Volatility Prediction Project Summary

Our project worked on building a machine learning model to predict whether the SPY (S&P 500 ETF) would have high or low volatility each day. The goal was to help traders make better investment and risk management choices. We used past market data like closing prices, highs, lows, opening prices, and trading volumes, plus extra features we created, such as returns, rolling volatility, Average True Range (ATR), skewness, kurtosis, Bollinger Bands, and average volume. Each feature was an average of the previous 20 days (later changed to 14 days), and we predicted the volatility for the next day. We used a Random Forest Classifier, which gave us a 97% accuracy on both training and testing data. Cross-checks and other tests showed the model was solid. We also made charts, like correlation matrices, to see how features were related. This project showed that machine learning can spot patterns in market volatility, making it a useful tool for traders dealing with up-and-down markets.

We hit some bumps along the way. At first, we tried a simpler model (logistic regression), but it wasn't good at predicting low volatility. Switching to Random Forest gave us 97% accuracy, which seemed too good to be true, so we thought the model might be overfitting. But when we compared training and testing scores, they were close, and other tests backed this up, proving it wasn't overfitting. Then we realized predicting next-day volatility was too easy because markets usually stay similar day-to-day. To make it more useful, we decided to predict volatility 3, 5, or 7 days ahead, which is harder but more valuable. Some features didn't connect strongly to volatility. Tuning the Random Forest model was tricky too—we had to adjust settings like tree depth and leaf size to keep it balanced.

Moving forward, we're making the model better. We're switching to a 14-day lookback period to make features more relevant and trying out new features. We're also visualizing features to find connections and picking the best ones with tests. In the future, we want to add data like economic news or market sentiment to improve predictions. We plan to update the model in real-time to keep up with market changes and test it in real trading to see how it performs. We also want to make the model easier to understand so users know why it makes certain predictions. By tackling these challenges and adding these improvements, our project aims to create a reliable and clear tool for predicting volatility, helping traders make smarter decisions in fast-changing markets.