
Stock Market

Australian Mining Stock Market Analysis

Anuja
Daniel
Maria Florencia
Ohana

Project Overview

Objective:

This project aims to analyze and predict the stock performance of mining companies in Australia using various machine learning methods to ensure optimal results. The Australian stock market, known for its potential for high returns and inherent risks, necessitates a rigorous analysis to make informed investment decisions. Considering the pivotal role the mining sector plays in the Australian economy, our research endeavors to provide critical insights into market dynamics and emerging trends.

Methodology:

Our analysis uses a combination of statistical modeling and machine learning. By employing methods LSTM for time series forecasting and Random Forest Regression for prediction, we aim to capture the nuances and complexities of stock movements. Additionally, we utilize technical indicators such as the Relative Strength Index, Simple Moving Average, and Standard Deviation.

Targeted Company:

BHP

Machine Learning Workflow

Data Collection & Exploration

Feature selection

Data Visualization

Split the data into Train, Test and Validation sets

Implementing model prediction and evaluation

Data Collection

Data Collection

To obtain the essential market data for our stock prediction model, we will employ the yFinance library in Python. This library is made for fetching pertinent data for any given ticker symbol from the Yahoo Finance website. The yFinance library allows us to seamlessly acquire the most recent market data and integrate it into our model.

Yfinance

- Importance of Financial Data Analysis
- Informed Investment Decisions
- What is yfinance?
- Retrieving Historical Stock Price Data
- Yahoo Finance + Python
- Easy-to-Use Interface
- Integration with Pandas and Matplotlib
- Access to Wide Range of Financial Data

Features Analysed

First Part

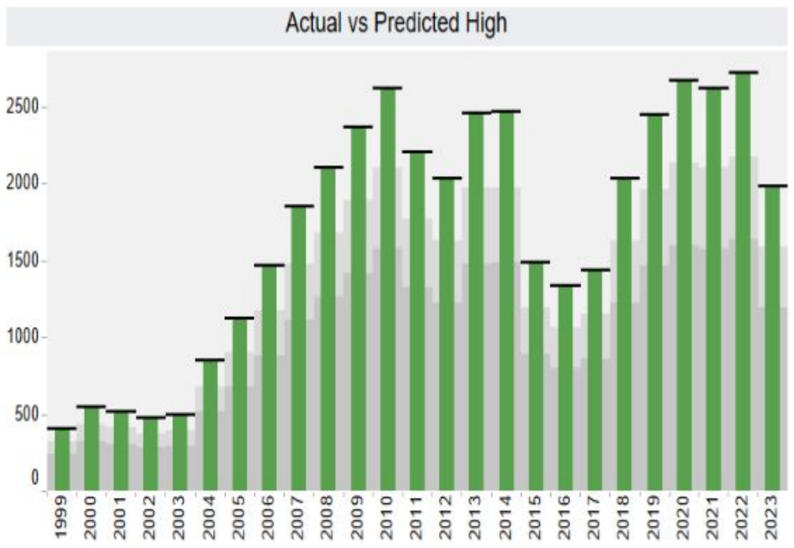
Date : Calendar date of the trading day.
Open : Opening price of the trading day.
High : Highest price of the stock traded during the day.
Low : Lowest price of the stock traded during the day.
Close : Closing price of the trading day.
Adj Close : Adjusted closing price of the trading day.
Volume : Number of shares traded in exchange during the day.

Second Part

In addition to the features we have used in the first part we have added these technical indicators to improve the model performance:

- RSI measures momentum
- SMA calculates trend,
- Standard Deviation quantifies volatility.

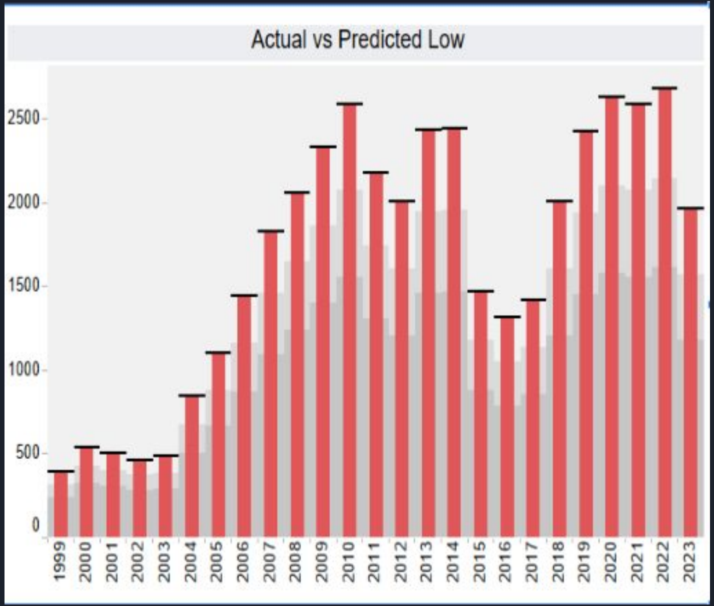
Data Visualization & Method



Both visualisations, the actual and predicted 'High' and 'Low' prices closely aligned, showing minimal differences.

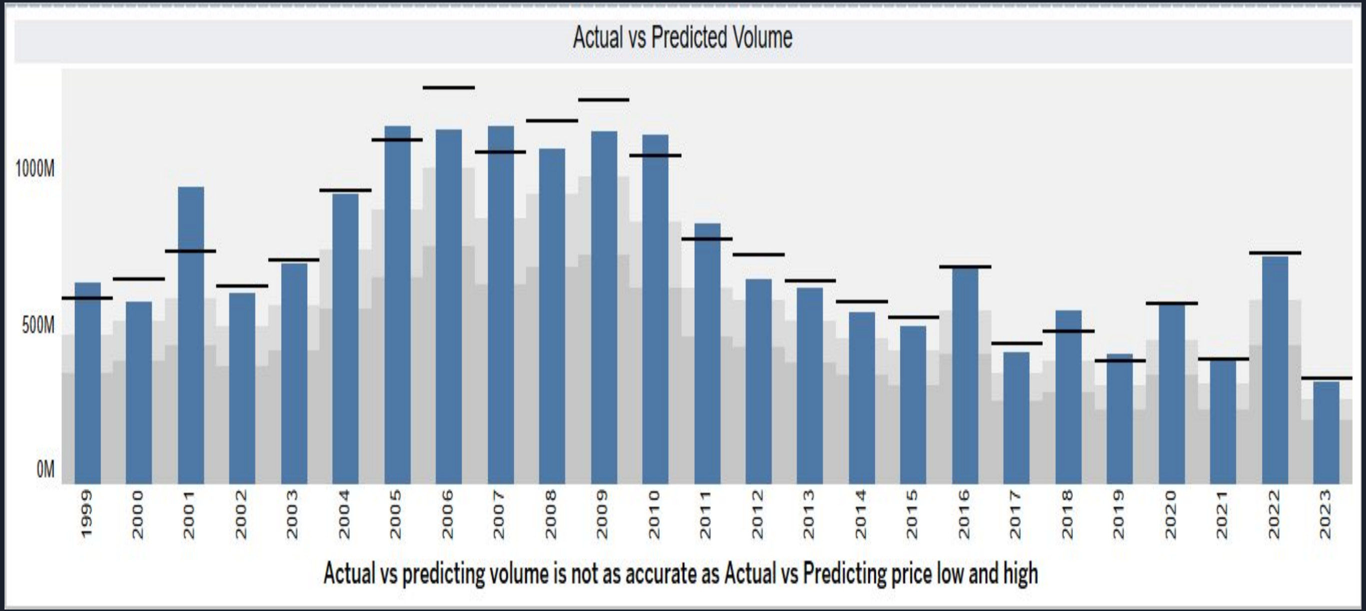
- Selected features: "Open", "High", "Low", "Close", and "Adj Close".
- To forecast the stock's 'High' price, the data was split into training sets (for model creation) and testing sets (for evaluation).
- To standardize the data, the Random Forest Regressor model was used. After training it with the known 'High' prices, the model predicted 'High' prices for the test data
- On the final visualization, both the actual and predicted 'High' prices closely aligned, showing minimal differences.

Data Visualization & Method

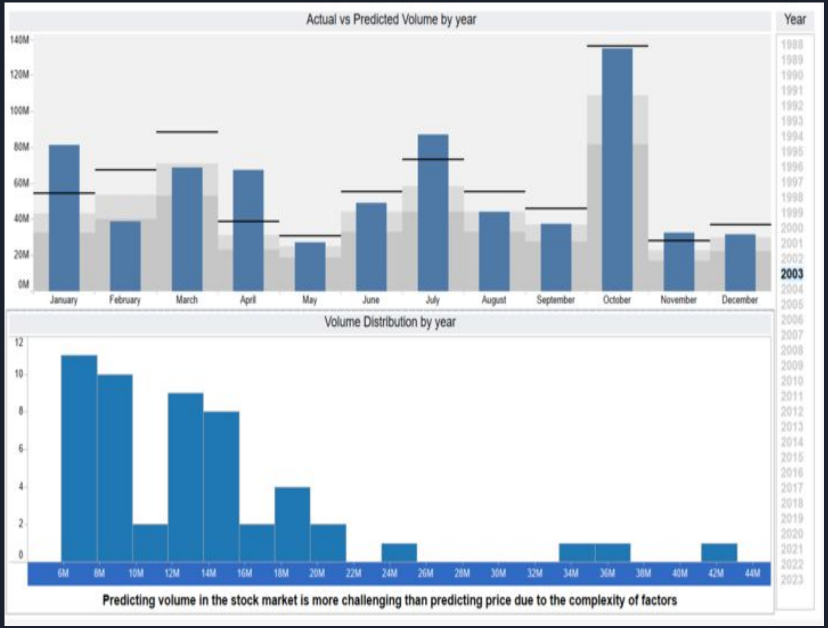


- Selected features: "Open", "High", "Low", "Close", and "Adj Close".
- To forecast the stock's 'Low' price, the data was split into training sets (for model creation) and testing sets (for evaluation).
- To standardize the data, the Random Forest Regressor model was used. After training it with the known Low prices, the model predicted Low prices for the test data
- On the final visualization, both the actual and predicted Low prices closely aligned, showing minimal differences.

Data Visualization

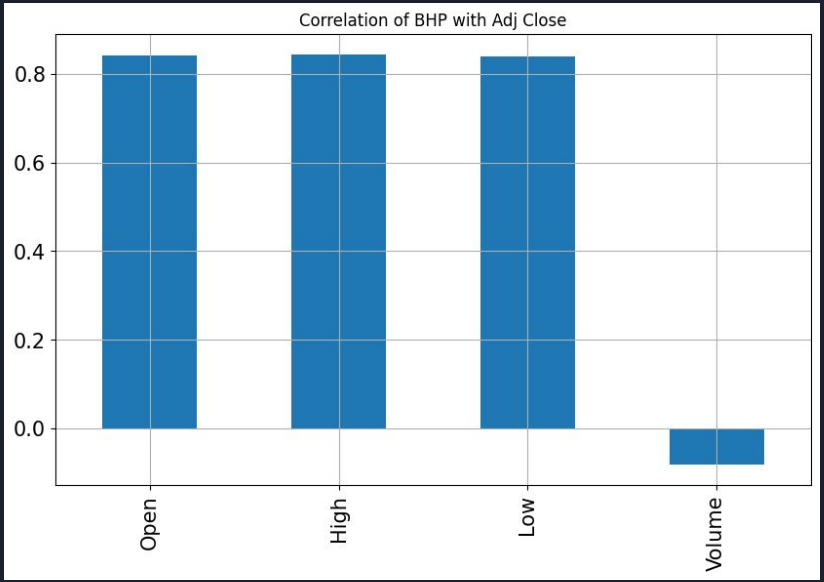


Data Visualization



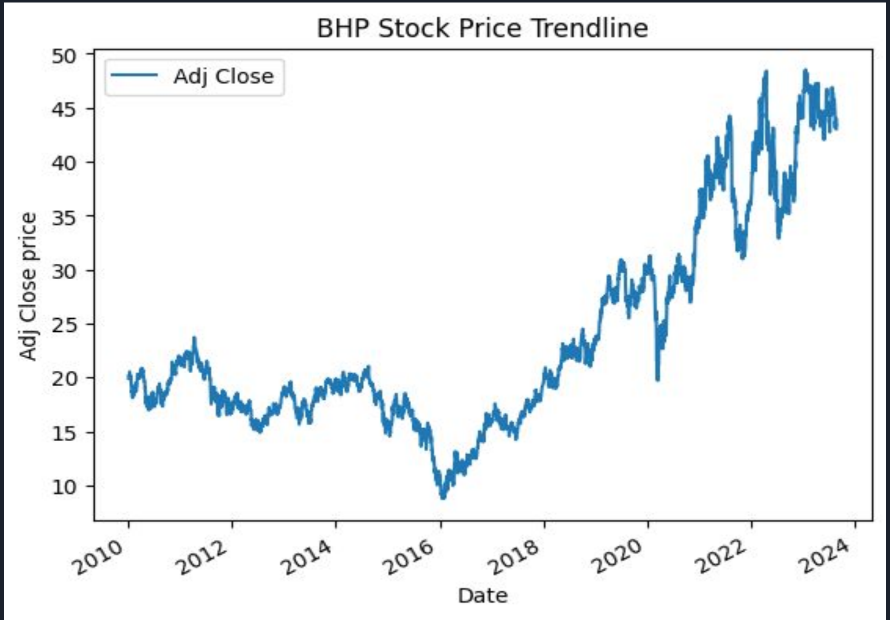
- Selected Features: "Open", "High", "Low", "Close", and "Adj Close".
- Target: Volume, which represents shares traded.
- Split the dataset into training and testing sets.
- Standardized features using StandardScaler to achieve zero-mean and unit-variance.
- Modeling was initialized and trained using a RandomForestRegressor.

Correlation Analysis



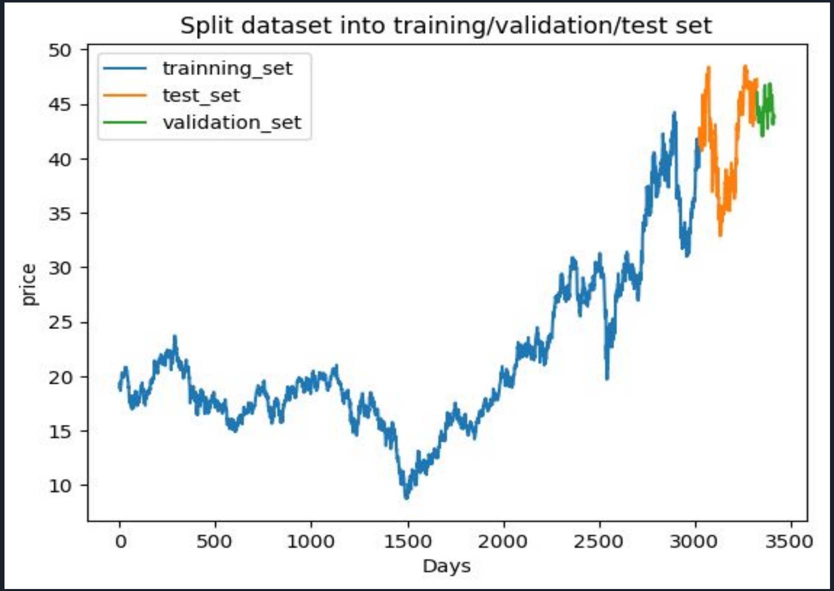
Open, High, and Low prices have a strong correlation with the Adjusted Close price, indicating they're reliable for predicting its value. On the other hand, while trading volume might reflect market sentiment, it doesn't closely relate to the Adjusted Close, suggesting it's not a main predictor.

Data Exploration



By focusing on the Adjusted Close Price and plotting its values over time, we can observe the stock's trajectory. The plot confirms that the BHP stock's adjusted closing price has been on an upward trend.

Split dataset into training, validation, and test



We're employing the Time Series Split to handle and validate our time series data. Our primary training dataset consists of 3,025 entries, each characterized by 7 unique features. This data will serve as the foundation for training our model. For evaluation purposes, a distinct testing dataset of 302 entries has been set aside. Before this final evaluation, we'll adjust our model using a validation set comprising 89 entries. Every data entry, across all these datasets, corresponds to a specific outcome we're trying to predict.

Models Explored

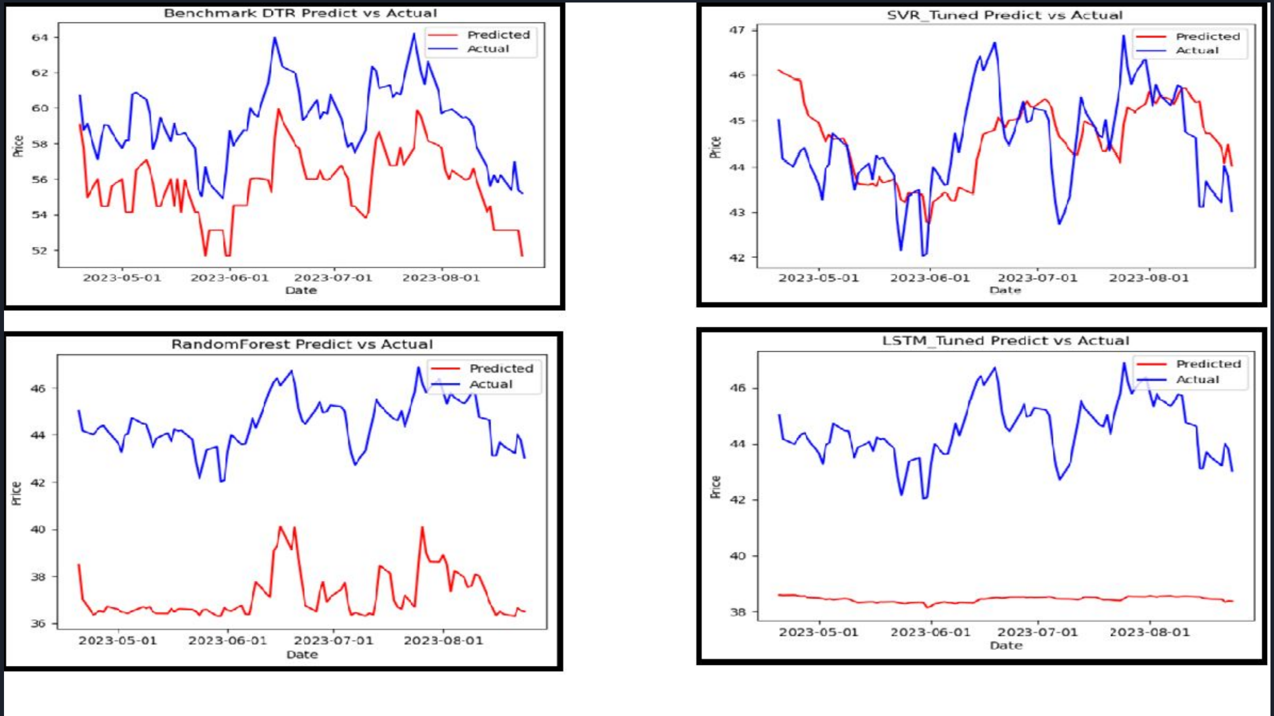
We have explored 4 models as below:

1. Decision Tree Regressor: It models decisions based on splitting data into branches, where each branch represents a decision path. It's simple and can help us understand what's driving a certain outcome.
2. SVR (Support Vector Regression): It works by finding the best-fitting line that comes as close as possible to our data points, allowing for a little bit of wiggle room. It's good for figuring out trends in data, even if they're not a simple straight line.
3. LSTM (Long Short-Term Memory): a smart system that remembers information from the past to predict what comes next. It's great for handling things like predicting future values in time-based data (like stock prices) or understanding patterns in text.
4. Random Forest: is an ensemble learning algorithm that combines multiple decision trees to make more accurate predictions. It's like asking a bunch of experts for their opinions and then making a decision based on what they all say. By listening to different viewpoints and putting them together, we can often predict things more accurately.

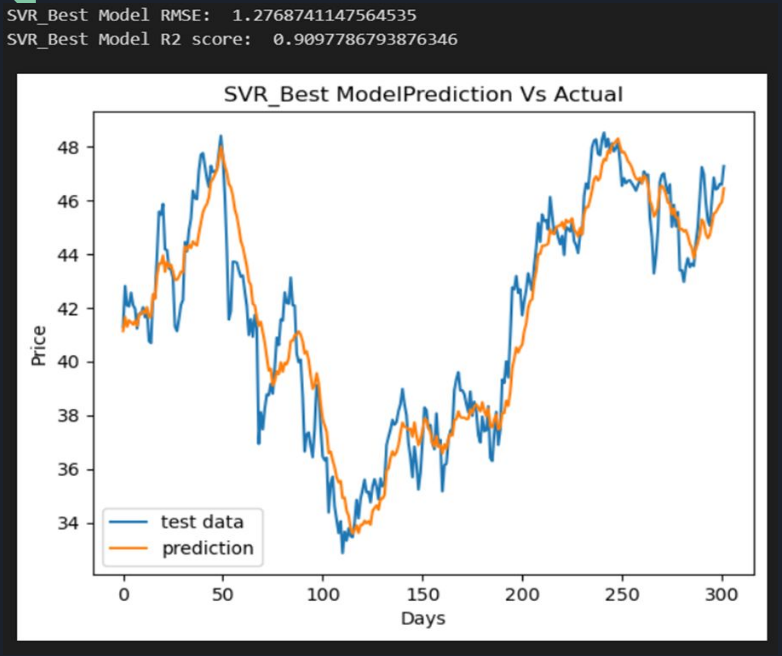
We have used graphs along with RMSE and R2 scores to judge the model performance.
RMSE tells us how spread out the errors are between predicted and actual values. Smaller RMSE means the predictions are closer to reality.
R2 shows how well our model fits the data. A higher R2 means the model explains more of the data's variation, indicating a better fit.

We will now explore different models and their performance in the next slides

Data Visualization

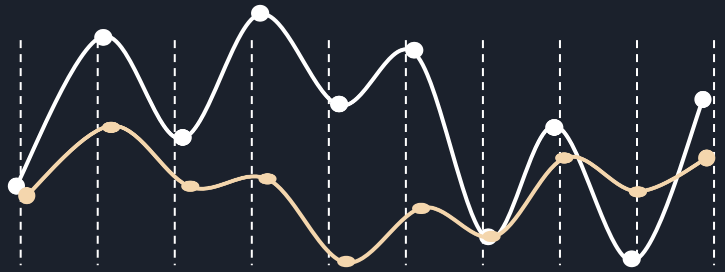


Optimal Model Selection



Based on the RSME Score and R2 Score, SVR model is the best performing model.
As our time was limited, we concluded our investigation at this point. However, additional research holds potential for enhancing model accuracy. For instance, incorporating additional technical indicators or delving deeper into time series analysis could contribute significantly to refining the model's performance.

Limitations of stock Prediction Model



The most foundational limitations is that past performance does not guarantee future results . Even if our model captures historical trends perfectly, unpredictable events can always affect stock prices. Sudden shifts in the economy, like recessions or booms can change the stock game .

Conclusion

Based on our label data and the features considered, the SVR emerged as the best model, especially when evaluating the RMSE and R2 scores.

While predictive models offer valuable insights into stock market trends, they should be viewed as one component of an investor's decision-making process. It's essential to integrate these model insights with expert guidance, thorough research, and an understanding of the broader market context.