## Project Task Log Summary

#### **Team Member 1: Lawrence Qiu**

**Week 1-2: Project Planning and Setup**

Collaborated on defining the project scope and deliverables

Set up the development environment and tools required for the project

**Week 3-4: Feature Engineering**

Developed TF-IDF feature extraction for sentiment analysis

Implemented feature selection techniques to improve model accuracy

**Week 5-6: Advanced Modeling Techniques**

Researched and integrated FastText for text classification

Began development on rating prediction using Linear Regression

**Week 7-8: Integration and Testing**

Integrated sentiment analysis with rating prediction models

Conducted thorough testing to ensure accuracy and reliability

**Week 9-10: User Interface Development**

Developed a user-friendly interface for model interaction

Implemented input validation and error handling

**Week 11-12: spaCy Integration for Similarity Analysis**

Integrated spaCy for advanced text processing and similarity analysis

Developed methods to compare review texts and calculate similarity scores

Fine-tuned model parameters to improve performance

Explored advanced techniques like word embeddings and deep learning

**Week 13-14: Final Testing and Refinement**

Created visualizations to understand data distribution

Analyzed model performance and identified areas for improvement

Performed final tests on the complete system

Refined code and interface based on test feedback

Prepared presentation materials and finalized project documentation

#### Team Member 2: Gurnoor

**Week 1-2: Project Kick-off and Initial Research**

Identified project objectives and goals

Conducted preliminary research on sentiment analysis methodologies

Compiled a list of potential datasets to use

**Week 3-4: Data Acquisition and Preprocessing**

Acquired IMDB movie reviews dataset from Kaggle

Preprocessed data by cleaning and tokenizing text

**Week 5-6: Model Exploration**

Explored different machine learning models suitable for text classification

Implemented baseline models using Logistic Regression and Naive Bayes

**Week 7-8: Model Training and Evaluation**

Trained sentiment analysis models on the preprocessed dataset

Evaluated models using metrics such as accuracy, precision, recall, and F1 score

**Week 9-10: Integration and Testing**

Integrated sentiment analysis with rating prediction models

Conducted thorough testing to ensure accuracy and reliability

**Week 11-12: Model Optimization**

Fine-tuned model parameters to improve performance

Explored advanced techniques like word embeddings and deep learning

**Week 13-14: Finalization and Documentation**

Finalized sentiment analysis models

Documented the research process, model development, and evaluation results

Prepared presentation