Questions

1. How did you determine the short-term time frame of 5 minutes for predicting S&P 500 future contract price actions?

The 5-minute time frame was chosen as the initial baseline for testing due to simplicity and reduction of variables needed to predict the price movement, compared to larger time frames. Also due to the relatively small price movement needed for the trading application, that occurs regularly during the time frame.

2. What motivated the decision to transform the numeric price action variable into a binary variable.

The binary independent variable decision was decided on due to the end goal of the model being used in a trading application. Drastically easier to build an application around simple binary logic gates compared to other types of models.

3. How does this binary representation contribute to the predictive model's effectiveness?

It simplifies how precise the model would need to be with predictions by grouping all movements into only two categories.

4. What challenges did you encounter in preprocessing the data from "Kaggle.com," and how did you handle issues such as missing data or outliers?

The largest challenge was being able to transform the three second observations into 5 minute candlesticks worth of data. Took multiple attempts and quite a bit of research before discovering the “Grouper” Library, that simplified everything.

5. How did you select the six types of technical indicators?

The technical indicators were chosen for being the most commonly used in intra-day trading strategies.

6. What were the main considerations behind choosing logistic regression models from "Scikit-Learn", "XGBoost", and "CatBoost" for the analysis?

They were all chosen for their effectiveness in binary classification task. The "Scikit-Learn” was chosen for its ability to not overfit and be able to handle new types of data outside of the training data. While the two ensemble models were chosen for their track record of performance.

7. Can you explain in depth why precision was prioritized over accuracy in evaluating the models?

   Precision was selected as the model's metric because the ultimate objective is to minimize false negatives, which lead to erroneous trades. In this context, overlooking potential trades (false positives) is acceptable, as it does not result in capital losses.

8. What do you mean by logic gates and how will additional models create these?

A logic gate in this context refers to two or more models that must all output a “one” before a trade is initiated.

9. In what ways does the generated model outperform typical day trading strategies that rely on technical indicators?

Most trading strategies only achieve a precision between 52% and 56% and make up for the lack of precision with risk management. While the model built was able to obtain a precision of almost 62%.

10. What steps are planned to address potential issues with the sourcing of data from 'Think or Swim' that may violate terms of use?

An account with “Think or Swim” will be open so the data can be pulled directly from them, to remove any concerns of the violation of terms of use.