Bangalore and the ratings, votes and other data about the restaurants. | EDA on location-based data in the city of Bangalore. | Data Visualizations | It highlights all the locations linked with certain cuisines and also describes whether a new restaurant will survive or not in that area. |
| 4. | Food and Restaurant recommendation system using hybrid filtering | North American Academic Research (NAAR) | U.S. govt. recognized dataset on food and restaurants. | Content-Based and Collaborative based filtering system using AI, KNN, linear regression and decision trees | Accuracy  Linear regression 74%  KNN 70%  Decision trees 78%  Random Forest 85% | Using Collaborative and Content based filtering with the help of ML algorithms, the model was achieving 83% accuracy which is pretty high as for actual food recommendation system thus proving the methodology quite effective. |
| 5. | Prediction of best Cuisines in a Country Capital using ML algorithms | NCI | Zomato Bangalore Dataset | Linear Regression, Decision Tree, Random Forest | Accuracy  Linear Regression 98%  Random Forest 98%  Decision Tree 97% | The Model does perform well to list out the best cuisines by the Country, the best model came out to be random forest to predict the score for the likability of the cuisine in the country. |
| 6. | Zomato Restaurants Data Analysis Using Machine Learning Algorithms | 2021 JETIR February 2021, Volume 8, Issue 2 | Zomato Bangalore Dataset | Linear Regression, Random Forest, Decision Tree | Accuracy  Linear Regression 24%  Random Forest 87%  Decision Tree 85% | In this paper, the authors are predicting the aggregate rating on the basis of the other columns/variables. The random forest and decision tree are quite good with almost 85%+ accuracy while the linear regression has low accuracy. |

# 3. METHODOLOGY

1. Discovery

At the beginning we understood the current food delivering apps market in India. After going through many reports and statistics we understood the current prominent food delivering apps in India are Swiggy and Zomato. Later, we understood about both companies and the way each platform works and how long they were working in India.

Below is the figure of the Market Distribution of Food Delivering Companies in India



2. Data Preparation

In the next step, we started collecting data from various resources such as online websites such as Kaggle, Github to find relevant resources. In Kaggle we were able to the required data which can be used for our project which had data about swiggy and Zomato both in India and in the city of Bangalore. The datasets which we had chosen had some missing values and were also suitable for our project. So, we had to prepare the data for our project.

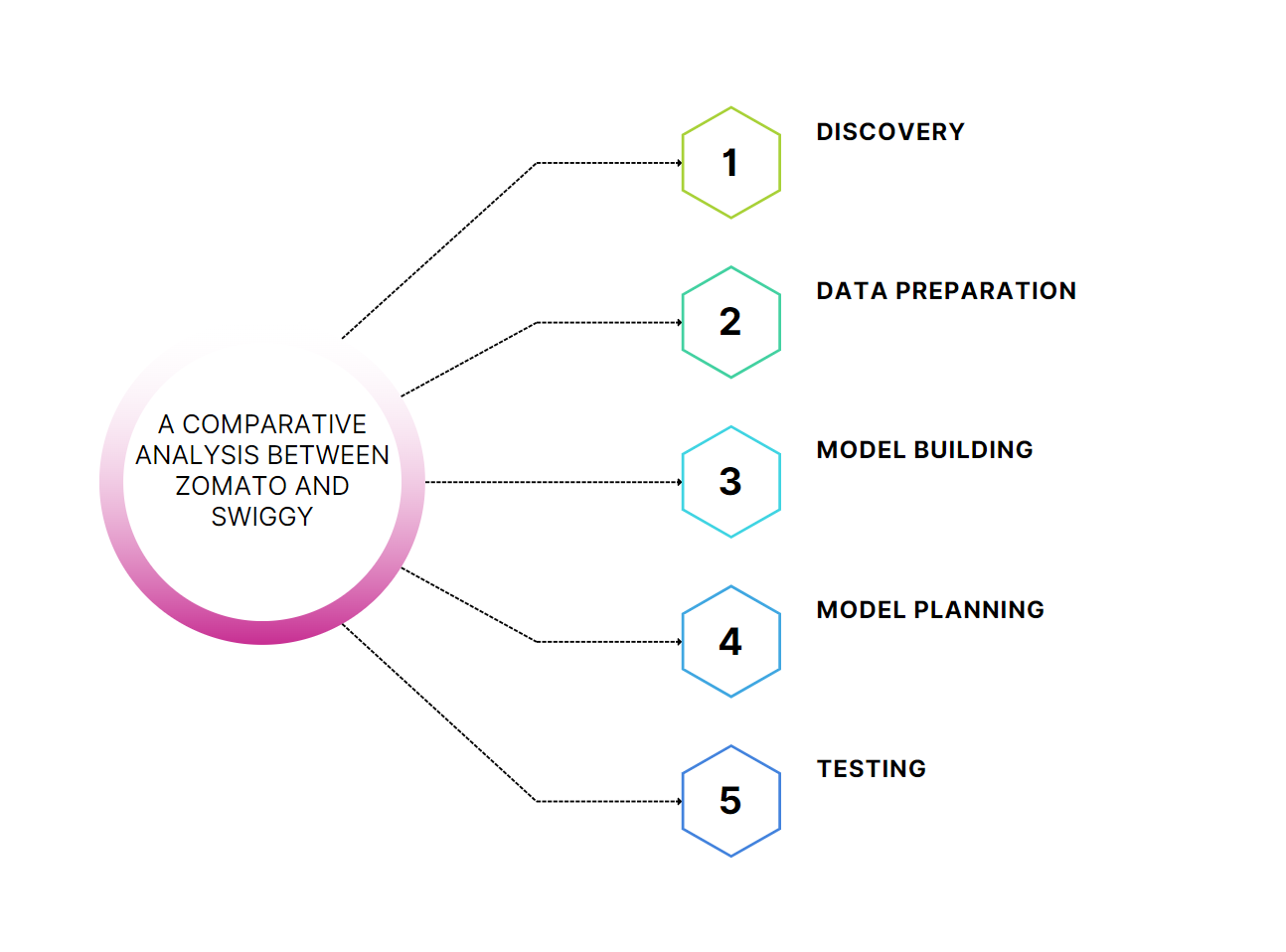
3. Model Planning

Next, we started planning model to predict various outcomes based on the user preferences such as their preferable cuisine, preferable restaurant, preferable delivery time or combination of various factors such as dish, type of food such as vegetarian food and non-vegetarian food, delivery time, restaurant. So, we decided to use algorithms such as Random Forest, Pattern Matching for finding the patterns between the data, Support Vector Machine (SVM) for classification and so on.

4. Model Building

After planning the model, we started building the model with chosen algorithms to predict, classify or find patterns by using the chosen algorithms and we built the platform to help the user.

**Keywords:** Random Forest, SVM, Pattern matching



**ARCHITECTURE**

# 4. MODULES

# There are three modules in our project:

# 1. Data

# 2. Model

# 3. Visualization

# 1. Data

# In this module, we have the data that we are using for this project which contains data of Swiggy and Zomato that we have preprocessed using various techniques such as imputation, data cleaning etc.

# 

# Image Source https://www.analyticsvidhya.com

# 2. Model

# In this module, we have the model which is used to predict the better app in various scenarios given by the user using various algorithms including K-means, Pattern Matching, SVM, Logistic Regression etc.

# 3. Visualization

# This module contains various graphs used to comparison the statistics between both the food delivering apps.

# 5. DATASET

# Zomato Dataset

The collected data has been stored in a csv file . Each restaurant in the dataset is uniquely identified by its Restaurant Id. Every Restaurant contains the following variables:

• Restaurant Id: Unique id of every restaurant across various cities of the world

• Restaurant Name: Name of the restaurant

• Country Code: Country in which restaurant is located

• City: City in which restaurant is located

• Address: Address of the restaurant

• Locality: Location in the city

• Locality Verbose: Detailed description of the locality

• Longitude: Longitude coordinate of the restaurant's location

• Latitude: Latitude coordinate of the restaurant's location

• Cuisines: Cuisines offered by the restaurant

• Average Cost for two: Cost for two people in different currencies 👫

• Currency: Currency of the country

• Has Table booking: yes/no

• Has Online delivery: yes/ no

• Is delivering: yes/ no

• Switch to order menu: yes/no

• Price range: range of price of food

• Aggregate Rating: Average rating out of 5

• Rating color: depending upon the average rating color

• Rating text: text based on rating of rating

• Votes: Number of ratings casted by people.

Swiggy Dataset

This dataset contains swiggy registered restaurants details of major metropolitan cities of India. In this dataset it contains the swiggy past orders and restaurants in cities such as

Mumbai

Delhi

Kolkata

Chennai

Bangalore

Hyderabad

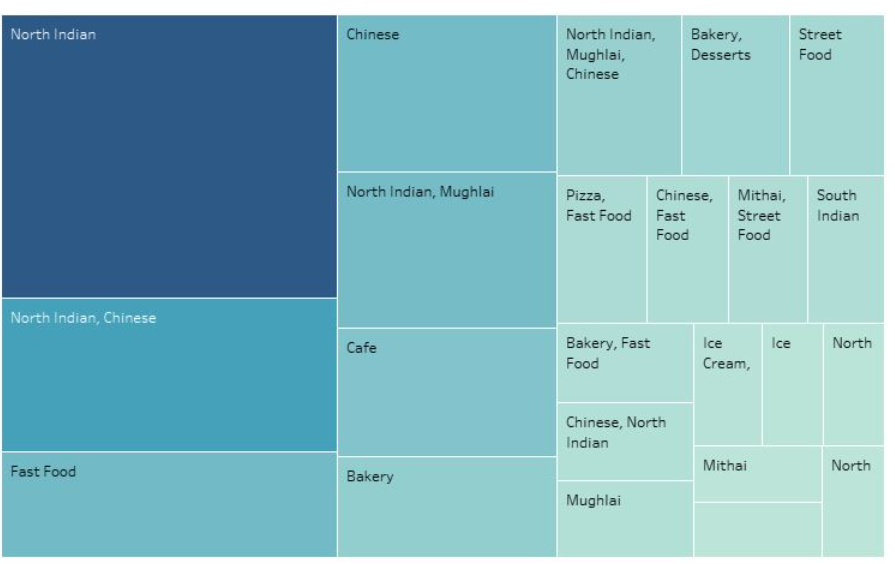
Ahmedabad

Pune

Surat

The Swiggy dataset contains variables such as:

* Restaurant Name
* Category
* Rating
* Cost for two Veg
* city
* Area
* Locality
* Address
* Long Distance Delivery.



# 6. ALGORITHM

# Support Vector Machine Algorithm (SVM): It is a supervised algorithm which is used to classify the values in the Swiggy and Zomato dataset. This is used to classify various parameters such as type of cuisine, type of food, restaurants based on their ratings.

# Logistic Regression: Logistic regression is a supervised algorithm learning. It is used to calculate or predict the probability. This was used to predict the restaurants that user would prefer based on their preferences etc. This was also used to predict the delivery time and ratings.

# K Means: K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters from the data available from the Swiggy and Zomato datasets. This was used to cluster the restaurants which have similar ratings, similar cuisine, similar location, or similar delivery time etc. which was helpful for the user to select the restaurant.

# 

# Clustering