The influence of stock market indexes (S&P500 and Dow Jones) on cryptocurrencies prices

Gorast Angelovski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
gorast.angelovski@students.finki.ukim.mk

Ana Todorovska
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
ana.todorovska@finki.ukim.mk

Ivan Rusevski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
ivan.rusevski@students.finki.ukim.mk

Jovana Marojevikj
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
jovana.marojevikj@students.finki.ukim.mk

Eva Spirovska
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
eva.spirovska@students.finki.ukim.mk

Hristijan Peshov
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
hristijan.peshov@students.finki.ukim.mk

Irena Vodenska
Laboratory for Interdisciplinary Finance and Economics
(LIFE) Research, Administrative Sciences Department,
Financial Management
Metropolitan College, Boston University
Boston, USA
vodenska@bu.edu

Lubomir T. Chitkushev
Laboratory for Interdisciplinary Finance and Economics
(LIFE) Research, Computer Science Department
Metropolitan College, Boston University
Boston, USA
ltc@bu.edu

Dimitar Trajanov
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
dimitar.trajanov@finki.ukim.mk

***Abstract*—In this paper we analyze openly available time series data for the prices of 18 cryptocurrencies and 2 stock market indexes (S&P500 and Dow Jones).**
First, we calculate the correlation values between the cryptocurrencies and indexes datasets.
Then, we use a state of the art time series prediction library (XGBoost) in order to make prediction models for the daily prices of all the cryptocurrencies, using the stock market index datasets as input features in the training model.
We calculate metrics for the difference between the actual prices and the prices predicted using our models.
Finally, we show the feature importance score that our model attributed to each prediction model, and compare the score between the three input features (S&P500 dataset, Dow Jones dataset, and the actual cryptocurrency dataset).

***Index Terms*—cryptocurrencies, stock market indexes, correlations, RMSE, feature importance**

I. Introduction

Cryptocurrencies have become a global phenomenon. With the rise of digitalisation in recent years, cryptocurrencies introduced a novel medium, and their market is growing rapidly and breaking records by the day. However, this market is also categorized as highly volatile, meaning that it is subjected to rapid changes in price and demand, sometimes overnight. Some of these changes are attributed to isolated events which have the ability to trigger an effect of mass speculation, usually through social media such as Reddit and Twitter. This presents investors and traders with a major challenge when it comes to analyzing and predicting the cryptocurrency market.

On the other hand, there are the major stock indexes, such as the S&P500 and DowJones, which are the most popular index funds and are the so called 'standard' for stock investing since the 1900s. In the field of economics these indexes, among others, are considered as classical economic indicators - implying that through them we can track and monitor the overall status of the economy.

In this paper we are investigating the relationship between these two US stock market indexes and the price of cryptocurrencies. We are using time series analysis in order to try and predict the fluctuations in the prices of cryptocurrencies, in correlation with the price levels of the S&P500 and DowJones. Our idea is to find a correlation between the price movements of these two indexes and some of the most popular cryptocurrencies, and thus predict the general trend in which the cryptocurrency market is going.

II. S&P500 and Dow Jones indexes

The S&P 500 Index, or Standard & Poor's 500 Index, is a market-capitalization-weighted index of 500 leading publicly traded companies in the U.S. It is not an exact list of the top 500 U.S. companies by market cap because there are other criteria that the index includes. Still, the S&P 500 index is regarded as one of the best gauges of prominent American equities' performance, and by extension, that of the stock market overall.

The Dow Jones Industrial Average (DJIA), also known as the Dow 30, is a stock market index that tracks 30 large, publicly-owned blue-chip companies trading on the New York Stock Exchange (NYSE) and the Nasdaq. The Dow Jones is named after Charles Dow, who created the index in 1896 along with his business partner Edward Jones.

III. Data

In this study the structured data we use for time series forecasting is publicly available.

A. Cryptocurrency prices

We collect historical prices for eighteen cryptocurrencies from Yahoo Finance for the period between March 2019 and March 2021.
For each cryptocurrency, the historical prices dataset contains a date, open price, high and low prices, close price, adjusted closing price, and volume.
The eighteen cryptocurrencies we picked based on popularity are the following: Bitcoin, Monero, Cardano, Stellar, Ripple, Iota, Chainlink, Litecoin, Ethereum, Neo, Dogecoin, Celsius, Dash, EOS, Nano, NEM, Maker, and VeChain.

B. Stock Market Indexes

We collect historical prices for the two biggest stock market indexes in the US, The S&P500 and The Dow Jones Industrial Average from Yahoo Finance for the period between March 2019 and March 2021.
For each stock market index, the historical prices dataset contains a date, open price, high and low prices, close price, adjusted closing price, and volume.

IV. Methodology

The methodology of this research consists of four phases, which are: data preprocessing, finding correlation between the indexes (S&P500 and DowJones) and cryptocurrencies, prediction of cryptocurrency prices, and analyzing the feature importance of the resulting time series.
The steps will be further analyzed in the text below.

A. Data preprocessing

The data used in this research includes the prices of the 18 most popular cryptocurrencies and 2 US stock market indexes (DowJones and S&P 500) in the date range from 01/03/2019 to 01/03/2021. Since the US stock market does not operate during weekends and holidays, these days are filtered out.

B. Correlation between stock market indexes and cryptocurrency prices

For the purpose of finding relations and patterns between the prices of the cryptocurrencies and indexes, it is relevant to measure the pairwise correlation of the time series data for the prices.
The two cases observed are: Correlation between S&P500 and each cryptocurrency, and correlation between DowJones and each cryptocurrency.
The pairwise correlation of DowJones daily prices and S&P500 daily prices is 0.937, meaning the prices are following the same trend.
This implies that their corresponding correlations to the cryptocurrencies will also be similar.

C. Time series predictions

For the purpose of predicting the daily prices, firstly the data is split such that 80 percent of it is used for training, and the remaining 20 is used for testing. Then, the data is transformed using a logarithimic function, as it is an effective method of removing exponential variance with time series data.
XGBoost is used as a model for prediction, as it is a state of the art model that we found shows the best results for time series forecasting in our previous research.
In order to get the optimal parameters for training the model, GridSearchCV is used. After the predictions are made, the data is transformed back to the original format using inverse log. As a accuracy metric, RMSE is used.

D. Feature importance

As for the feature importance, we calculate the F-score function, in order to quantify the extent to which a feature has contributed to the model's output.

V. Results

A. Correlations

All of the resulting correlations fall in the range between 0.10 and 0.24, except for one outlier with a correlation of 0.00 with both indexes - Nem. This indicates a weak correlation between the prices of the cryptocurrencies and the stock market indexes.
The S&P 500 has an average correlation of 0.175, and the Dow Jones has an average correlation of 1.40

B. RMSE between predicted and actual prices

The RMSE represents the difference between predicted and actual values measured in USD.
The results vary greatly due to the inconsistent pricing of cryptocurrencies.

C. Feature importance using F-score

The highest feature importance is most often attributed to the price of the cryptocurrency itself - this occurs in 9 out of 18 cases.
The next highest feature importance is attributed to the price of the S&P 500 - in 8 out of 18 cases.
In 16 out of 18 cases, the S&P 500 has a higher feature

TABLE I
CORRELATIONS

Cryptocurrency	S&P500_corr	DowJones_corr
Bitcoin	0.17	0.13
Monero	0.19	0.14
Cardano	0.21	0.17
Stellar	0.21	0.18
Ripple	0.14	0.11
Iota	0.23	0.19
Chainlink	0.20	0.17
Litecoin	0.18	0.14
Ethereum	0.22	0.18
Neo	0.19	0.18
Dogecoin	0.12	0.10
Celsius	0.13	0.09
Dash	0.14	0.11
Eos	0.17	0.13
Nano	0.23	0.19
Nem	0.00	0.00
Maker	0.20	0.14
VeChain	0.22	0.16

TABLE II
RMSE BETWEEN PREDICTED AND ACTUAL PRICES

Cryptocurrency	RMSE
Bitcoin	1248.4452
Monero	7.8988
Cardano	0.0184
Stellar	0.0214
Ripple	0.0478
Iota	0.0203
Chainlink	1.4736
Litecoin	7.7578
Ethereum	62.0107
Neo	1.2886
Dogecoin	0.0038
Celsius	0.3219
Dash	10.2230
Eos	0.1604
Nano	0.3105
Nem	0.0550
Maker	86.4395
VeChain	0.0018

importance compared to the Dow Jones.
The Dow Jones index has the lowest feature importance scores, the lowest F-score of the three features in 14 out of 18 cases.

VI. CONCLUSION

This research was conducted as an extention of a larger research topic that we are working on.
Our research presents an opportunity to understand the relationship between these different markets and utilize this knowledge as an investment asset.
Based on the results obtainen, we conclude that these indexes have an insignificant impact on the price of cryptocurrencies.
It could be implied that the impact of the S&P500 on the price of cryptocurrencies is arguably larger than that of the Dow Jones. However, more research is needed to in order to justify this conclusion.
In the future we intend to integrate the findings of this research with our other ongoing research, as well as answering some of the peculiarities of the model like why does the previous day price of some cryptocurrencies have the lowest F-score?

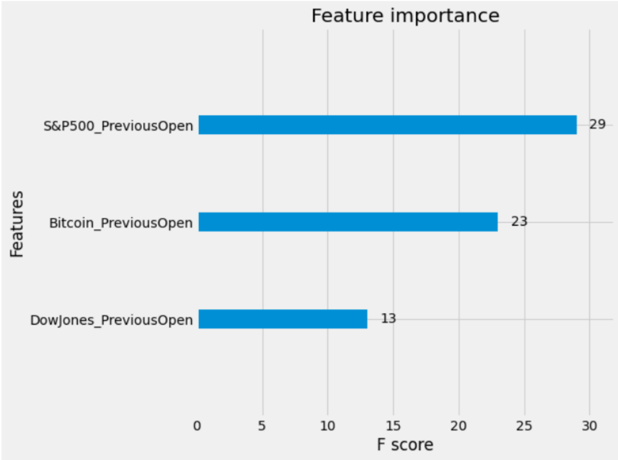

Fig. 1. Feature importance using F-score (Bitcoin)

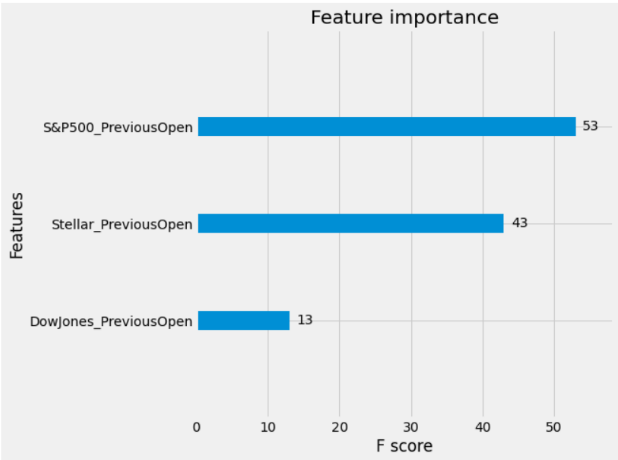

Fig. 2. Feature importance using F-score (Stellar)

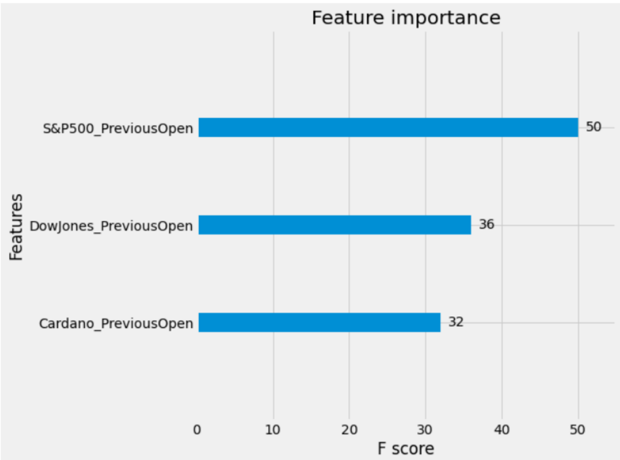

Fig. 3. Feature importance using F-score (Cardano)

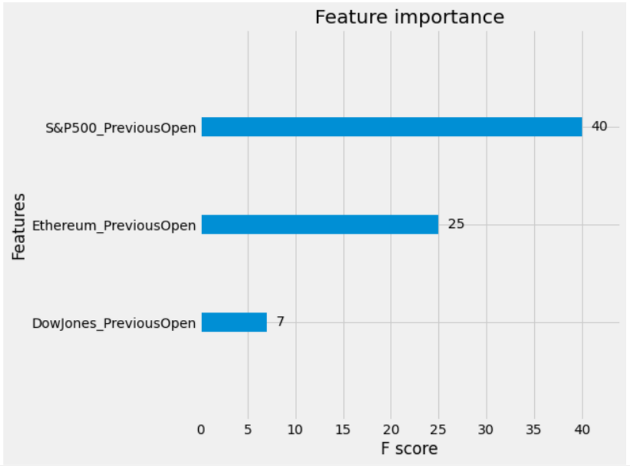

Fig. 4. Feature importance using F-score (Ethereum)

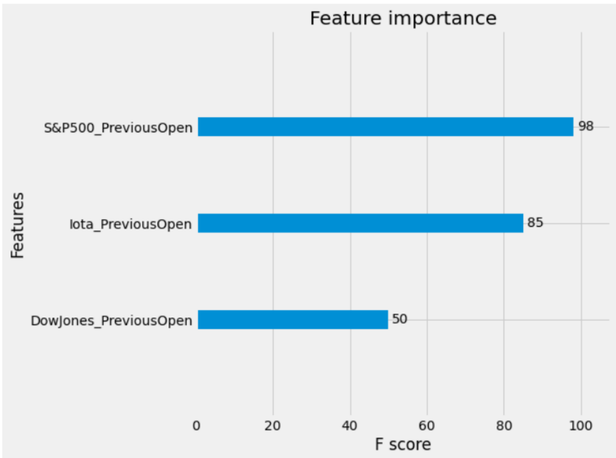

Fig. 5. Feature importance using F-score (Iota)

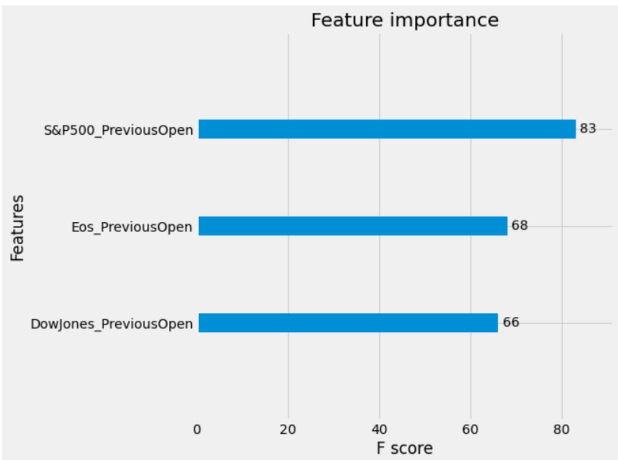

Fig. 6. Feature importance using F-score (Eos)

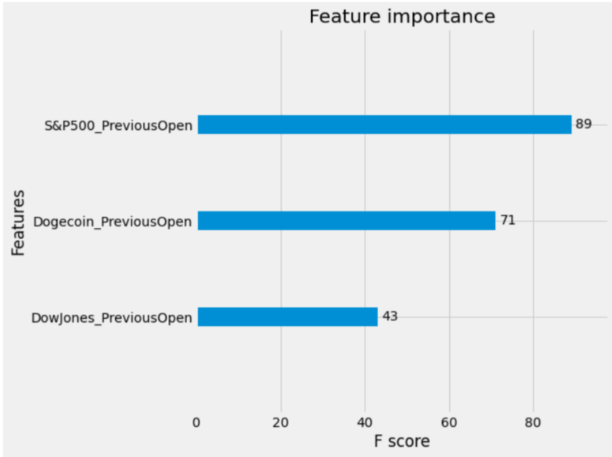

Fig. 7. Feature importance using F-score (Dogecoin)

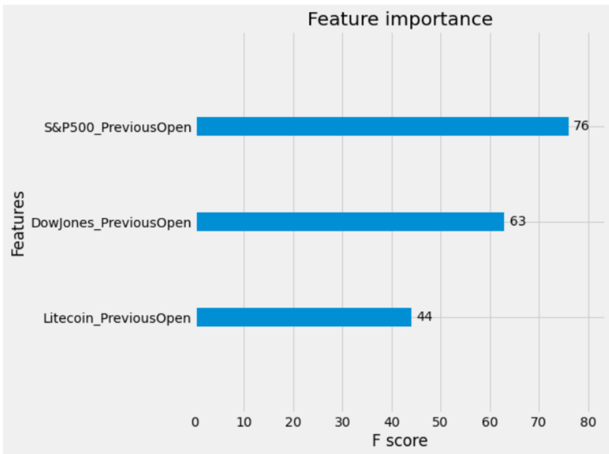

Fig. 8. Feature importance using F-score (Litecoin)

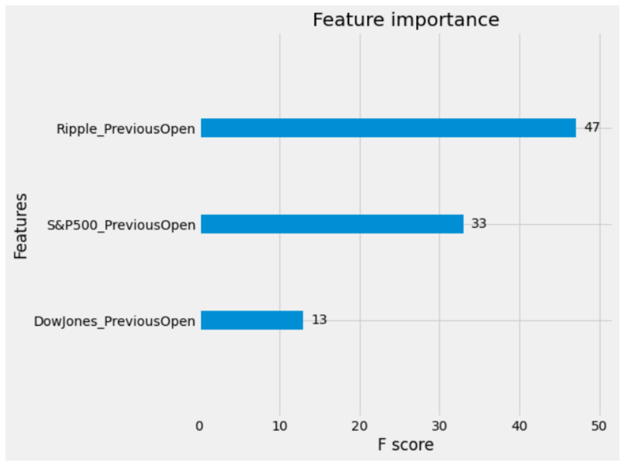

Fig. 9. Feature importance using F-score (Ripple)

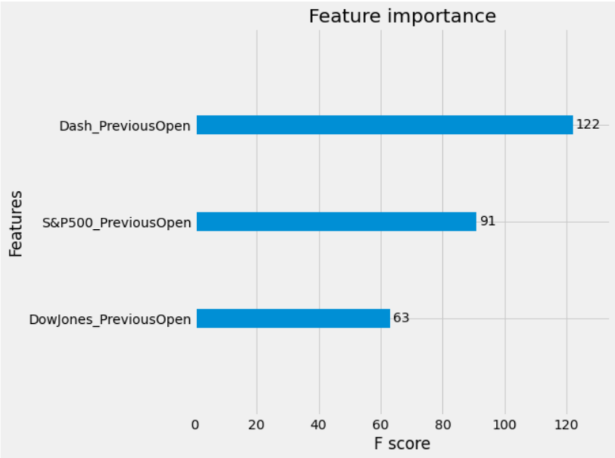

Fig. 10. Feature importance using F-score (Dash)

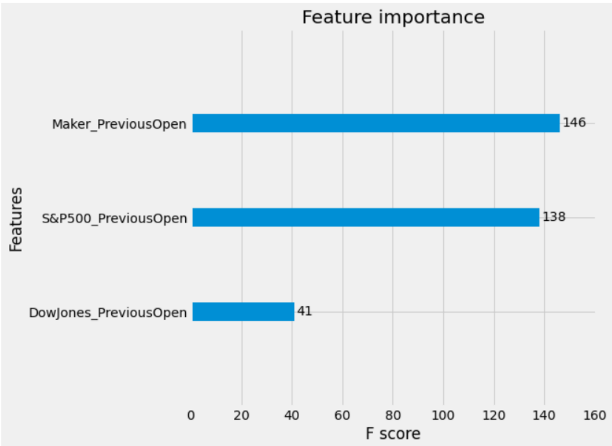

Fig. 11. Feature importance using F-score (Maker)

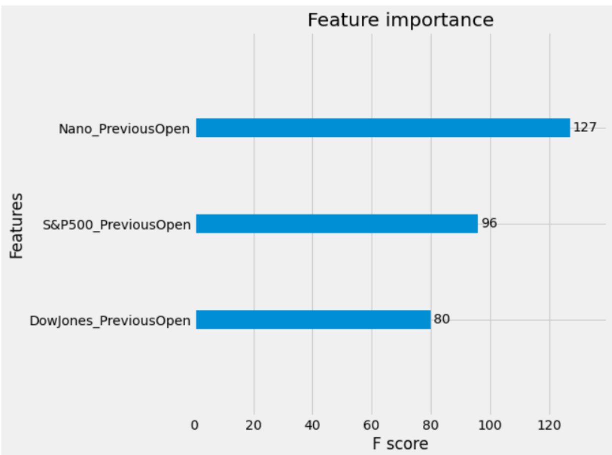

Fig. 12. Feature importance using F-score (Nano)

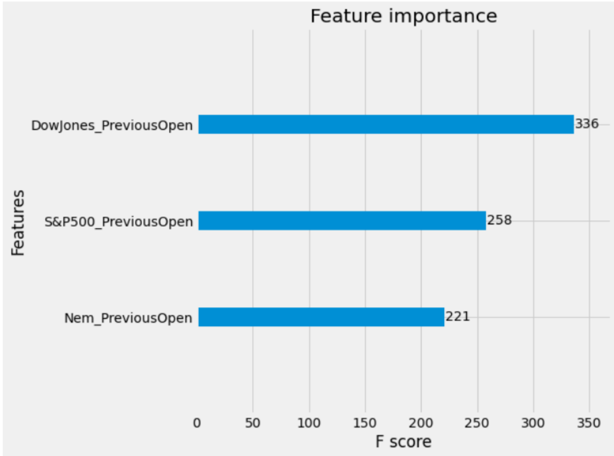

Fig. 13. Feature importance using F-score (Nem)

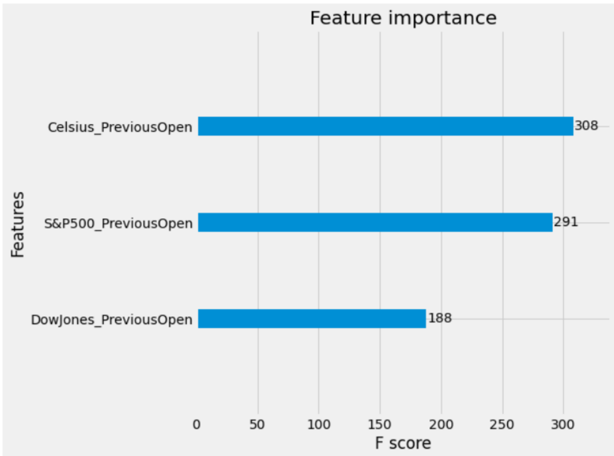

Fig. 14. Feature importance using F-score (Celsius)

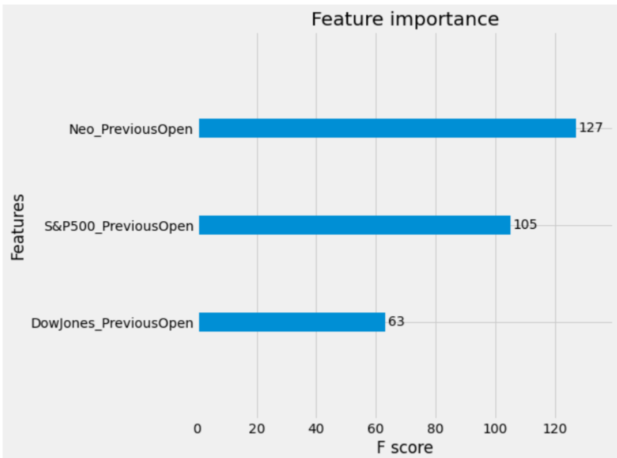

Fig. 15. Feature importance using F-score (Neo)

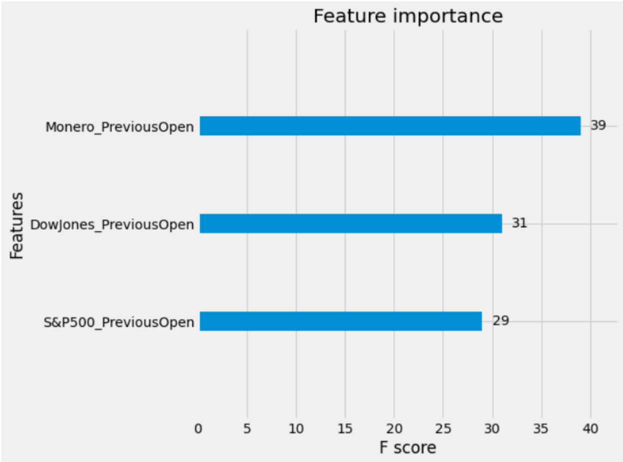

Fig. 16. Feature importance using F-score (Monero)

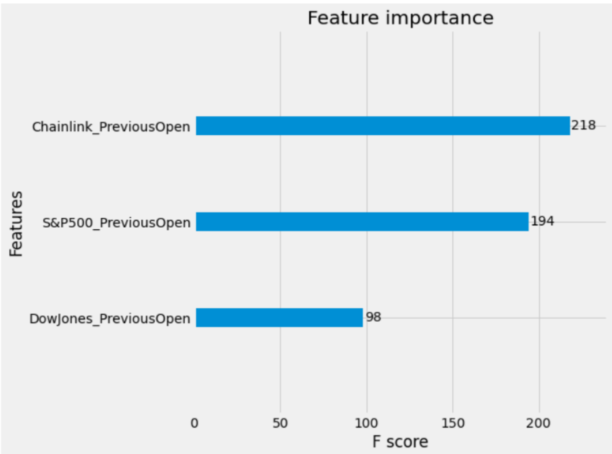

Fig. 17. Feature importance using F-score (Chainlink)

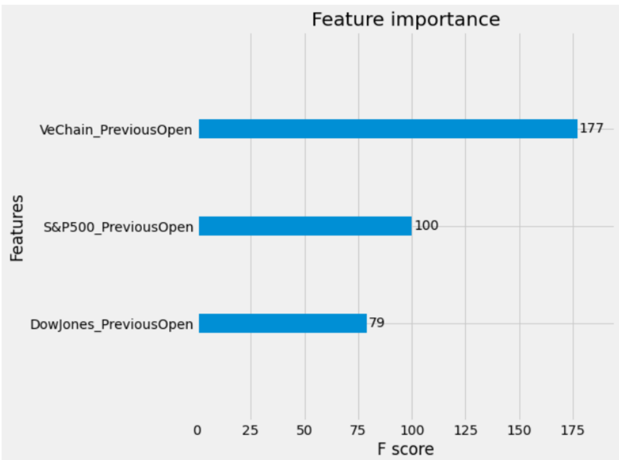

Fig. 18. Feature importance using F-score (VeChain)

REFERENCES

[1] Nakamoto Satoshi, “Bitcoin: A peer-to-peer electronic cash system.” Manubot, 2019.

[2] Chen, Tianqi and Guestrin, Carlos. ”Xgboost: A scalable tree boosting system” Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp.785-794, 2016.

[3] Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. ”Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation.” Australasian joint conference on artificial intelligence. Springer, Berlin, Heidelberg, 2006.

[4] Todorovska, Ana, Eva Spirovska, Gorast Angelovski, Hristijan Peshov, Ivan Rusevski, Jovana Marojevikj, Irena Vodenska, Lubomir T. Chitkushev, and Dimitar Trajanov, ”Analysis of cryptocurrency interdependencies.” In Proceedings of Blockchain in Kyoto 2021 (BCK21) (p. 011004).

[5] Lundberg, Scott M., and Su-In Lee. ”A unified approach to interpreting model predictions.” Advances in neural information processing systems 30 (2017).

[6] Jovan Davchev, Kostadin Mishev, Irena Vodenska, Lou Chitkushev, and Dimitar Trajanov. Bitcoin price prediction using transfer learning on financial micro-blogs. In The 16th Annual International Conference on Computer Science and Education in Computer Science, September 5, Bulgaria (virtual), 2020.

[7] Y. Jiao and J. Jakubowicz, ”Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks,” 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 4705-4713, doi: 10.1109/BigData.2017.8258518.

[8] Cortez, Alexandra, and Ana Tulcanaza. ”Bitcoin: its influence on the global World and its relationship with the stock exchange.” Revista Chakiñan de Ciencias Sociales y Humanidades 5 (2018): 54-72.