**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email, and Contribution:** |
| Contributor Roles :  1. Jayanth V (Email – [arambamvj@gmail.com](mailto:arambamvj@gmail.com)) :   * 1. Data Overview.   2) Null values treatment.  3) Mapping string columns into numeric features  4) Functions for Text Preprocessing.  5) Exploratory Data Analysis.  6) Data Visualization.  7) Preprocessing the dataset for clustering models (Using TF -IDF Vectorizer)  8) Project structuring.        2. **Gowthaam Kumarasamy (Email -** [**gowthaam02@gmail.com**](file:///C:\Users\Admin\Downloads\gowthaam02@gmail.com)**) :**   1. Implementing Elbow Method for finding the K value. 2. K- Means Clustering. 3. Hierarchical Clustering. 4. Agglomerative Clustering. 5. DBSCAN Clusteirng. 6. VADER model and TEXT BLOB model. 7. Model - Comparison. 8. Technical Documentation. |
| **Please paste the GitHub Repo link.** |
| GitHub Link:- <https://github.com/jayV1999/ZOMATO-RESTAURANT-CLUSTERING-AND-SENTIMENT-ANALYSIS>  Google Drive Link:  <https://drive.google.com/drive/folders/11WTKlWwRT9QGCbdHGI_K2FYrCOSAH_u_?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)** |
| We are provided with two datasets, One is the metadata dataset and the other one is the Reviews dataset.  The Project focuses on Customers and companies, we have to analyze the sentiments of the reviews given by the customer in the data and make some useful conclusions in the form of Visualizations. Also, cluster the zomato restaurants into different segments. The data is visualized as it becomes easy to analyze data in an instant. The Analysis also solve some of the business cases that can directly help the customers find the Best restaurant in their locality and for the company to grow up and work in the fields they are currently lagging in.  At first, we performed Data cleaning by creating a data frame and removing unnecessary columns with null values, dropping the duplicate rows in this dataset, and started getting basic insight from the cleaned dataset.  We have performed some text preprocessing on some columns to make them efficient for us to implement our clustering models, And also performed visualization to get basic insights on the features.  We went with feature conversion in which the object datatype features are converted to numeric features in our dataset.  After feature selection, we tend to split the data into clusters and started implementing basic Machine Learning Models for clustering  The Models Were :  a) K – means clustering  b) Hierarchical clustering  c) Agglomerative clustering  d) DBSCAN clustering  After merging two datasets we tend to perform sentiment analysis on the clusters that we have separated using VADER MODEL and TEXT BLOB MODEL.  The main objective of sales forecasting is to paint an accurate picture of expected sales by targeting the clusters. Sales teams aim to either hit their expected target or exceed it. When the sales forecast is accurate, operations go smoothly and future planning for the company's growth is done efficiently. |