UPDATE -1

1. **Introduction Revision:** "In our project, we use Yahoo Finance's Adjusted Close data for major stocks, including Amazon, Apple, Google, Intel, and Nvidia, from 2016 to 2024. Our goal is to forecast future stock prices using machine learning techniques."

2. **Technical Indicators Description:** "We employ several technical indicators as features in our predictive model, including the Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Bollinger Bands, Williams %R, Commodity Channel Index (CCI), and Percentage Price Oscillator (PPO)."

3. **Model Performance and Adjustments:** "Our gradient boosting regression model has achieved a Root Mean Square Error (RMSE) of 0.71 to 0.85 on training data and 2.4 to 4.5 on testing data, indicating promising predictive capabilities."

4. **SARIMA Model Integration:** "To further refine our predictions, we integrated a SARIMA model with exogenous variables from our initial machine learning model to capture seasonal trends and cyclic patterns. We utilized an Autoarima function for each stock ticker after selecting appropriate differencing and transformation methods."

5. **NLP for Sentiment Analysis:** "To complement our quantitative models, we are applying NLP techniques to analyse sentiment in news headlines. We aim to understand how sentiment correlates with stock price movements. Initial tests with pre-trained models like DistilBERT have shown promising results in sentiment classification."

6. **Challenges and Data Preprocessing:** "We encountered challenges in data preprocessing, particularly in standardizing the formats of dates and headlines across different sources. Ensuring consistency in our data is crucial for accurate model training and forecasting."

7. **Future Work on SARIMA and LSTM Integration:** "Our next steps include developing a universal SARIMA model applicable to all stock data and integrating an LSTM model to account for random fluctuations not captured by other models."

8. **Data Sources Expansion:** "To enhance our data sources for sentiment analysis, we are incorporating news data from Alpaca and Hugging Face, which provide extensive datasets that can help in identifying market sentiment and potential price movements."

9. **Conclusion and Future Work:** "Looking ahead, we plan to integrate machine learning, SARIMA, and NLP into a cohesive predictive pipeline. This will improve our forecasts and allow us to dynamically adjust our models based on real-time data and sentiment analysis results."