CS 410 Project Proposal

Topic: Sentiment Analysis on Steam Game Reviews (Theme 5: Free Topics)

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.

   Team Members:
   Xinyu Lian (lian7)
   Yu Zhang (yz93)
   Yuankai Wang (yuankai4) (captain)
   Yutao Zhou (yutaoz3)

2. 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?

   Topic:
   We will do a sentiment analysis on the steam game review using the dataset from kaggle (https://www.kaggle.com/datasets/smeeeow/steam-game-reviews).

   Tasks:
   - Perform analysis on the review text and see if it shows a positive or negative attitude towards the game. Then compare the result with the actual review rating (like or dislike).
   - Identify common positive and negative feedback themes by analyzing the text. By identifying recurring themes and patterns in positive and negative reviews, helping game developers understand what aspects of their games players appreciate or dislike.
   - Do the same analysis (as previous two tasks) about what review can get more upvotes or considered funny.

   Importance:
   - **Consumer Behavior Insight**: Understanding what players value can guide developers in the right direction, ensuring better user satisfaction in future game developments.
   - **Improved Recommendation Systems**: If we can capture sentiment from textual reviews accurately, recommendation systems can be enhanced by considering more than just binary (like/dislike) feedback.
   - **Community Management**: Detecting prevalent sentiments can aid community managers in addressing issues or controversies in real-time.
   - **Broad Applicability**: Techniques honed here can be transferred to other domains, making our approach versatile.

Approach:

- Data Preprocessing:
  - Processing on the dataset
  - Processing on the review text. Including tokenizing and removing special characters, converting to lowercase, and stemming or lemmatizing the words.
- Feature Extraction:
  - Use techniques like TF-IDF or word embeddings to convert the text data into numerical features.
- Model Selection and Training:
  - Choose a sentiment analysis model such as Naive Bayes, Support Vector Machine (SVM), or a deep learning model.
  - Split the dataset into a training and testing set.
  - Train the model on the training set and evaluate its performance using accuracy, precision, recall, and F1-score on the testing set.
- Predict Sentiments:
  - Apply the trained model to predict the sentiment of user reviews.

Datasets: https://www.kaggle.com/datasets/smeeeow/steam-game-reviews

Expected Outcome:
We will be able to classify user reviews as positive, negative, or neutral, providing insights into how players feel about the games

Evaluation:
Compare our expected result with the actual rating and see how close they match.

3. Which **programming language** do you plan to use?
   Python. Python is used in course MP and is renowned for its capabilities in data analysis, offering a rich library and tools.

4. 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**.

Task	**Estimated Time Cost**
Environment Setup	5h
Data Preprocessing	5h

Build and Train the model based on like or dislike information	20h
Find out the keywords affecting like or dislike	10h
Build and Train the model based on upvotes and funny votes on the review	20h
Find out the keywords affecting the upvotes and funny votes on the review	15h
Evaluate the model performance	10h