**Project Title:** Stock Sentiment Analysis Web Crawler

**Team Members:** Dmitry Smirnov (dmitrys3), Charles Yu (cyu66), Haosen Yao (haoseny2), Lambodar Panigrahy (lp18)

**Project Coordinator:** Dmitry Smirnov

1. **Progress Made:** We have assigned tasks for each person in our group. Person1 is going to research and identify APIs and data sources for financial articles and social media platforms (e.g., Kaggle, Twitter API). Person2 is going to develop and implement a topic modeling pipeline (e.g., using Latent Dirichlet Allocation (LDA) or BERT). Person3 is going to integrate Microsoft’s AutoGen toolkit or an equivalent framework to create the intelligent task agent. Person4 is going to design and develop a user-friendly web interface. So far, we have completed the first step to collect data for topic modeling. By using Tweepy API, we collected tweets and store those into a CSV file, which then will be cleaned by removing noise and stopwords and saved it to a cleaned CSV file. We removed unnecessary elements such as URLs, mentions, hashtags, special characters, and stopwords by using libraries like re for regex-based cleaning and nltk for stopwords and tokenization.
2. **Remaining Tasks:**
3. **Any Challenges/Issues:**