

Data Article

Title: Short-Term Forecasting of the JSE All Share Index Using Gradient Boosting Machines

Authors: Mueletshedzi Mukhaninga, Thakhani Ravele and Caston Sigauke

Affiliations: Department of Mathematical and Computational Sciences, University of Venda, Private Bag X5050, Thohoyandou, 0950, South Africa.

Contact email: caston.sigauke@univen.ac.za

Abstract

The paper discusses an application of Gradient Boosting Machines (GBM) to forecast the Johannesburg Stock Exchange (JSE) All Share Index (ALSI) closing price for the next trading day based on different training-testing split ratios. Daily data from the Wall Street Journal website, from 14 October 2009 to 16 April 2024, is used. Two models, Gradient Boosting Machines (GBM) and Principal Component Regression (PCR), are used for short-term prediction of the all-share index. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE) are used to assess the performance of the models. The results show that GBM consistently outperforms PCR across all split ratios, delivering more accurate predictions. These findings highlight the superior accuracy of GBM, making it a more effective model for short-term forecasting of the JSE ALSI closing price. The data are stored in an Excel file.

Specifications Table

Subject area	<i>Short-Term Forecasting</i>
More specific subject area	<i>Financial Market Modelling and Forecasting</i>
Type of data	<i>Excel file</i>
How data was acquired	<i>Provided and from the internet</i> https://www.investing.com/
Data format	<i>Filtered and analysed.</i>
Experimental factors	<i>N/A</i>
Experimental features	<i>N/A</i>
Data source location	<i>Wall Street Journal Markets, FTSE/JSE Top 40 and S&P500 index Data portal webpage</i>
Data accessibility	<i>The data is hosted on GitHub https://github.com/csigauke</i>
Related research article	<i>Short-Term Forecasting of the JSE All Share Index Using Gradient Boosting Machines</i>

Value of the Data

The data can be used for financial time series forecasting in cross-sectional, temporal, or cross-temporal frameworks.

Data

The data used in this study is from Wall Street Journal Markets, FTSE/JSE Top 40 and S&P 500 index, and it can be accessed from <https://www.investing.com/>. The data comprises Date, Close, Day, Month, diff1, diff2, diff5, Oilprice, Goldprice, Platprice, SandP and UsdZar.

Experimental Design, Materials, and Methods

Data used in the study is from Wall Street Journal Markets, FTSE/JSE Top 40 and S&P 500 index.

Acknowledgements

The authors acknowledge Wall Street Journal Markets, FTSE/JSE Top 40 and S&P 500 index for providing the data.

Funding sources

The 2024 NRF Honours Postgraduate Scholarship support towards this research is hereby acknowledged.

References

Wall Street Journal Markets website

<https://www.wsj.com/marketdata/quotes/index/ZA/XJSE/ALSH/historical-prices> (accessed on 7 October 2024).

The FTSE/JSE Top 40 from <https://za.investing.com/indices/ftse-jse-top-40-historical-data> (accessed on 7 October 2024).

The S&P 500 index data from <https://www.wsj.com/market-data/quotes/index/SPX/historical-prices> (accessed on 7 October 2024).

Gold <https://www.investing.com/commodities/gold-historical-data> (accessed on 10 July 2025).

Platinum <https://www.investing.com/commodities/platinum-historical-data> (accessed on 10 July 2025).

Crude Oil (WTI/Brent) <https://www.investing.com/commodities/crude-oil-historical-data> (accessed on 10 July 2025).

USD/ZAR <https://www.investing.com/currencies/usd-zar> (accessed on 10 July 2025).