Proposal: Analyzing Financial News to Predict Stock Sentiment and Price Trends

1. Introduction
Financial news significantly influences stock prices by shaping market sentiment and investor behavior. This project will focus on identifying sentiment in financial news articles and correlating it with stock price trends for a few key companies across multiple industries. The aim is to build a straightforward predictive model to gauge price movement based on sentiment data from financial news sources.

2. Objectives
Primary Objectives:
- Analyze sentiment around stock tickers based on financial news.
- Correlate sentiment with daily stock price movements.
- Develop a simple predictive model to forecast price trends based on sentiment.
- Compare the impact of financial news sentiment across different industries.

3. Data Collection
Data Sources:
- **Financial News**: Articles from one or two major financial news platforms (e.g., Reuters, Bloomberg).
- **Stock Prices**: Historical stock data from free sources like Yahoo Finance.

Scope:
- **Focus on 10 companies per industry, covering three industries**:
  - **Technology**: For example, Apple (AAPL), Microsoft (MSFT), Tesla (TSLA), NVIDIA (NVDA).
  - **Healthcare**: For example, Pfizer (PFE), Moderna (MRNA), Johnson & Johnson (JNJ), AbbVie (ABBV).
  - **Finance**: For example, JPMorgan Chase (JPM), Goldman Sachs (GS), Bank of America (BAC), Morgan Stanley (MS).

- **Goal**: Collect approximately 200 articles per industry, ensuring balanced representation across companies.

4. Methodology
4.1 Data Preparation
- **Text Preprocessing**: Clean text data by removing irrelevant symbols, stopwords, and non-informative content.
- **Data Alignment**: Align stock price data with daily sentiment scores derived from news articles.
  - **Handling Missing Sentiment Scores**: Apply time series imputation techniques or use the average sentiment score across similar companies to fill missing values.

4.2 Sentiment Analysis

- **Sentiment Model**: Use a pretrained sentiment model (e.g., VADER or FinBERT if time allows) to classify sentiment as positive, negative, or neutral.
  - **Model Comparison**: Compare the effectiveness of VADER vs. FinBERT for capturing sentiment in financial text.
- **Aggregation**: Calculate average daily sentiment for each stock by aggregating results across news sources.

4.3 Feature Engineering
- **Feature Selection**: Combine daily sentiment scores with historical price features, such as daily returns, volatility, trading volume, lagged sentiment scores, and moving averages.

4.4 Predictive Model
- **Simple Time Series Model**: Use a basic model like linear regression, Random Forest, or a lightweight LSTM to predict price direction (up/down) based on sentiment and past prices.
- **Advanced Model Exploration**: Experiment with more advanced models like XGBoost and Random Forests to see which performs best in predicting price movements.
- **Evaluation**: Use metrics such as accuracy, directional accuracy, mean absolute error (MAE), and time-series specific metrics like Mean Squared Logarithmic Error (MSLE) to assess model performance on a small test set.

4.5 Visualization
- **Sentiment Trends**: Create line charts to display sentiment changes over time for each stock.
- **Price and Sentiment Correlation**: Plot sentiment against stock price to visualize relationships.
- **Industry Comparison**: Use bar charts or scatter plots to compare the strength of sentiment impact across industries.
- **Price Sensitivity Analysis**: Create visualizations to show how different levels of sentiment impact stock prices.
- **Stock Movement Prediction**: Show model predictions for future price movements based on sentiment data.

5. Evaluation
5.1 Model Accuracy
- **Metrics**: Evaluate the model's accuracy, directional accuracy, MAE, and MSLE in forecasting stock movement direction (up or down).
- **Robustness Check**: Test model stability across different training and test splits using cross-validation techniques tailored for time series, such as TimeSeriesSplit.

5.2 Sentiment Correlation Check
- **Time Lag Analysis**: Evaluate the time lag between sentiment and stock price impact to understand how quickly market sentiment affects stock movements.
- **Cross-Correlation Analysis**: Use cross-correlation techniques to evaluate the lag effect of sentiment.

5.3 Industry Comparison
- **Statistical Testing**: Perform statistical tests (e.g., ANOVA, Kruskal-Wallis) to determine if the impact of sentiment differs significantly across industries.
- **Sensitivity Analysis**: Use sensitivity analysis to evaluate how different sentiment scores impact stock movement predictions.

5.4 Error Analysis
- **Analysis of Failure Cases**: Conduct an analysis of the model's mispredictions by reviewing specific companies and market events. This may offer insights into model weaknesses and ways to improve.

6. Expected Outcomes
This project is expected to provide:
- **Insights into Sentiment and Stock Price Trends**: Understanding how financial news sentiment influences stock price trends.
- **Predictive Model for Stock Movement**: A simple, yet interpretable predictive model for stock movement based on sentiment and historical price data.
- **Visualization Tools**: Visualizations that effectively communicate sentiment trends and their impact on stock prices across industries.
- **Industry-Specific Insights**: Comparative analysis of sentiment impact across sectors, offering insights into industry-specific market behavior.