Torrin Pataki

False Signal Filtering for Technical Trading Strategies using Machine Learning

Independent technical analysts applying trend-following and momentum-based strategies in equity and options markets often face the challenge of false signals generated during mean-reverting price action. For an analyst utilizing technical indicators to identify breakouts, these false signals can lead to a significant number of failed trades, increased transaction costs, wasted analyst time and lower overall strategy performance. This project addresses these key pain points by training a custom machine learning model to filter false breakout signals, with the goal of selecting a more profitable subsample of trades while simultaneously lowering the strategy's time in market.

Data Preparation:

To address the non-IID nature of financial time series when used for training a statistical learning model, I created dollar bars from daily time bars. This approach aims to create a more statistically consistent and uniform representation of price action while preserving the underlying granularity of the prices' path. This is critical for the application of the triple-barrier labelling method as introduced by Lopez de Prado (2018), which I implemented to generate labels based on robust stop-loss and take-profit barriers tailored to the analyst's existing strategy. I also employed meta-labelling, layering my model to predict the probability of the binary classification labels generated from the primary technical strategy borrowed from FerencFinance. Taking this step ensures the versatility of my system as an analyst's needs change: the underlying primary strategy is modular, allowing it to be replaced and then immediately used to create labels for the same meta-labeling model.

Technical Feature Engineering:

I calculated a variety of technical factors on my dataset, focusing on momentum, volatility, and trend indicators. As suggested by Jansen (2020), I derived these indicators using TA-Lib and pandas-ta. I used specific indicators and a setup that the user of FerencFinance employs in their existing proprietary technical analysis. Due to the constraints imposed by the proprietary nature of this setup, I will omit certain details; however, the accessibility of these indicators and modularity of feature selection in the model development process reinforce the robustness of my system to changes in an analyst's strategy.

Model & Backtesting:

I chose to use Sklearn's implementation of a random forest model as the initial baseline due to its relative robustness to overfitting through bagging. I intend to test other models however, for illustration purposes, I believe a random forest is sufficient and relatively intuitive to understand as an ensemble of decision trees. A key issue with FerencFinance was the use of an in-sample backtest to evaluate strategy performance, and as suggested by Carver (2015), I chose to use a walk-forward optimization to prioritize realistic out-of-sample evaluation. This involved training the model on an expanding window of historical data and testing its performance on a rolling basis. The rolling window ensures that the model is being evaluated on data it has not seen before, though I acknowledge it is not perfect as information leakage from overlapping labels may still inflate my results.

Results:

While backtest overfitting is an obvious and ubiquitous concern, my initial results are very promising, suggesting my model's potential to improve the analyst's existing strategy. The confusion matrix seen in Figure 1 demonstrates the model's ability to successfully distinguish between profitable and unprofitable trades, with an accuracy of 0.649 and an area under the ROC curve of 0.703.

```
All Trades PF 0.8274039694286777
All Trades Avg 0.013219709872033171
All Trades Win Rate 0.4981132075471698
All Trades Time In Market 0.5093167701863354
Meta-Label Trades PF 0.8726319303603768
Meta-Label Trades Avg 0.014849459005242927
Meta-Label Trades Win Rate 0.6041666666666666
Meta-Label Time In Market 0.3064182194616977
Confusion Matrix:
[[45 19]
 [21 29]]

Accuracy: 0.6491228070175439

Balanced Accuracy: 0.6415625

Classification Report:
              precision    recall  f1-score   support

        SELL       0.68      0.70      0.69        64
         BUY       0.60      0.58      0.59        50

    accuracy                           0.65       114
   macro avg       0.64      0.64      0.64       114
weighted avg       0.65      0.65      0.65       114

AUC: 0.7028125000000001
Log loss: 0.6560579452943371
```

Figure 1

Importantly, the backtest shown in Figure 2 indicates a noticeable reduction in trades during periods of a downward trend or mean-reversion price movement, leading to an increase in profit factor and a promising decrease of time in market. These results are unrealistically high and further refinements are needed to address potential biases due to overlapping labels. Though the results certainly will provide inflated expectations for live model deployment, they still serve as an effective demonstration of the model's effectiveness in selecting trades preceding upward trends.

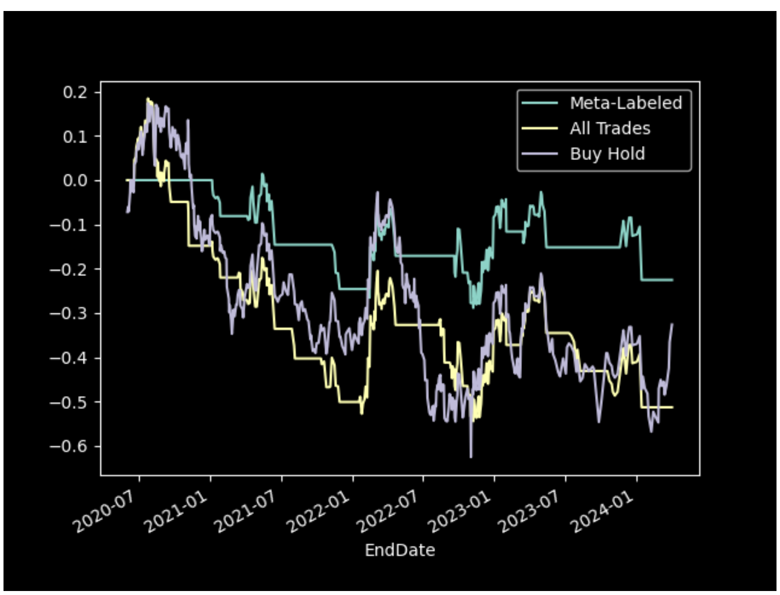

Figure 2

Lessons Learned:

The most important lesson that I have learned is the importance of data preparation and creating IID examples for machine learning in finance. Luckily, I am sufficiently informed to identify that the lack of independence between labels is likely causing my unrealistically high performance; however, a less experienced practitioner may view this backtest and be eager to deploy the model to live trading. To help amend the information leakage between labels, further research and implementation of sample weighting techniques are needed to fully mitigate the impact of overlapping labels and enhance the data quality. As a secondary lesson, I feel my results validate my hypothesis that a machine model can effectively classify trends and filter out potentially unprofitable mean-reversion periods, which is a critical challenge for the FerencFinace user's core strategy. Verifying this hypothesis has set me on a path of continuous development, and I have an ambitious list of improvements and integrations I want to add to this project.

Future Considerations:

Firstly, I will address my previously discussed independence issues by implementing sample weighting to enhance data quality and reduce the correlation between individual trees in my random forest. Additionally, I am interested in supplementing my walk forward backtesting with cross-validation techniques namely the Combinatorial Purged Cross-Validation Method as introduced by Lopez de Prado (2018), where an exhaustive number of combinations of backtest paths are tested rather than only the historical path. After cross-validating and reevaluating my model, I will have the confidence to deploy it on a paper trading platform such as Interactive Broker, where the model will automatically buy and sell using Interactive Broker's live datasets. This will also be far superior to yfinance as it will have higher granularity tick data than is offered on yfinance's free library. To tie my entire vision for this project together, I would add the model as a selectable strategy within the FerencFinance application, providing the analyst with a customized and streamlined ML-driven analysis tool. Finally, as general areas of interest to explore with this project, I would explore the use of sentiment analysis to gauge overreaction to news events and identify potential periods of mean reversion. I would also like to evaluate different modelling techniques, such as a boosting algorithm, for potential performance improvements.

Conclusion

This project demonstrates the power of combining financial domain knowledge with modern computer science and data analyst skills. Inspired by leaders in quantitative analysis like Marcos Lopez de Prado (2018) and Stephen Jansen (2020), I am aiming to empower independent analysts by creating sophisticated tools typically reserved for larger institutions. If these filtered signals help avoid even a small number of losing trades for the analyst, it translates into real monetary value, making this an endeavor with both intellectual satisfaction and tangible impact.

References

Carver, R. (2015). *Systematic trading: A unique new method for designing trading and investing systems*. Harriman House Ltd.

Jansen, S. (2020). *Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and Alternative Data for systematic trading strategies with python*. Packt.

Prado, M. L. de. (2018). *Advances in financial machine learning*. Wiley.