**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Gayatri Surapaneni  [gayatrisurapaneni08@gmail.com](mailto:gayatrisurapaneni08@gmail.com)  Zomato restaurant clustering and sentiment analysis ML Project was completely contributed by me. |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/Gayatri-Surapaneni/Zomato-restaurant-clustering-and-sentiment-analysis |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Abstract**:  India is well-known for its unique multi-food cuisine, which is offered in a huge number of restaurants and hotel resorts and symbolizes  unity in variety. In India, the restaurant industry is changing rapidly. More People are appealing to the concept of eating restaurant meals, whether they dine outside or have food delivered to their homes. The increasing number of restaurants in  every Indian state has encouraged an analysis of the information to gain some insights, noteworthy facts, and statistics about the Indian food sector. As a result, the purpose of this study is to analyze Zomato restaurant data in  Hyderabad. Zomato is a restaurant aggregator  and food delivery service based in India. With the use of unsupervised and supervised machine learning algorithms, the work here clusters restaurants into distinct segments and evaluates the sentiments in customer reviews. The analysis also resolves several business cases that can directly assist customers in locating the best restaurant in their area, as well as the company's growth and development in areas where it is  currently underperforming.  **Problem Statement:**   * The Project focuses on Customers and Company, you have t**o analyze the sentiments of the reviews given by the customer in the data and made some useful conclusion in the form of Visualizations**. Also, **cluster the zomato restaurants into different segments**. * This could help in clustering the restaurants into segments. Also the data has valuable information around cuisine and costing which can be used in cost vs. benefit analysis.   **Data Sets Given**  Zomato Restaurant names and Metadata   * + This dataset gives the information about the restaurants, cuisines, collections available   Zomato Restaurant reviews   * + This Dataset gives the information of the reviewers and followers of the every restaurants which are available   **Methodologies used:**   * 1. Loading the data   2. Exploratory Data Analysis   3. Treating missing values and outliers   4. Natural Language Processing   5. Clustering   6. Sentimental analysis   7. Building models   8. Hyper parameter Tuning.   **Challenges faced:**   * In the metadata(for clustering), we had only 100 rows and 4 variables to learn. After building the models, we found the silhouette score different number of clusters. We mainly focused on the silhouette score to evaluate the models. We were able to secure only around 0.6 silhouette score from all different models whose optimal number of clusters were to be 6 or 7. * In the reviews data(for sentiment analysis), to find the feature for the analysis was a tedious task. We made the split for rating and created another feature which is further used for sentimental analysis. But while creating this another feature, we first took 3 partitions for rating (average, good and best). But we did not get good result from the feature with 3 splits. Later another feature was created with 2 splits from rating(1 and 0) * Overall, we succeeded with good Silhouette score of 0.6 with 6 optimal clusters in clustering. And in Sentiment Analysis, we got the train and test accuracy as good as 0.9 and 0.84. |