Executive Summary

This project made a RandomForestClassifier to predict whether MSFT's next day close will be higher than today's, using five technical features (weekly momentum, 21-day volatility, 14-day RSI, MACD, MACD signal). A train/test split is used to split the data and the last twenty trading days are used to evaluate the performance of the model. The model made an accuracy of 60%, which is an improvement over the random guess.

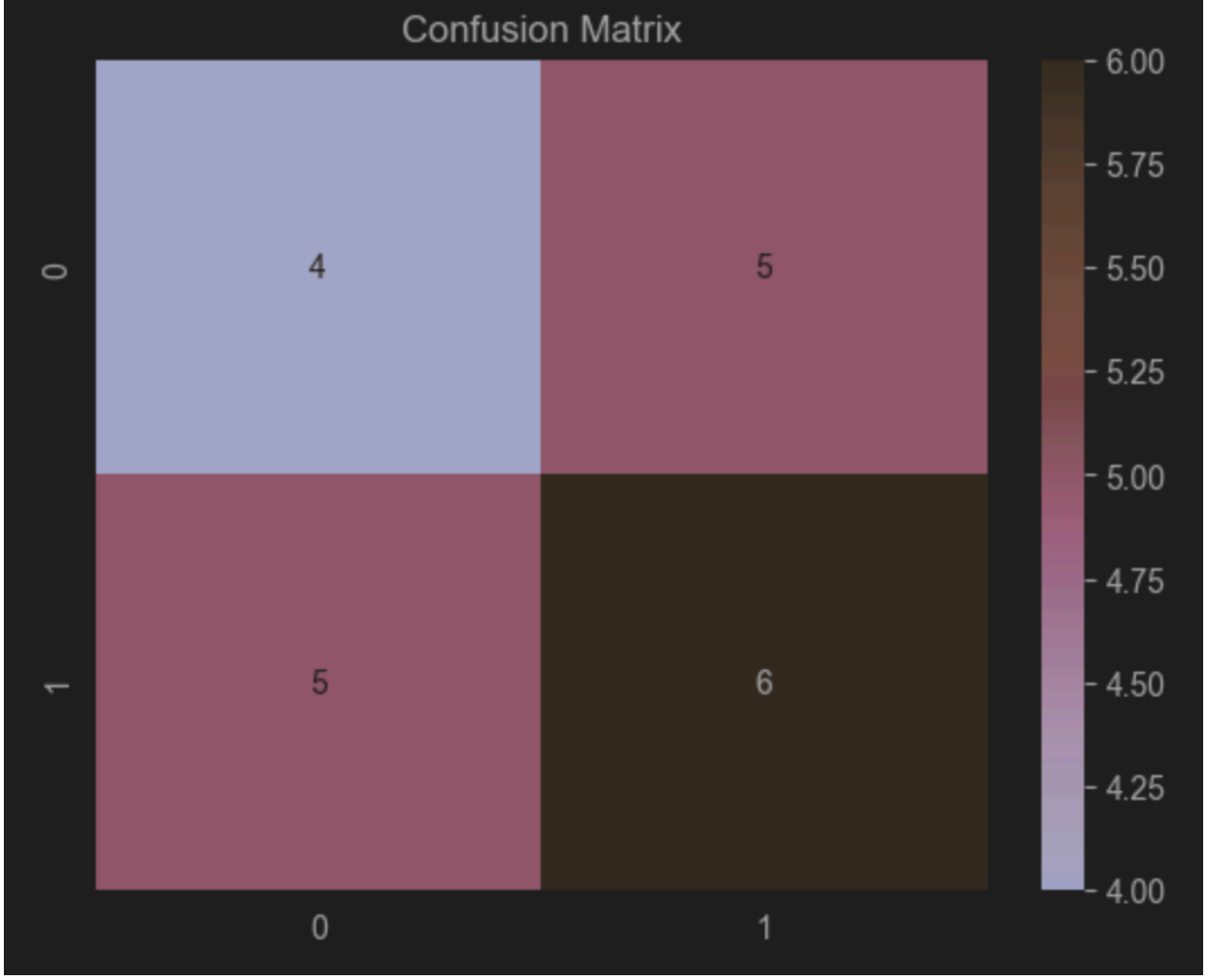
Charts with Assumptions & Risks

The model assumes that the market can be predicted, past data have predictive power of the future, and there is no structural break in between. Transaction cost is assumed to be 0.

- When the model predicts "down", it's right 44% of the time.
- When it predicts "up", it's right 73% of the time.
- The model assumes this accuracy to hold for the future.

Sensitivity Analysis

	precision	recall	f1-score	support
0	0.44	0.44	0.44	9
1	0.55	0.55	0.55	11
accuracy			0.50	20
macro avg	0.49	0.49	0.49	20
weighted avg	0.50	0.50	0.50	20



When the RSI feature was removed, the forecast accuracy decreased to 0.5, which is the same as a random guess. It shows that RSI is a useful feature.

Next Steps

The next step is to try different features while controlling the risk of overfitting. Different models, like logistic regression or XG Boost can be used in comparison to the random forest model. Transaction cost should also be considered. The model should also be tested on more datasets to ensure generalization.

Conclusion

While the random forest model offers a better performance compared to randomly guessing, it is prone to overfitting and requires further optimization. The model serves as a good starting point for more complex research. A forecasting accuracy of 60% is not high in the machine learning field, but given the complexity of forecasting in finance, it is a reasonable result.