**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1)Name : Vikramaditya sah  Email id : [vsah625@gmail.com](mailto:vsah625@gmail.com)  Contributions : 1)Writing introduction and Conclusion summary  2)Assist in feature engineering  3)Technical documentation  4) Contribution in PPT  2)Name : Sivaramaguhan S  Email id : [ssivaramaguhan@gmail.com](mailto:ssivaramaguhan@gmail.com)  Contributions : 1)Data wrangling  2)EDA and Visualization parts  3)Contribution in PPT |
| **Please paste the GitHub Repo link.** |
| Github Link : https://github.com/vikram0050/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)**  In today’s digitized modern world, the popularity of food apps is increasing due to their functionality to view, book, and order food with a few clicks on the phone for their favorite restaurant or cafes, by surveying the user ratings and reviews of the previously visited customers. Food app like Zomato provides a secular part where user can rate their experience of the visited restaurant or café. Zomato is a site where someone can give a review of a restaurant, how the restaurant is, and someone's opinion about the restaurant. Restaurant customer satisfaction can be analyzed by their review on Zomato. Sometimes, restaurants see the reviews in Zomato, but they don't get if the reviews are positive or negative to their restaurants. Reviews on Zomato are still in the form of text and can be classified with positive, negative, or neutral ratings. |
| Starting with EDA we found the most and least expensive restaurants and also found out the critics which have rated more than 100 restaurants and have more than 10000 followers on which restaurant staff should focus.  For clustering, we have decided on 3 clusters after the Silhouette score plot and elbow plot where we used KMeans clustering and Hierarchical clustering algorithms. During clustering, we created a superset of cuisine and cost as features for the clusters.  For sentiment analysis, we used both supervised and unsupervised techniques where we used Linear Discriminant Analysis & Non-negative Matrix Factorization for unsupervised and Multinomial NB, Decision trees, and its ensembles for supervised learning. In Sentiment Analysis, we converted the ratings above 3.5 as positive and the rating below 3.5 as negative. After tuning the hyperparameters and using the cross-validation technique, the optimum hyperparameters were chosen for the models and the best model was chosen as Lightgbm classifier for supervised learning technique. |