DSCI 510 – Fall 2025

Final Project Progress Report

Student: Shivam Shukla
Project Title: Analyzing the Relationship Between Financial News Sentiment and Market Volatility (2017–2020)
Date: November 13, 2025

Project scope update
The original goal was to study how financial news sentiment relates to stock market behaviour. I have refined the scope to focus on the 2017–2020 period and on market-level behaviour instead of individual stocks. The project now analyzes how sentiment in timestamped financial news headlines relates to S&P 500 index returns and VIX volatility levels. I plan to compute sentiment for each headline using TF–IDF features and a Linear SVC classifier trained on labelled financial text, then aggregate these predictions into daily sentiment scores that can be compared against market performance and volatility over time.

Data sources
• **Financial News Headlines (Kaggle):** A CSV dataset of approximately 50,000 timestamped financial news headlines from 2017 to 2020. The headlines are stored locally in the *data/* directory (not tracked in Git) and will be used as input for text preprocessing and sentiment classification.
• **S&P 500 Index (^GSPC):** Daily Open, High, Low, Close, and Volume data for 2017–2020 obtained programmatically from the Yahoo Finance API using the yfinance Python library. This provides market return information.
• **VIX Index (^VIX):** Daily VIX closing values for 2017–2020, also retrieved via yfinance. This serves as a measure of expected market volatility (the “fear index”) and will be used as a control or secondary dependent variable.

Issues / difficulties
So far the main challenges involve correctly aligning dates between the news headlines and market data, since markets are closed on weekends and holidays and some dates may have few or no headlines. Another anticipated difficulty is handling neutral or ambiguous headlines during model training and evaluation, and tuning the Linear SVC with high-dimensional TF–IDF features while keeping training time reasonable. I have successfully implemented and tested an API loader using yfinance (for ^GSPC and ^VIX) and added a test function in src/tests.py, so the next steps are to finish loading the news dataset, build the sentiment model, and begin exploratory analysis of the merged time series.