
CAPSTONE PROJECT

STOCK MARKET PREDICTION

Presented By:

1. Aman Srivastava-IIT (ISM) Dhanbad-Mining Engineering

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

You have recently joined an Asset Management Firm as an Associate Analyst. One of the clients of your firm are keenly interested in Amazon, Domino's, Netflix and Bitcoin as an asset. But the financial markets are highly volatile and complex, making accurate prediction of stock prices and cryptocurrency values a challenging task. So, you have been tasked to develop a predictive model for forecasting the stock market values of Amazon (AMZ), Domino's Pizza Inc (DPZ), Netflix (NFLX), and the value of Bitcoin (BTC) using historical data.

The primary objective is to explore and analyze the historical data to identify patterns and trends that can be leveraged to predict future prices. The secondary objective is to evaluate the performance of different machine learning algorithms in predicting stock and cryptocurrency values and to determine which model provides the most accurate forecasts.

By addressing these challenges, your firm aims to provide valuable insights and tools for their clients to make informed decisions in the stock and cryptocurrency markets. The outcome of this task could potentially enhance investment strategies and improve financial decision-making processes.

PROPOSED SOLUTION

After we extract the data of stock market values of Amazon (AMZ), Domino's Pizza Inc (DPZ), Netflix (NFLX), and the value of Bitcoin (BTC) from May 2013-May2019 and store it in a CSV file. This Time Series data then goes through various algorithms trying to predict the value. In our case we will be measuring the accuracy of these predictions by **Symmetric mean absolute percentage error (SMAPE)**, that is given by, the absolute difference between actual value and predicted value is divided by half the sum of absolute actual value and predicted value, and then average all such values across all the fitted points:

$$\text{SMAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

where A_t is the actual value and F_t is the forecast value.

The prediction model with minimum SMAPE value is the most fit algorithm. The models can then be tuned to better predict the values. Enhancements like Hyperparameter optimization, feature engineering, etc.

SYSTEM APPROACH

Time Series data is a series of data points indexed in time order. Time series data is everywhere, so predicting these data can be done using various algorithms with varying degree of accuracy. This process of selecting the correct algorithm with highest accuracy can often be time consuming for the analyst hence we will be using the AutoAI feature of Watson Studio provided by IBM Cloud.



ALGORITHM & DEPLOYMENT

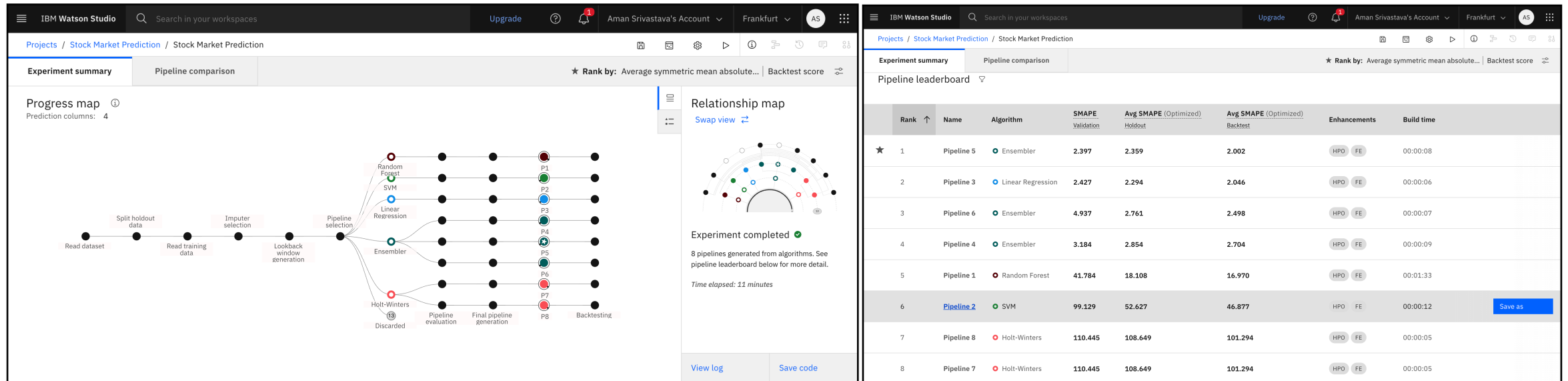
The screenshot shows the IBM Watson Studio interface for configuring an AutoAI experiment titled "Stock Market Prediction". The top navigation bar includes the IBM Watson Studio logo, a search bar, an "Upgrade" button, a help icon, a notification bell with a red "1", the user's account "Aman Srivastava's Account", the location "Frankfurt", and a profile icon "AS". The breadcrumb trail is "Projects / Stock Market Prediction / Stock Market Prediction".

The main content area is divided into two panels:

- Add data source:** This panel on the left contains a dashed box with the text "Add files such as tabular data (CSV)." and two buttons: "Browse" and "Select data from project". Below this, a file named "portfolio_data.csv" is listed with a size of "79.39 KB" and "Columns: 5".
- Configure details:** This panel on the right is titled "What do you want to predict?". It includes a "Prediction columns" dropdown menu showing "4 x AMZN, DPZ, BTC, NFLX" and a "Date/time column (optional)" dropdown menu showing "Date". Below these, it displays "Prediction columns: AMZN, DPZ, BTC, NFLX" and "CUH remaining: 11.92 CUH". The "PREDICTION TYPE" is set to "Time Series Forecast", and it is "OPTIMIZED FOR Avg SMAPE". At the bottom, there is a "Run experiment" button.

- i. The dataset was uploaded on Watson Studio.
- ii. AutoAI was configured to predict the values of AMZN, DPZ, BTC and NFLX.

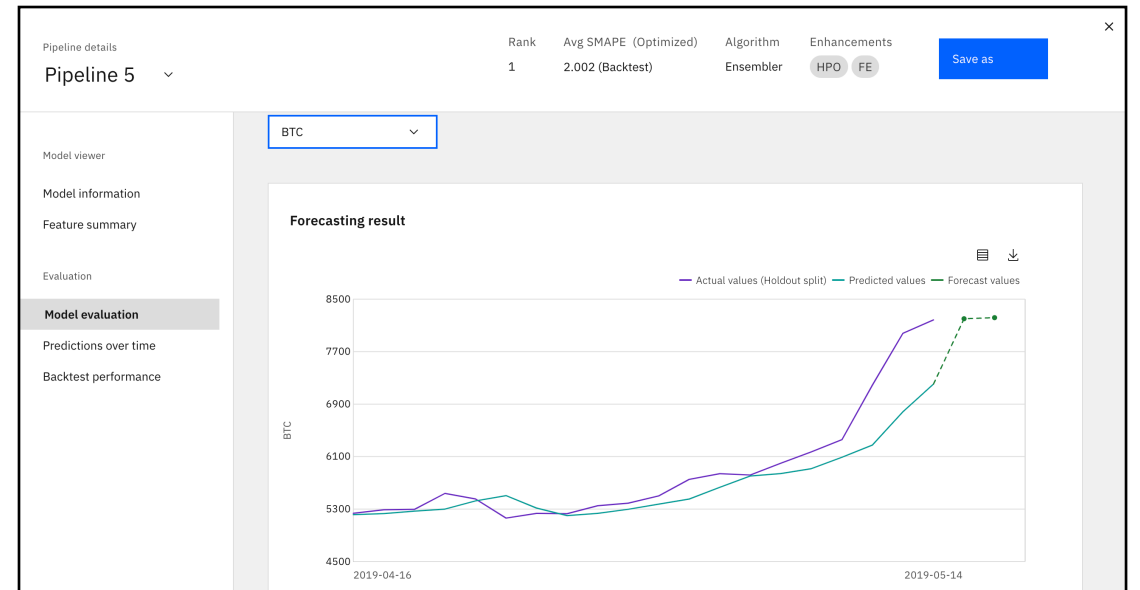
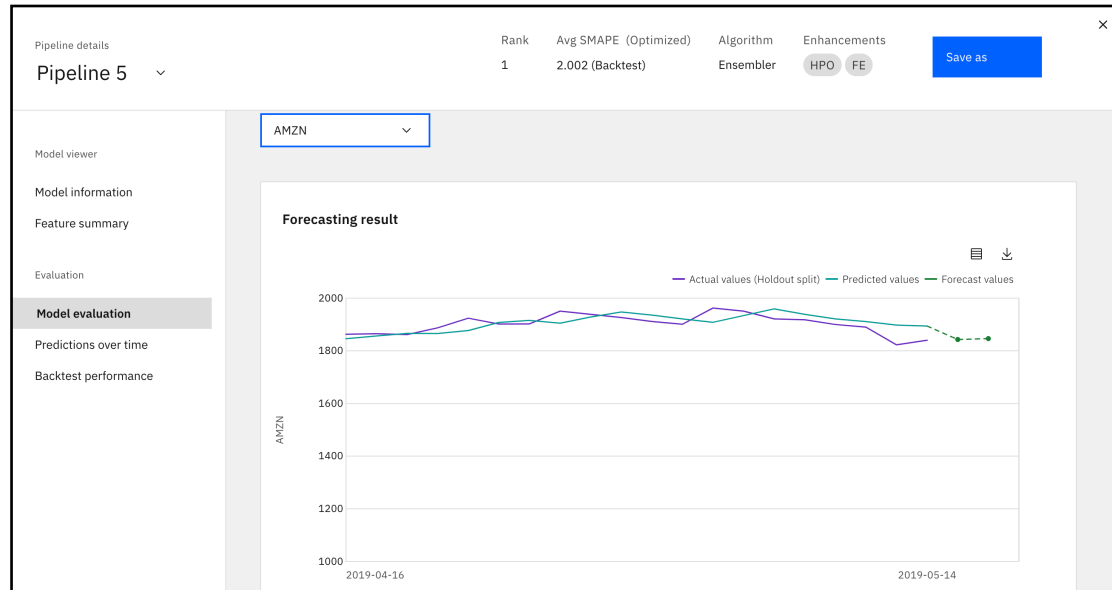
ALGORITHM & DEPLOYMENT



- i. AutoAI goes through various algorithms and enhancements to and finalises 8 pipelines to treat the dataset.
- ii. AutoAI ranks these pipelines on minimum SMAPE
- iii. The code for these algorithms can be saved directly in .ipynb format

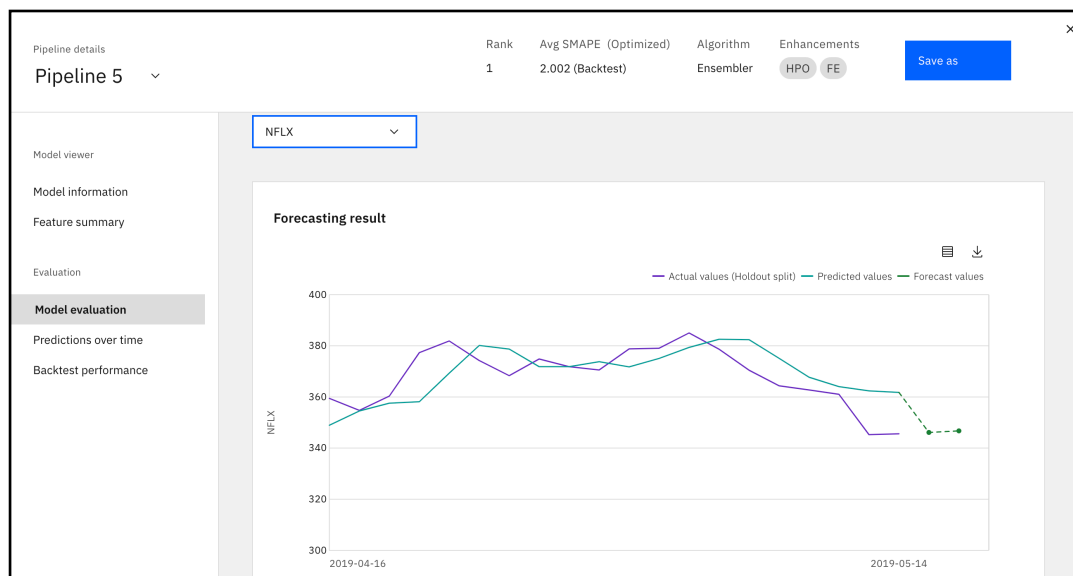
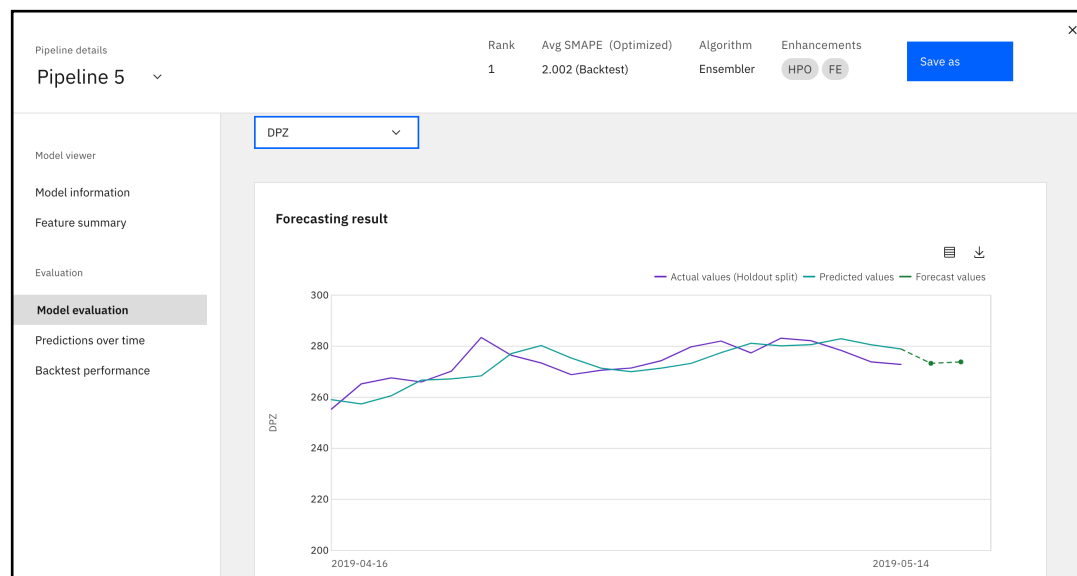
<https://github.com/aman-sri90/IBM-Cloud/tree/main>

RESULT



Forecast for Amazon and Bitcoin using ensembler algorithm and Hyper Parameter Optimisation and Feature Engineering Enhancements

RESULT



Forecast for Domino's and Netflix using ensembler algorithm and Hyper Parameter Optimisation and Feature Engineering Enhancements

CONCLUSION

Through our analysis, we demonstrated the potential of machine learning to enhance investment strategies and improve financial decision-making processes for our clients. The predictive models developed in this project offer valuable insights and tools, allowing clients to make more informed decisions in the highly volatile stock and cryptocurrency markets.

By leveraging the AutoAI feature of IBM's Watson Studio, we were able to automate the process of selecting and tuning machine learning algorithms, thereby efficiently identifying models with the highest predictive accuracy.

The application of AutoAI facilitated the analysis of time series data, enabling us to uncover patterns and trends that significantly impact future price movements. This approach not only reduced the time and effort required for model selection and optimization but also provided a robust framework for analyzing complex financial data.

Overall, the project underscores the importance of leveraging advanced technologies and data-driven approaches in asset management, providing a competitive edge in navigating the complexities of financial markets.

FUTURE SCOPE

This project lays a strong foundation for utilizing machine learning in financial forecasting, and several opportunities exist to expand and enhance the work further:

1. **Incorporating Additional Data Sources:** Future work could include integrating more diverse data sources, such as social media sentiment analysis, macroeconomic indicators, and news analytics, to improve prediction accuracy and provide a more comprehensive view of market dynamics.
2. **Real-Time Predictions:** Developing a system for real-time data collection and prediction would enable continuous monitoring and forecasting of stock and cryptocurrency values, offering clients timely insights and decision-making support.
3. **Expanding the Asset Portfolio:** Including additional stocks, commodities, and cryptocurrencies in the analysis could diversify the investment portfolio and provide broader insights across various market segments.
4. **Enhanced Algorithm Selection:** Exploring and experimenting with a wider range of machine learning and deep learning algorithms, including neural networks and ensemble methods, could further refine the predictive models and enhance their performance.
5. **Risk Assessment and Management:** Incorporating risk analysis tools and methodologies to assess potential investment risks and develop strategies for mitigating them could add significant value to the predictive models.
6. **User-Friendly Interfaces and Dashboards:** Developing interactive dashboards and visualization tools would enable clients to easily interpret and act on the insights generated by the predictive models, making the system more accessible and user-friendly.

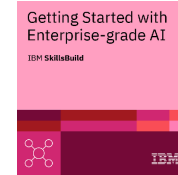
By exploring these future directions, the project can continue to evolve and adapt to the changing landscape of financial markets, ensuring that clients receive the most relevant and actionable insights for their investment strategies.

REFERENCES

- <https://towardsdatascience.com/supervised-machine-learning-feature-engineering-and-hyper-parameter-tuning-a3da583dd7b9>
- https://en.wikipedia.org/wiki/Symmetric_mean_absolute_percentage_error
- Dataset: <https://www.kaggle.com/datasets/hershyandrew/amzn-dpz-btc-ntfx-adjusted-may-2013may2019/data>

CERTIFICATE1

In recognition of the commitment to achieve
professional excellence



Aman Srivastava

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI



Issued on: 09 JUL 2024

Issued by IBM

Verify: <https://www.credly.com/go/pFfd5JGv>



CETIFICATE 2

In recognition of the commitment to achieve
professional excellence



Aman Srivastava

Has successfully satisfied the requirements for:

Cloud Computing Fundamentals



Issued on: 29 JUL 2024

Issued by IBM

Verify: <https://www.credly.com/go/WYfa0KpF>





THANK YOU