1. *What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.*

The name and primary developer for the course project is me, Hiranmayi Palanki. My NetID is: palanki2. I’ll serve as the captain as well as the team member as I’ll be developing the project individually / solo.

1. *What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?*

My topic is “Sentiment Analysis of Restaurant Reviews”. The objective of the task is to develop a machine learning model capable of discerning / predicting the sentiment behind restaurant reviews.

This topic is particularly interesting in gaining deeper insights from customer feedback than merely relying on numerical ratings. By interpreting the sentiment expressed in written reviews, restaurants can enhance their services and overall business performance.

The importance of this approach lies in the fact that individuals can convey their genuine experiences more comprehensively through words, which often surpasses the limitations of a numerical rating system. Consider, for instance, two patrons who share an identical dining experience; yet their respective numerical scores may differ – one might assign a 10, while the other opts for an 8. This variance can impact the overall rating system. However, by encouraging patrons by articulating their experiences in their own words and subsequently employing sentiment analysis to gauge the sentiment, we can establish a more standardized method of collecting customer feedback for businesses, particularly within the restaurant industry.

The tools employed for the completion of this project will include:

* + Python
  + A selection of libraries tailored for specific project tasks, as below:
* Machine Learning: Scikit-Learn
* Data Loading, Handling and Manipulation: Pandas, NumPy
* Visualization: Matplotlib, Seaborn
* Saving, Loading Machine Learning models and associated files: Pickle

The dataset which will be utilized for this project is the “Malaysia Restaurant Review Dataset”, which was meticulously compiled through web scraping of restaurant reviews from both Trip Advisor and Google. This dataset was focused on restaurants situated in Malaysia and was filtered according to the restaurant’s specific geographic locations. The dataset can be accessed at the following link: <https://www.kaggle.com/datasets/choonkhonng/malaysia-restaurant-review-datasets>

The expected outcome of this project is the development of a well-trained machine learning model capable of accurately categorizing restaurant reviews as positive, neutral, or negative. This outcome is intended to serve as a valuable tool for restaurants, enabling them to conduct in-depth analyses of their customer reviews.

Evaluation of the course project hinges on two key aspects. Firstly, it involves an evaluation of the model’s performance by gauging various metrics. Secondly, it encompasses the real-time performance of the model, which is assessed through user inputs fed into the system as part of the model’s ongoing evaluation.

1. *Which programming language do you plan to use?*

· Python

1. *Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.*

I will be the sole contributor to this project, and I expect the time allocation for various project tasks to be as follows:

* Initial project research: 3 hours
* Text retrieval and data cleaning: 1 hour
* Initial algorithm implementation: 16 hours
* Testing and enhancing model performance: 20 hours
* Preparing final reports, documentation, and conducting a demo: 5 hours