# **Machine learning** and **statistical** methods can't outperform the **naïve** forecasting method

## Forecasting Accuracy Comparison of Various Machine Learning and Statistical Models on Stock Market Price Movements

***Ruan Pretorius***, *University of The Witwatersrand*

*Supervisors: Prof. Terence van Zyl, Dr. Farai Mlambo*

**ABSTRACT**

Accurate financial time series forecasts can assist investors in gaining a competitive edge over other participants in capital markets. No empirical conclusion existed on what the most accurate model(s) were for forecasting stock market price movements over different forecast horizons. Limitations from previous studies were addressed in this study by compared the forecasting accuracy of 20 different models on 403 time series of stocks/indices. These included machine learning (ML), statistical, and benchmark models. The naïve benchmark model outperformed all other models in this study for nearly all accuracy metrics and forecast horizons tested.

## PROBLEM STATEMENT

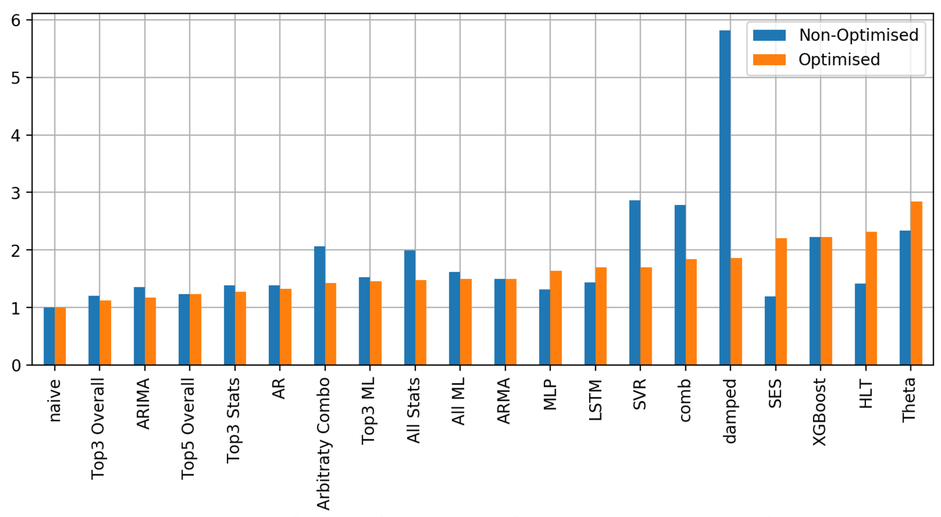
- Previous research yielded contradicting results on superiority of ML over statistical models [1]
- No empirical conclusion on most accurate model(s) for forecasting stock market price movements over different forecast horizons [2],[3],[4],[5]
- Existing studies lacked comparisons to benchmarks, lacked multi-step-ahead forecasts, and used limited data sets [1],[2],[3]

## METHODOLOGY

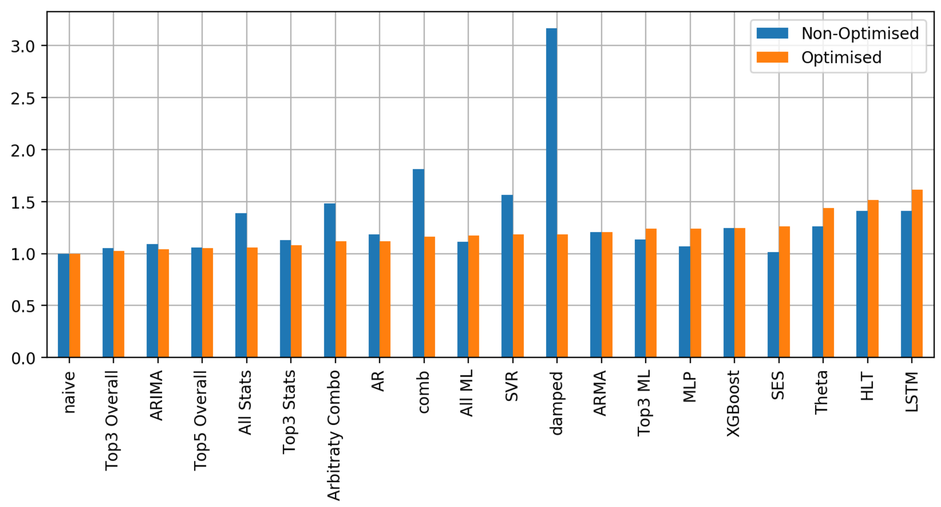
- Compared forecasting accuracy of 20 different models on 403 time series of stocks/indices on NYSE and Nasdaq exchanges
- Included ML, statistical, and benchmark models as well as combinations of these
- Included 12 of the most successful models from previous studies and stock market forecasting surveys [4],[5],[6],[7]
- Forecast errors assessed: 18 time steps, three different accuracy measures (sMAPE, MASE, OWA)
- Non-optimised models based on those of previous studies [1]
- Some models further optimised using grid-search

## RESULTS

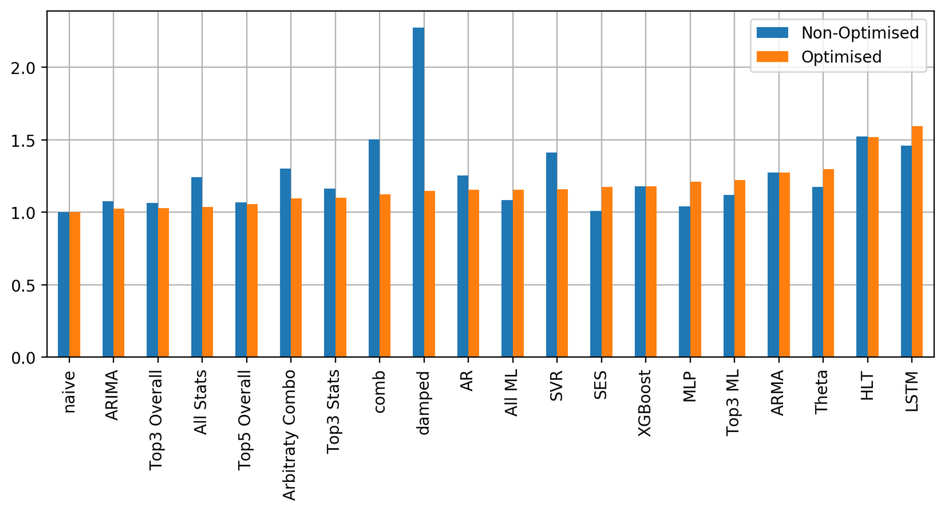
*OWA error for short-term forecasts (1-6 months)*




*OWA error for medium-term forecasts (7-12 months)*




*OWA error for long-term forecasts (13-18 months)*




## CONCLUSIONS

- No conclusion could be drawn on superiority of pure ML over pure statistical models
- The top performing models included combinations of both pure ML and pure statistical models
- The naïve benchmark model outperformed all other models in this study for nearly all accuracy metrics and forecast horizons tested
- When forecasting monthly stock market price movements between one and 18 months, no model from this study outperformed a simple random walk (naïve) model

## REFERENCES

[1] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward". In: PLOS One 13.3 (2018).
[2] Lukas Ryll and Sebastian Seidens. "Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting: A Comprehensive Survey". In: arXiv preprint arXiv:1906.07786 (2019).
[3] Bjoern Krollner, Bruce J Vanstone, and Gavin R Finnie. "Financial Time Series Forecasting with Machine Learning Techniques: A Survey". In: European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges, Belgium: ESANN, Apr. 2010. URL: http://works.bepress.com/bjoern_krollner/1/
[4] G Atsalakis and Kimon P Valavanis. "Surveying Stock Market Forecasting Techniques - Part I: Conventional Methods". In: Journal of Computational Optimization in Economics and Finance 2.1 (2010), pp. 45–92.
[5] George S Atsalakis and Kimon P Valavanis. "Surveying Stock Market Forecasting Techniques - Part II: Soft Computing Methods". In: Expert Systems with Applications 36.3 (2009), pp. 5932–5941.
[6] Spyros Makridakis and Michele Hibon. "The M3-Competition: Results, Conclusions and Implications". In: International Journal of Forecasting 16.4 (2000), pp. 451–476.
[7] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. "The M4 Competition: 100,000 Time Series and 61 Forecasting Methods". In: International Journal of Forecasting 36.1 (2020), pp. 54–74.

**ACKNOWLEDGEMENTS**

- Research sponsor : DSI-NICIS National e-Science Postgraduate Teaching and Training Platform
- Poster template: Mike Morrison