

ДОДАТОК А

Перелік джерел посилання за науковими напрямками керівника та науковців
кафедри програмної інженерії

17. Smelyakov, Kirill, Oleksandr Bizkrovnyi, Natalia Sharonova, Serhii Smelyakov and Anastasiya Chupryna. “Building of Regression Models for Cryptocurrency Price Prediction.” International Conference on Computational Linguistics and Intelligent Systems (2022)., URL: <https://ceur-ws.org/Vol-3171/paper90.pdf>

38. Gordiievych, A., Shubin, I. Forecasting of airfare prices using time series. 2015 Information Technologies in Innovation Business Conference, ITIB 2015 - Proceedings, 2015, pp. 68-71, 7355055.

39. Afanasieva I.V., Naumov A.B., Onyshchenko K.G. Bitcoin price prediction using the boosting algorithm. The 2nd International scientific and practical conference “Science and society: modern trends in a changing world” (January 22-24, 2024) MDPC Publishing, Vienna, Austria. 2024. 662 p.

40. Наумов А.Б., Смеляков К.С., Чуприна А.С. “Емоційна оцінка новин для прогнозування курсу біткоїну”. 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 6., – Харків: ХНУРЕ. 2024. – 958 с.

48. Dmytro Nazarenko, Irina Afanasieva, Natalia Golian, Vira Golian Investigation of the deep learning approaches to classify emotions in texts / al. Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021). Volume I: Main Conference, Lviv, 22–23 April 2021. 2021. pp. 206–224.

49. Ворочек О.Г, Дударь В.В. Исследование интеллектуальных средств анализа и прогнозирования. Журнал «Вестник» Херсонского национального технического университета №4(27), 2007.

ДОДАТОК Б

Тексти наукових публікацій за темою кваліфікаційної роботи

ISBN 978-3-954754-01-4

**BITCOIN PRICE PREDICTION USING THE BOOSTING
ALGORITHM**

Afanasieva I.V.,

Naumov A.B.,

Onyshchenko K.G.

Kharkiv National University of Radioelectronics

Kharkiv, Ukraine

anton.naumov.cpe@nure.ua

Abstract: This article shows the use of the boosting algorithm to predict Bitcoin prices, offering comprehensive analysis at the intersection of finance and machine learning. The research includes data preprocessing, feature engineering, and model training, highlighting boosting's effectiveness in identifying complex relationships in cryptocurrency markets. Metrics tuning and hyperparameter estimation further improves model accuracy. Despite the challenges posed by market dynamics, the results highlight boosting's potential to provide valuable insight into Bitcoin price movements, which is critical for investors and researchers navigating the volatile digital asset landscape.

Keywords: Bitcoin, cryptocurrency, boosting algorithm, machine learning, price forecasting, financial forecasting, hyperparameter tuning, cryptocurrency market, predictive analytics.

Introduction

For several years, the authors have been engaged in research in the field of NLP, text vectorization, emotion detection and machine learning. The volume of problems solved over the years shows the excellent applicability of artificial intelligence algorithms for processing and structuring big data and obtaining results in a short time.

The use of large and powerful packages from leading providers such as Google, Facebook and OpenAI has fully proven its effectiveness on large-scale tasks in the field of financial, linguistic, sociological, political and other research.

However, it is not always necessary to use such powerful tools as Whisper, Google Sentiment Analysis or Azure Machine Learning Studio or combine many powerful tools to solve a problem. For simpler tasks, training, comparing and evaluating new models, simpler models can and should be used, such as boosting algorithms.

In recent years, the cryptocurrency market has witnessed unprecedented growth and volatility. Among various digital assets, Bitcoin has become a prominent player. Forecasting price movements has become a critical aspect for both investors and researchers. This article examines the application of the boosting algorithm to Bitcoin price forecasting, shedding light on its potential at the intersection of finance and machine learning.

Aim

The purpose of this work is to review the effectiveness of forecasting the Bitcoin price using technology XGBoost.

XGBoost (eXtreme Gradient Boosting) is a powerful and efficient machine learning algorithm known for its accuracy and speed. It belongs to a family of boosting algorithms that iteratively combine weak learners to create a strong

predictive model. Boosting excels at dealing with complex relationships within data, making it an ideal candidate for financial forecasting.

Although the XGBoost model often provides higher accuracy than a single decision tree, it sacrifices the internal interpretability of decision trees. For example, following the path that a decision tree takes to make a decision is trivial and self-evident, but following the paths of hundreds or thousands of trees is much more difficult. To achieve both performance and interpretability, some model compression techniques allow XGBoost to be converted into a single "regenerated" decision tree that approximates the same decision function.

Features of XGBoost that differentiate it from other gradient boosting algorithms include:

- Smart penalization of trees
- A proportional shrinking of leaf nodes
- Newton Boosting
- Extra randomization parameter
- Implementation on single, distributed systems and out-of-core computation
- Automatic Feature selection
- Theoretically justified weighted quantile sketching for efficient computation
- Parallel tree structure boosting with sparsity
- Efficient cacheable block structure for decision tree training

XGBoost works like Newton-Raphson in function space, unlike gradient boosting which works like gradient descent in function space, the loss function uses a second order Taylor approximation to relate to the Newton-Raphson method.

Methodology

1. Data Preprocessing:

- Collection of historical Bitcoin price data.
- Cleaning and normalization to handle missing values and outliers.

2. Feature Engineering:

- Identification of relevant features influencing Bitcoin prices.
- Transformation of data to enhance predictive capabilities.

3. Training the boosting Model:

- Splitting the dataset into training and testing sets.
- Employing boosting to train the model on historical data.

4. Hyperparameter Tuning:

- Optimizing model parameters for enhanced performance.
- Cross-validation to avoid overfitting.

5. Evaluation Metrics:

- Utilizing metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess model accuracy.

After training the boosting model on historical Bitcoin price data, the accuracy of predicting future prices was a pretty decent one. The algorithm has demonstrated an ability to find relationships in data, demonstrating its usefulness in predicting complex market dynamics. The predictive power of the model was demonstrated by its ability to adapt to changing trends, offering valuable information for long-term investment strategies.

To further optimize the model, hyperparameters were tuned. This process involved systematically adjusting the algorithm parameters to achieve the best possible performance. Through cross-validation techniques, the tuned boosting

model demonstrated enhanced generalization capabilities, reducing overfitting and providing reliable predictions. Fine-tuning the parameters not only improved accuracy, but also made the model more resilient to fluctuations in the cryptocurrency market.

Comparisons with other time series forecasting methods highlighted boosting's ability to handle complex cryptocurrency price data. The learning approach used by boosting algorithm has allowed it to match and in some cases outperform other methods, demonstrating its adaptability to non-linear relationships and dynamic market conditions. This observation supports the view that advanced machine learning algorithms, especially those belonging to the boosting family, offer a significant advantage in predicting cryptocurrency prices.

The implications of using boosting to predict Bitcoin prices are far-reaching. Investors can use the model's accuracy to make informed decisions, potentially maximizing profits and minimizing risks. On the other hand, researchers can benefit from a deeper understanding of the complex relationships in the cryptocurrency market, contributing to the development of financial forecasting.

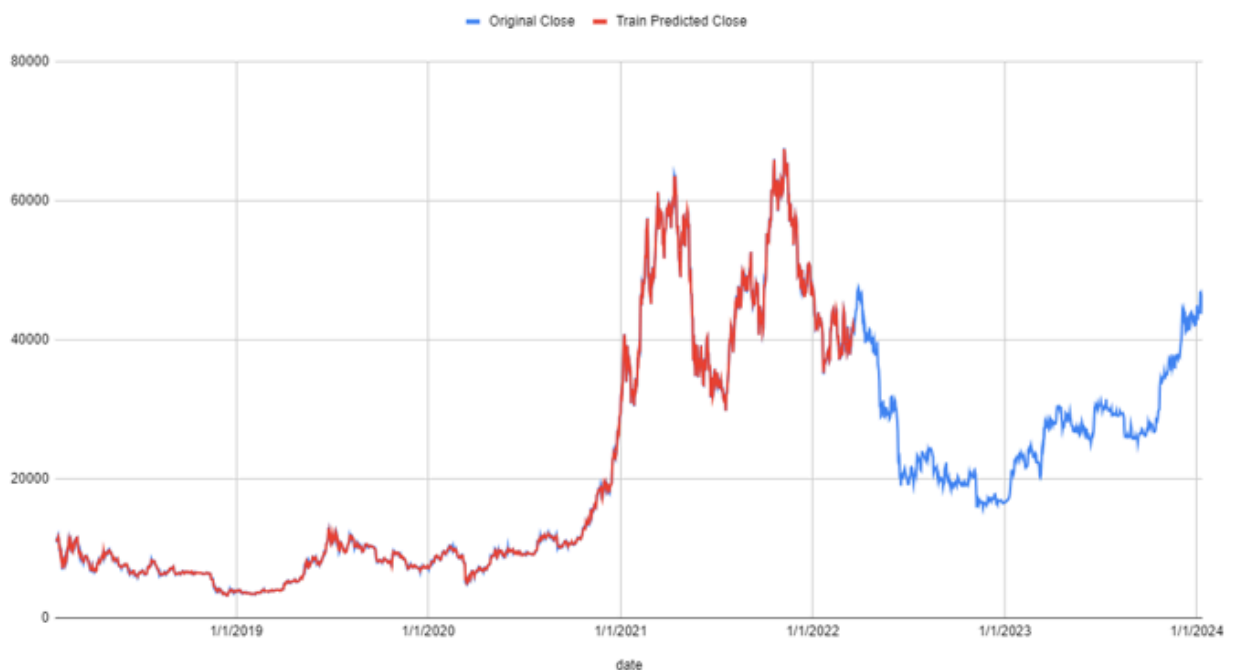
Results and Discussion

The use of the boosting algorithm to predict Bitcoin prices yielded profound results, highlighting its effectiveness in identifying nuanced patterns in the cryptocurrency market. The model's performance was assessed using a comprehensive set of metrics, including mean absolute error (MAE) and root mean square error (RMSE), providing a reliable assessment of its predictive capabilities.

The model has demonstrated a high level of accuracy in identifying complex patterns and trends in the cryptocurrency market. Hyperparameter tuning further improved performance by ensuring reliable prediction of price fluctuations.

Despite the promising results, it is important to acknowledge the challenges associated with cryptocurrency price forecasting. External factors such as regulatory changes, market sentiment and global events can significantly affect prices, adding an element of unpredictability. Although the boosting algorithm excels at capturing historical patterns, its effectiveness in navigating unexpected events remains an area of ongoing research.

XGBoost Bitcoin Price Prediction Chart



Picture 1. XGBoost Bitcoin Price Prediction Chart

Conclusion

In conclusion, the results and discussion highlight the effectiveness of the boosting algorithm in predicting Bitcoin prices. The model's accuracy, adaptability,

and improved performance through hyperparameter tuning make it a valuable tool in the arsenal of both cryptocurrency investors and researchers. While challenges remain in the dynamic cryptocurrency market, the use of sophisticated machine learning algorithms such as boosting offers a promising path to overcome the challenges of digital asset price prediction.

In the dynamic world of cryptocurrencies, using advanced machine learning algorithms like boosting for price prediction becomes mandatory. This article demonstrated the effectiveness of boosting in predicting Bitcoin prices, highlighting its potential to provide valuable information to investors and researchers navigating the volatile world of digital assets.

As we continue to explore the intersection of finance and machine learning, the use of sophisticated algorithms becomes essential to staying ahead in understanding and predicting cryptocurrency market dynamics. The integration of boosting into Bitcoin price forecasting serves as a testament to the advancements in predictive analytics in the ever-evolving blockchain ecosystem.

List of references

1. Chen, T., & Guestrin, C. (2016). boosting: A Scalable and Accurate Implementation of Gradient Boosting Machines. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
2. Brownlee, J. (2018). Introduction to Time Series Forecasting with Python. Machine Learning Mastery.
3. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
5. Huang, G., Zhu, Q. Y., & Siew, C. K. (2006). Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 70(1-3), 489-501.
6. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
8. Oliphant, T. E. (2006). A Guide to NumPy. Trelgol Publishing.
9. Brown, M., Pelosi, M., & Dirska, H. (2013). Volatility Forecasting for Risk Management. *Journal of Forecasting*, 32(7), 561-576.
10. Fan, J., & Yao, Q. (2017). Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

УДК 004.9:336.741.24]:159.942.5 DOI: <https://doi.org/10.30837/IYF.IIS.2024.303>

ЕМОЦІОНАЛЬНА ОЦІНКА НОВИН ДЛЯ ПРОГНОЗУВАННЯ КУРСУ БІТКОЇНУ

Наумов А. Б., Смеляков К. С.

Науковий керівник – к.т.н., доцент Чуприна А. С.

Харківський національний університет радіоелектроніки, каф. ПІ,
м. Харків, Україна

e-mail: anton.naumov.cpe@nure.ua

This article showcases the use of news sentiment analysis to predict Bitcoin prices. It offers a comprehensive analysis at the intersection of finance and machine learning. The research includes data collection and pre-processing, sentiment analysis, feature development, and model training. The article highlights the effectiveness of improved efficiency in identifying complex relationships in cryptocurrency markets. Tuning and estimating hyperparameters further improves the accuracy of the model. Although market volatility presents challenges, the results emphasize the importance of forecasting for investors and cryptocurrency researchers.

У світі фінансових ринків впевнене прогнозування курсів криптовалют є ключовим чинником для інвесторів і трейдерів. Одним з факторів, що впливають на динаміку цін на біткоїн, є громадська думка, виражена через новинні та соціальні медіа. Новини та публікації у соціальних медіа можуть мати значний вплив на психологію учасників ринку, формуючи тренди та напрямки руху цін. У цьому контексті аналіз емоціональних оцінок у джерелах новин набуває важливого значення для прогнозування курсу біткоїну.

Незважаючи на активні дослідження криптовалютних ринків, відносно мало робіт присвячено дослідженню впливу емоціональних оцінок новин на цінову динаміку біткоїну. У зв'язку з цим актуальність цього дослідження полягає у необхідності розробки моделей, здатних враховувати та аналізувати інформацію з джерел новин для прогнозування курсу біткоїну з високою точністю.

Метою даної роботи є дослідження використання методів аналізу емоціональних оцінок новин для прогнозування курсу біткоїну. Ми розглянемо методи збирання, обробки та аналізу даних, що використовуються для отримання емоціональних оцінок новин, а також розробимо модель прогнозування курсу біткоїну на основі цих даних та оцінимо її точність.

У цьому дослідженні використовувалася така методологія:

1. Збір даних:
 - збір даних зі стрічок новин Google і Yahoo (рисунок 1).
2. Обробка даних:

– очищення даних (видалення пунктуації, стоп-слів, нормалізація стемером та лематайзером, видалення цифр).

3. Аналіз емоціональних оцінок:

– поляризація сентиментів за допомогою бібліотек Vader та TextBlob (рисунок 2).

4. Отримання історичних даних для біткоїна:

– використання відкритого джерела <https://finance.yahoo.com/>

5. Навчання моделі з урахуванням отриманих метрик сентиментів:

– поділ даних на тренувальну та тестову частину.

– застосування бібліотеки TensorFlow та високорівневого API Keras для навчання моделі.

6. Оцінка метрик:

– використання метрик MAE та RMSE для оцінки точності.

Дана модель, заснована на нейронних мережах з використанням шарів LSTM [1], продемонструвала високу точність у прогнозуванні динаміки ціни біткоїну на основі емоціональних оцінок (рисунок 3), виражених у новинах (RMSE: train – 0.01009, test – 0.00876, MAE: train – 0.00840, test – 0.00783).

У порівнянні з прогнозуванням курсу біткоїна за допомогою моделі на базі алгоритму XGBoost [2], за допомогою даної методики було отримано на порядок вищу точність.

Цей успіх можна пояснити кількома факторами: використання глибокого навчання дозволяє моделі враховувати часові залежності даних, а також попередня обробка текстів та векторизація даних дозволили ефективно подати текстову інформацію у вигляді числових ознак, що покращило процес навчання моделі.

| | Title | Source | Description | Link | Time |
|---|---|-------------------|--|---|-------------|
| 0 | Bitcoin support levels to watch as BTC price a... | The Cointelegraph | Bitcoin bulls should be able to defend \$28,000... | https://cointelegraph.com/news/bitcoin-support... | 4 hours ago |
| 1 | Bitcoin bulls risk trading range loss as BTC p... | The Cointelegraph | Bitcoin falls on Fed minutes, inflation outloo... | https://cointelegraph.com/news/bitcoin-bulls-r... | 9 hours ago |

Рисунок 1 – Формат даних зібраних зі стрічок новин

| date | source | subject | text | title | url | Subjectivity | Polarity | Analysis |
|---------------------|---------------|------------|--|---|---|--------------|----------|----------|
| 11/15/2022 11:00 | CoinTelegraph | blockchain | roster full entertainment industry notable | Aether Games Sets Sights on Dark Fantasy DeFi | https://cointelegraph.com/news/aether-games-se... | 0.362500 | 0.112500 | 1 |

Рисунок 2 – Формат даних після обробки та поляризації

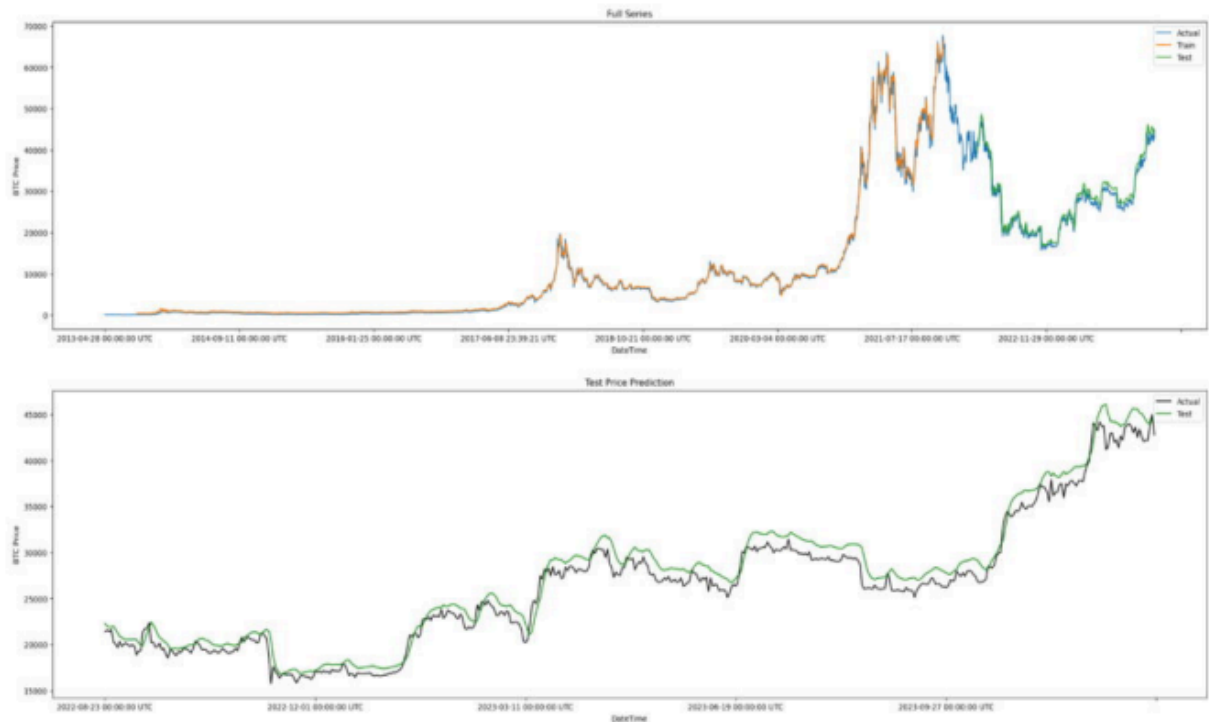


Рисунок 3 – Графіки реального та спрогнозованого курсу біткоїна, побудованого на основі аналізу емоціональних оцінок

На завершення, наше дослідження показало, що аналіз емоціональних оцінок новин із застосуванням інструментів машинного навчання, таких як нейронні мережі, що використовують TensorFlow та Keras, має значний потенціал для прогнозування курсу біткоїну. Було продемонстровано високу точність моделі у передбаченні ціни біткоїну на основі емоціональних оцінок, що підтверджує ефективність цього підходу. Однак, для подальшого розвитку цієї методики та підвищення можливостей її застосування у реальних умовах, необхідно продовжити дослідження у напрямку покращення застосованих методів обробки тексту, збільшення обсягів та різноманітності джерел даних та оптимізації моделей для роботи в реальному часі.

Список використаних джерел:

1. Smelyakov, K., Bizkrovnyi, O., Sharonova, N., Smelyakov, S., Chupryna, A. Building of Regression Models for Cryptocurrency Price Prediction CEUR Workshop Proceedings, 2022, 3171, pp. 1216–1232.
2. Afanasieva I.V., Naumov A.B., Onyshchenko K.G. Bitcoin price prediction using the boosting algorithm. The 2nd International scientific and practical conference “Science and society: modern trends in a changing world” (January 22-24, 2024) MDPC Publishing, Vienna, Austria. 2024. 662 p.

ДОДАТОК В

Слайди презентації



Дослідження ефективності
використання інформаційних
технологій на основі штучного
інтелекту для прогнозування
курсу криптовалют

Наумов А.Б.
Науковий керівник:
к.т.н., доцент кафедри ПІ Чуприна А.С.

21 червня 2024



Актуальність

Ринок криптовалют швидко зростає та розвивається. За останні кілька років вартість біткоіну та інших криптовалютів значно зросла, що привернуло увагу інвесторів і трейдерів.

Прогнозування криптовалют є складним завданням. Через волатильність ринку та відсутність централізованого регулювання ціни криптовалют можуть різко вагатися.

Штучний інтелект може бути використаний для аналізу великих обсягів даних та виявлення закономірностей, які можуть допомогти у прогнозуванні цін на криптовалюту.





<https://www.livecoinwatch.com/crypto-market-cap>

Проблематика

Нестабільність і непередбачуваність ринку: Ринок криптовалют характеризується високою волатильністю, що затрудняє точне прогнозування майбутніх цін. Ціни можуть різко змінюватися під впливом різних факторів, таких як новини, набування інвесторів, зміни в регулюванні та ін.

Відсутність централізованого регулювання: ринок криптовалют не регулюється ні одним центральним органом, що може призвести до маніпулювання ринком та іншими недобросовісними практиками.

Вплив соціальних мереж і новин: Настрої і думки, виражені в соціальних мережах і новинних статтях, можуть впливати на ціни на криптовалюту, що затрудняє аналіз даних і прогнозування. Складність розробки ефективних моделей ШІ:

Розробка моделей ШІ, здатних точно прогнозувати ціни на криптовалюту, є складною задачею, що вимагає глибокого розуміння ринку криптовалют, машинного навчання та обробки даних.



Напря́м дослідження

Машинне навчання. Алгоритми машинного навчання що можуть бути навчені на історичних даних про ціни на криптовалюти, щоб виявляти закономірності та робити прогнози про майбутні ціни.

Аналіз настроїв може бути використаний для оцінки настроїв інвесторів та трейдерів щодо криптовалют.

Обробка природної мови може бути використана для аналізу статей новин та інших текстових даних, щоб витягти інформацію, яка може вплинути на ціни на криптовалюти.



Перелік основних теорій у галузі

Аналіз технічних показників: Цей метод використовує технічні показники, такі як SMA, RSI, і MACD, для аналізу цінних графіків криптовалют і прогнозування майбутніх рухів цін.

Аналіз фундаментальних даних: Цей метод використовує фундаментальні фактори, такі як команда розробників, технологія блокчейну, і прийняття криптовалюти, для оцінки вартості криптовалюти.

Машинне навчання: Цей метод використовує алгоритми машинного навчання для аналізу великих обсягів даних і прогнозування майбутніх цін на криптовалюту.

Глибоке навчання: Цей метод використовує нейронні мережі для аналізу складних закономірностей у даних і прогнозування майбутніх цін на криптовалюту.

Обробка природної мови (NLP): Цей метод використовує NLP для аналізу тексту, пов'язаного з криптовалютою, таких як новини, статті та публікації в соціальних мережах, для виявлення настроїв і прогнозування майбутніх рухів цін.



Прогалини у наявних дослідженнях

Наразі було проведено багато досліджень прогнозування за допомогою різноманітних алгоритмів. Також існує чимало досліджень інструментів для аналізу настроїв (новини, соціальні мережі, тощо).

Але не було знайдено дослідження по порівнянню ефективності саме таких інструментів аналізу настроїв новин, як TextBlob vs Bert vs spaCy vs VADER vs Flair для прогнозування курсу криптовалют з використанням LSTM RNN мережі.



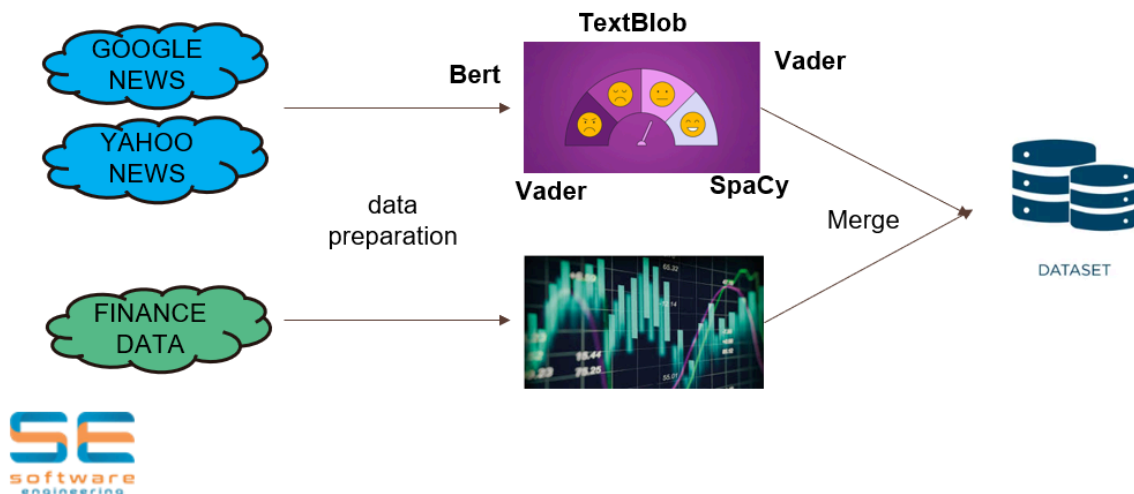
Постановка задачі

Точність прогнозування на довгостроковий період часу є досить низькою, тому було вирішено зупинитися на короткочасному прогнозуванні.

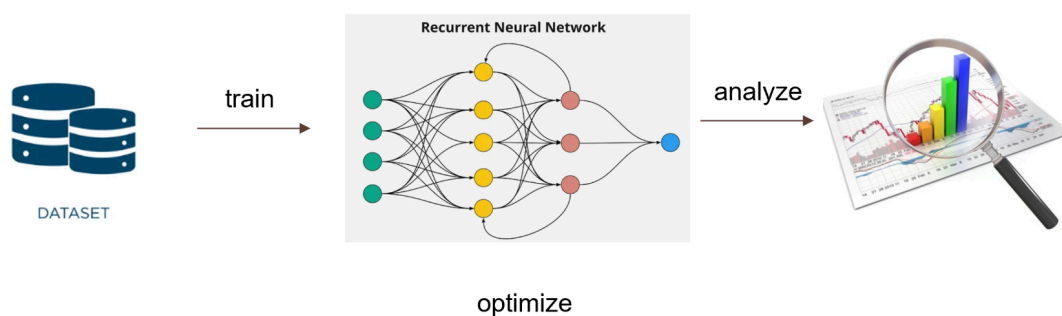
Метою даної роботи є проведення дослідження порівняння ефективності застосування поширених інструментів для аналізу настроїв та алгоритму LSTM для прогнозування ціни закриття криптовалюти на поточний день.



Архітектура системи для проведення експериментального дослідження



Архітектура системи для проведення експериментального дослідження



Методологія

- Задача багатокритеріального вибору криптовалюти для дослідження.
- Збір даних та препроцесинг.
- Аналіз настроїв (Bert, Flair, spaCy, TextBlob , VADER).
- Запропонована модель - LSTM.
- Вибір метрик (MSE, RMSE, MAE).



Опис програмного забезпечення, що було використано у дослідженні

Середовище розробки

- Jupyter Notebook 6.5.4.



Мова

- Python 3.11.5



Використані бібліотеки:

- pandas 2.0.3
- requests 2.3.0
- urllib 3.11.5
- BeautifulSoup 4.12.2
- nltk 3.8.1
- vaderSentiment 3.3.2
- textblob 0.15.3
- transformers 4.32.1 з bert-large-uncased моделлю
- flair 0.13.1
- spacy 3.7.4
- numpy 1.24.3
- matplotlib 3.7.2
- sklearn 1.4.1.post1
- keras 2.15.0
- scipy 1.11.1



Вибір криптовалюти для дослідження

Векторний опис альтернатив

| | Капіталізація, USD | Стабільність за останній місяць, % | Обсяг торгів / в день (середній за ост місяць), USD | Кількість транзакцій в секунду | Історичні дані, міс. |
|-------|--------------------|------------------------------------|---|--------------------------------|----------------------|
| BTC | 775 821 726 845 | 0,10 | 18 047 358 995,00 | 7 | 129 |
| ETH | 259 300 223 439 | 0,09 | 17 068 510 656,00 | 30 | 101 |
| SOL | 26 815 560 650 | 0,00 | 2 090 151 459,00 | 65000 | 44 |
| MATIC | 7 447 803 791 | 0,05 | 580 052 071,00 | 65000 | 56 |
| ARB | 1 383 455 156 | 0,08 | 315 842 978,00 | 40000 | 9 |
| GEL | 92 420 917 | 0,12 | 243413.1935 | 15000 | 27 |



Вибір криптовалюти для дослідження

Розрахунки та найоптимальніша альтернатива

- пошук невідомованих альтернатив за принципом Парето
- нормування оцінок за шкалами за еталоном з визначеними ваговими коефіцієнтами

| | Капіталізація, USD | Стабільність за останній місяць, % | Обсяг торгів / в день (середній за ост місяць), USD | Кількість транзакцій в секунду | Історичні дані, міс. | |
|-------------|--------------------|------------------------------------|---|--------------------------------|----------------------|-------------|
| Рез-т BTC | 0,10000 | 0,29167 | 0,10000 | 0,00001 | 0,35000 | 0,84 |
| Рез-т ETH | 0,03342 | 0,26250 | 0,09458 | 0,00005 | 0,27403 | 0,66 |
| Рез-т SOL | 0,00346 | 0,00000 | 0,01158 | 0,10000 | 0,11938 | 0,23 |
| Рез-т MATIC | 0,00096 | 0,14583 | 0,00321 | 0,10000 | 0,15194 | 0,40 |
| Рез-т GEL | 0,00001 | 0,35000 | 0,00000 | 0,02308 | 0,07326 | 0,45 |



Збір даних

<https://www.google.com>

Google search results for 'btc'. The search bar shows 'btc' and the results are filtered to 'News'. The page shows 'About 5,590,000 results (0.37 seconds)'. The main heading is 'News about Cryptocurrency'. A table of news items is displayed with columns for Title, Source, Description, Link, and Date.

| | Title | Source | Description | Link | Date |
|---|--|-----------------------|--|---|--------------|
| 0 | Bitcoin 'Halving' Due Next Year Spurs Predict... | Bloomberg.com | Bitcoin's rebound is just the start of a rally... | https://www.bloomberg.com/news/articles/2023-0... | Apr 23, 2023 |
| 1 | Crypto winter is over — and bitcoin could hit ... | CNBC | Crypto winter is over — and bitcoin could hit ... | https://www.cnbc.com/2023/04/24/bitcoin-btc-pr... | Apr 24, 2023 |
| 2 | Bitcoin could hit \$100000 by end-2024, Standar... | Reuters | Top cryptocurrency bitcoin could reach \$100000... | https://www.reuters.com/markets/us/bitcoin-cou... | Apr 24, 2023 |
| 3 | BTC Presents "Brand New" Spring Recital | City of Pittsburg, KS | Break The Ceiling Athletics & Dance Present Th... | https://www.pittks.org/event/btc-presents-bran... | Apr 3, 2023 |
| 4 | Bitcoin's Lightning Network: What It Is And Ho... | Forbes | It's a network of payment channels that allows... | https://www.forbes.com/sites/digital-assets/20... | Apr 27, 2023 |

<https://news.search.yahoo.com>

Yahoo News search results for 'btc'. The search bar shows 'btc' and the results are filtered to 'News'. An advertisement for 'Current Bitcoin Price - Bitcoin Exchange' is visible at the top of the results.



Збір даних

<https://finance.yahoo.com>

Yahoo Finance page for Bitcoin Futures, Jun-2024 (BTC=F). The page shows a price chart and a data table. The current price is \$69,750.00, down 5.00 (-0.01%) from the previous close. The chart shows the price movement from 2014 to 2024. The data table below the chart provides historical price and volume data.

| Date | Close | Volume | Change, % |
|----------------|--------------|--------------|-----------|
| 0 9/17/2014 | 457.334015 | 2.105680e+07 | NaN |
| 1 9/18/2014 | 424.440002 | 3.448320e+07 | -7.192558 |
| 2 9/19/2014 | 394.795590 | 3.791970e+07 | -6.984264 |
| 3 9/20/2014 | 408.903992 | 3.686360e+07 | 3.573492 |
| 4 9/21/2014 | 398.821014 | 2.658010e+07 | -2.465855 |
| ... | ... | ... | ... |
| 3479 3/27/2024 | 69455.343750 | 4.082711e+10 | -0.760835 |
| 3480 3/28/2024 | 70744.953130 | 3.437490e+10 | 1.856746 |
| 3481 3/29/2024 | 69892.828130 | 2.523085e+10 | -1.204503 |
| 3482 3/30/2024 | 69645.304690 | 1.713024e+10 | -0.354147 |
| 3483 3/31/2024 | 71333.648440 | 2.005094e+10 | 2.424203 |



Підготовка даних

Спочатку було об'єднано вміст колонок Title та Description в колонку Text з метою спрощення оцінки сентиментів.

Далі було видалено усі спеціальні символи та цифри, залишаючи тільки символи алфавіту.

Також було виконано приведення до нижнього регістру.

Видалення стоп-слів - це базова техніка обробки природної мови (NLP), яка полягає у виключенні з найпоширеніших і неінформативних слів. Було використано модуль stopwords бібліотеки NLTK для виконання цієї процедури.

Лематизація - це процес приведення словоформи до її базової чи нормальної форми, що називається лемою. Лема являє собою канонічну форму слова, яка є його словниковою формою або основою. Наприклад, лемою для слова "бігав" буде слово "бігати", для "кішок" - "кішка".

Для виконання лематизації було використано модуль WordNetLemmatizer бібліотеки NLTK.



Аналіз сентиментів

Аналіз сентиментів – це процес визначення тональності чи емоційного забарвлення тексту. Метою аналізу сентиментів є визначення того, чи є текст позитивним, негативним чи нейтральним.

Для наглядності дослідження було вирішено використовувати одну модель на різних датасетах. Задля цього було проведено оцінку сентиментів новин для кожного з вибраних інструментів аналізу (Bert, Flair, spaCy, TextBlob, VADER) та отримано 2 оцінки – полярність та сентимент на чотирьох часових інтервалах – з початку (2014р), за останні 4, 2 та один рік.



Підготовка датасетів

Групування sentimentів з даними BTC було виконано за середнім показником та по максимальному показнику.

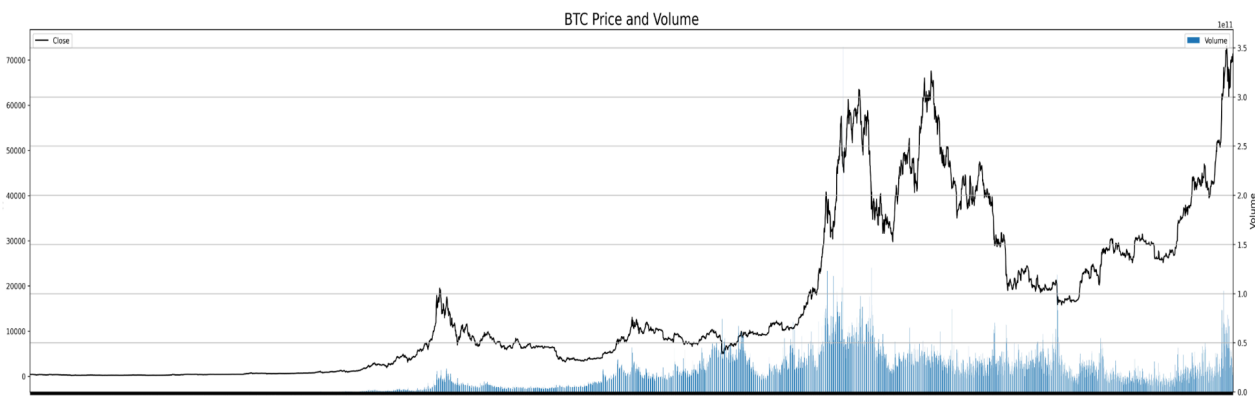
Всього маємо 40 датасетів (5 інструментів аналізу, 4 часові інтервали, 2 підходи до групування).

```
def analysis(score):
    if score < 0:
        return -1
    elif score == 0:
        return 0
    else:
        return 1
```

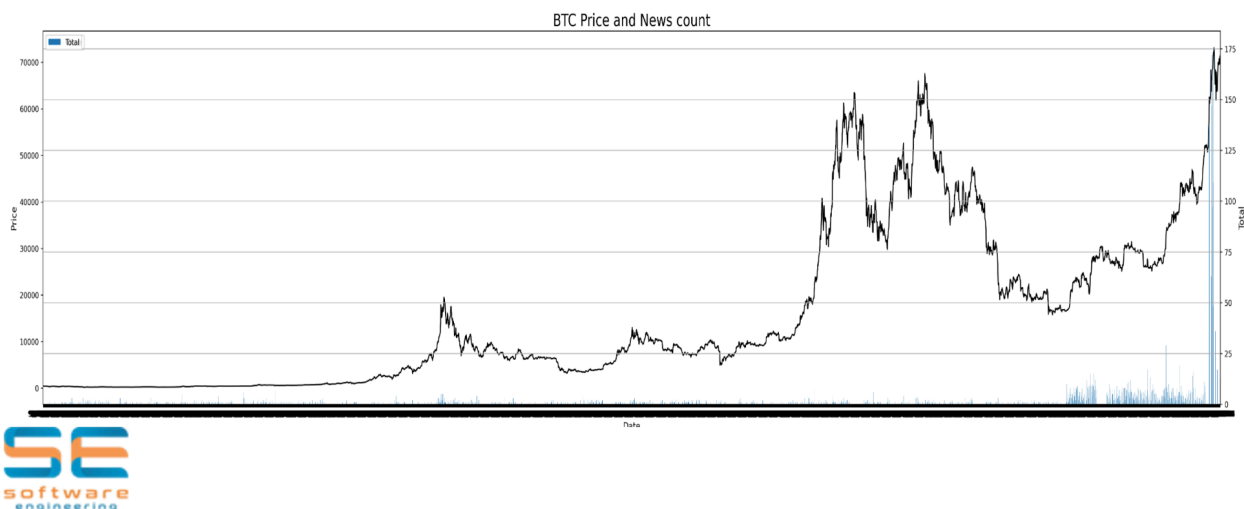
| | Date | Close | Volume | Polarity | Sentiment |
|-----|-----------|----------|----------------|----------|-----------|
| 0 | 1/1/2024 | 44167.33 | 18426978443.00 | -0.58 | -1.00 |
| 1 | 1/2/2024 | 44957.97 | 39335274536.00 | -0.60 | -1.00 |
| 2 | 1/3/2024 | 42848.18 | 46342323118.00 | -0.66 | -1.00 |
| 3 | 1/4/2024 | 44179.92 | 30448091210.00 | -0.60 | -1.00 |
| 4 | 1/5/2024 | 44162.69 | 32336029347.00 | -0.61 | -1.00 |
| ... | ... | ... | ... | ... | ... |
| 86 | 3/27/2024 | 69455.34 | 40827113309.00 | 0.00 | 0.00 |
| 87 | 3/28/2024 | 70744.95 | 34374900617.00 | 0.00 | 0.00 |
| 88 | 3/29/2024 | 69892.83 | 25230851763.00 | 0.00 | 0.00 |
| 89 | 3/30/2024 | 69645.30 | 17130241883.00 | 0.00 | 0.00 |
| 90 | 3/31/2024 | 71333.65 | 20050941373.00 | 0.00 | 0.00 |



Ціна BTC та обсяг торгів з 2014 року



Ціна BTC та кількість новин з 2014

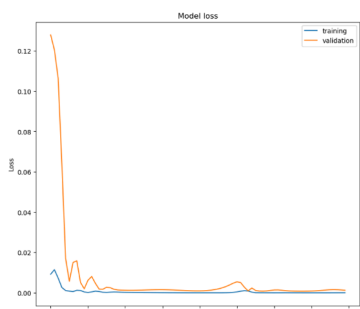


Тренування моделі

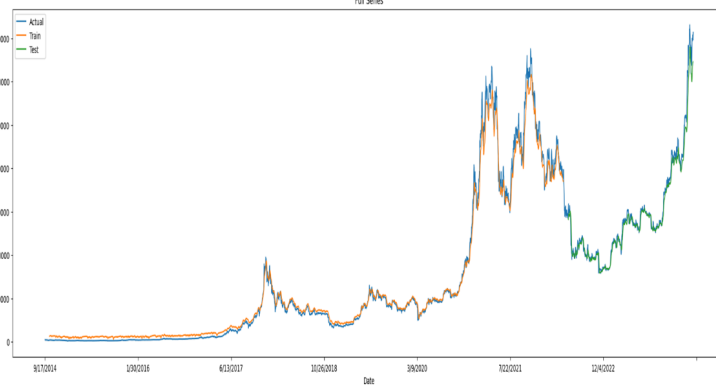
```

model = Sequential()
model.add(LSTM(num_units, input_shape=(look_back,num_of_features)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
history= model.fit(x_train, y_train, validation_split=0.30, epochs=epochs, batch_size=batch_size, shuffle=False)

```



Number of features 4
 Training dataset length 2787
 Testing dataset length 697



Аналіз отриманих результатів

| | | | TextBlob | | | Vader | | | Flair | | | Spacy | | | Bert | | |
|------|-----|-------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | | | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| max | max | train | 0.00063 | 0.02505 | 0.01801 | 0.00035 | 0.01871 | 0.01342 | 0.00054 | 0.02328 | 0.01414 | 0.00056 | 0.02362 | 0.01731 | 0.00047 | 0.02164 | 0.01428 |
| | | test | 0.00087 | 0.02957 | 0.0179 | 0.00046 | 0.02143 | 0.01307 | 0.00121 | 0.03472 | 0.02665 | 0.00077 | 0.0277 | 0.01672 | 0.00063 | 0.02509 | 0.01517 |
| | ave | train | 0.00033 | 0.01828 | 0.01268 | 0.00034 | 0.01853 | 0.01414 | 0.00052 | 0.02289 | 0.0142 | 0.00024 | 0.01544 | 0.00785 | 0.00046 | 0.02146 | 0.01455 |
| | | test | 0.00055 | 0.02335 | 0.01561 | 0.00042 | 0.02046 | 0.01274 | 0.001 | 0.03165 | 0.0229 | 0.00037 | 0.01933 | 0.0125 | 0.00057 | 0.02377 | 0.0139 |
| 2021 | max | train | 0.00107 | 0.03277 | 0.02383 | 0.00154 | 0.03926 | 0.03055 | 0.00117 | 0.0342 | 0.02485 | 0.00113 | 0.03369 | 0.02427 | 0.00266 | 0.05155 | 0.03821 |
| | | test | 0.00092 | 0.0304 | 0.02076 | 0.00296 | 0.05438 | 0.04725 | 0.00133 | 0.03652 | 0.02502 | 0.00107 | 0.0327 | 0.02224 | 0.00131 | 0.03619 | 0.02512 |
| | ave | train | 0.00121 | 0.03479 | 0.02493 | 0.00112 | 0.03339 | 0.02376 | 0.00114 | 0.03382 | 0.02447 | 0.00109 | 0.03304 | 0.02353 | 0.00323 | 0.05687 | 0.04039 |
| | | test | 0.00123 | 0.03504 | 0.02417 | 0.0013 | 0.0361 | 0.02482 | 0.00159 | 0.03988 | 0.02935 | 0.00109 | 0.03298 | 0.02261 | 0.002 | 0.04469 | 0.03054 |
| 2023 | max | train | 0.00101 | 0.03185 | 0.02321 | 0.00052 | 0.02276 | 0.01665 | 0.00064 | 0.02521 | 0.01808 | 0.0007 | 0.02648 | 0.01949 | 0.00065 | 0.02542 | 0.01774 |
| | | test | 0.02594 | 0.16107 | 0.14437 | 0.00411 | 0.06411 | 0.04998 | 0.01204 | 0.10973 | 0.09557 | 0.01951 | 0.13969 | 0.12352 | 0.01515 | 0.1231 | 0.10713 |
| | ave | train | 0.00061 | 0.02469 | 0.01781 | 0.00047 | 0.02168 | 0.01611 | 0.00046 | 0.02143 | 0.01554 | 0.0006 | 0.02453 | 0.01776 | 0.00068 | 0.02601 | 0.0181 |
| | | test | 0.00894 | 0.09456 | 0.0796 | 0.00585 | 0.07646 | 0.0638 | 0.00539 | 0.07341 | 0.06186 | 0.00926 | 0.09624 | 0.0812 | 0.0141 | 0.11876 | 0.10368 |
| 2024 | max | train | 0.01093 | 0.10453 | 0.06172 | 0.00699 | 0.08362 | 0.05092 | 0.01127 | 0.10616 | 0.06938 | 0.01019 | 0.10094 | 0.05957 | 0.01031 | 0.10156 | 0.06153 |
| | | test | 0.09253 | 0.30419 | 0.30369 | 0.05842 | 0.2417 | 0.24112 | 0.06982 | 0.26423 | 0.26389 | 0.0707 | 0.26589 | 0.26547 | 0.05069 | 0.22515 | 0.22475 |
| | ave | train | 0.01374 | 0.1172 | 0.06626 | 0.00739 | 0.08599 | 0.0525 | 0.00767 | 0.08756 | 0.05536 | 0.01477 | 0.12152 | 0.06796 | 0.01299 | 0.11399 | 0.06878 |
| | | test | 0.15626 | 0.3953 | 0.3949 | 0.05646 | 0.23762 | 0.23683 | 0.06097 | 0.24693 | 0.24646 | 0.17658 | 0.42021 | 0.41976 | 0.06156 | 0.24811 | 0.24781 |

Оптимізація датасетів

З'ясуємо, який з показників – сентимент чи полярність дає більшу ефективність прогнозування. Задля цього підготуємо ще 10 датасети. По 2 типу до 5 інструментів. Використаємо тільки найбільший часовий інтервал, тому що він показав найкращий результат.

| | Close | Volume | Sentiment |
|------|--------------|-------------|-----------|
| 0 | 457.334015 | 21056800 | 0.000000 |
| 1 | 424.440002 | 34483200 | 0.000000 |
| 2 | 394.795990 | 37919700 | 0.000000 |
| 3 | 408.903992 | 36863600 | 0.000000 |
| 4 | 398.821014 | 26580100 | 0.000000 |
| ... | ... | ... | ... |
| 3479 | 69455.343750 | 40827113309 | 0.000000 |
| 3480 | 70744.953130 | 34374900617 | 0.000000 |
| 3481 | 69892.828130 | 25230851763 | 0.000000 |
| 3482 | 69645.304690 | 17130241883 | 0.000000 |
| 3483 | 71333.648440 | 20050941373 | 0.461538 |

| | Close | Volume | Polarity |
|------|--------------|-------------|----------|
| 0 | 457.334015 | 21056800 | 0.000000 |
| 1 | 424.440002 | 34483200 | 0.000000 |
| 2 | 394.795990 | 37919700 | 0.000000 |
| 3 | 408.903992 | 36863600 | 0.000000 |
| 4 | 398.821014 | 26580100 | 0.000000 |
| ... | ... | ... | ... |
| 3479 | 69455.343750 | 40827113309 | 0.000000 |
| 3480 | 70744.953130 | 34374900617 | 0.000000 |
| 3481 | 69892.828130 | 25230851763 | 0.000000 |
| 3482 | 69645.304690 | 17130241883 | 0.000000 |
| 3483 | 71333.648440 | 20050941373 | 0.260269 |

Аналіз оптимізованих результатів

Використання **середньо-взваженого показника полярності** показує більш чутливу оцінку і більш точне прогнозування ніж інші показники настроїв, що використовувались в даному дослідженні. Найкращім інструментом для аналізу настроїв виявився **TextBlob**.

| | | TextBlob | | | Vader | | | Flair | | | Spacy | | | Bert | | |
|-----------|-------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| Original | train | 0.00033 | 0.01828 | 0.01268 | 0.00034 | 0.01853 | 0.01414 | 0.00052 | 0.02289 | 0.0142 | 0.00024 | 0.01544 | 0.00785 | 0.00046 | 0.02146 | 0.01455 |
| | test | 0.00055 | 0.02335 | 0.01561 | 0.00042 | 0.02046 | 0.01274 | 0.001 | 0.03165 | 0.0229 | 0.00037 | 0.01933 | 0.0125 | 0.00057 | 0.02377 | 0.0139 |
| Polarity | train | 0.0002 | 0.01429 | 0.00693 | 0.00024 | 0.01555 | 0.00976 | 0.00025 | 0.01591 | 0.00797 | 0.00025 | 0.01592 | 0.00778 | 0.00029 | 0.01689 | 0.00985 |
| | test | 0.00031 | 0.01757 | 0.01136 | 0.00034 | 0.01852 | 0.01137 | 0.00039 | 0.0197 | 0.01247 | 0.00041 | 0.0202 | 0.01242 | 0.00053 | 0.02299 | 0.01793 |
| Sentiment | train | 0.00028 | 0.01671 | 0.00938 | 0.0009 | 0.03 | 0.01395 | 0.00028 | 0.01677 | 0.0114 | 0.00067 | 0.00067 | 0.01727 | 0.00028 | 0.01674 | 0.0121 |
| | test | 0.00046 | 0.02146 | 0.01283 | 0.00204 | 0.04516 | 0.03489 | 0.00039 | 0.01978 | 0.01287 | 0.00117 | 0.00117 | 0.02348 | 0.00037 | 0.01911 | 0.01327 |

Висновки

| Absolute | | | | | |
|----------|-----|-------|----------|----------|----------|
| | | | MSE | RMSE | MAE |
| ave | min | train | 0.00004 | 0.00115 | 0.00092 |
| | | test | 0.00006 | 0.00176 | 0.00114 |
| | ave | train | 0.000132 | 0.003608 | 0.004226 |
| | | test | 0.000186 | 0.003916 | 0.00242 |

| % | | | | | |
|-----|-----|-------|---------|---------|---------|
| | | | MSE | RMSE | MAE |
| ave | min | train | 0.16667 | 0.07448 | 0.11720 |
| | | test | 0.16216 | 0.09105 | 0.09120 |
| | ave | train | 0.34921 | 0.18675 | 0.33318 |
| | | test | 0.00004 | 0.00115 | 0.00092 |

З цих даних можемо побачити, що оптимізація за полярністю оцінок TextBlob майже на 8-17 відсотків показало покращило прогноз відносно найкращих показників до оптимізації і майже до 35 відсотків від середніх показників полярності за всіма датасетами.

Також тренування моделі на більших інтервалах часу показує кращі результати, незважаючи на те, що кількість новин збільшувалася на протязі усього періоду та їх щільність також збільшувалася у датасетах з 2021, 2023 та 2024 років по зростанню.

Публікація результатів



ISBN 978-3-954754-01-4

BITCOIN PRICE PREDICTION USING THE BOOSTING ALGORITHM

Afanaseva I.V.,
Naumov A.B.,
Onyshchenko K.G.
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
anton.naumov.cpe@nure.ua

Abstract: This article shows the use of the boosting algorithm to predict Bitcoin prices, offering comprehensive analysis at the intersection of finance and machine learning. The research includes data preprocessing, feature engineering, and model training, highlighting boosting's effectiveness in identifying complex relationships in cryptocurrency markets. Metrics tuning and hyperparameter estimation further improves model accuracy. Despite the challenges posed by market dynamics, the results highlight boosting's potential to provide valuable insight into Bitcoin price movements, which is critical for investors and researchers navigating the volatile digital asset landscape.

Keywords: Bitcoin, cryptocurrency, boosting algorithm, machine learning, price forecasting, financial forecasting, hyperparameter tuning, cryptocurrency market, predictive analytics.

Публікація результатів

МАТЕРІАЛИ ХХVІІІ МІЖНАРОДНОГО МОЛОДІЖНОГО
ФОРУМУ

«РАДІОЕЛЕКТРОНІКА ТА МОЛОДЬ
У ХХІ СТОЛІТТІ»

16 – 18 квітня 2024 р.

Том 6

КОНФЕРЕНЦІЯ
«ІНФОРМАЦІЙНІ ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ»
INFORMATION INTELLIGENT SYSTEMS



УДК 004.9:336.741.24]:159.942.5 DOI: <https://doi.org/10.30837/IYF.IIS.2024.303>

ЕМОЦІОНАЛЬНА ОЦІНКА НОВИНИ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ БІТКОЇНУ

Наумов А. Б., Смеляков К. С.
Науковий керівник – к.т.н., доцент Чурина А. С.
Харківський національний університет радіоелектроніки, каф. ПІ,
м. Харків, Україна
e-mail: anton.naumov.cpe@nure.ua

This article showcases the use of news sentiment analysis to predict Bitcoin prices. It offers a comprehensive analysis at the intersection of finance and machine learning. The research includes data collection and pre-processing, sentiment analysis, feature development, and model training. The article highlights the effectiveness of improved efficiency in identifying complex relationships in cryptocurrency markets. Tuning and estimating hyperparameters further improves the accuracy of the model. Although market volatility presents challenges, the results emphasize the importance of forecasting for investors and cryptocurrency researchers.

У світі фінансових ринків введено прогнозування курсів криптовалют є ключовим чинником для інвесторів і трейдерів. Одним з факторів, що впливають на динаміку цін на біткоїні, є громадська думка, виражена через новинні та соціальні медіа. Новини та публікації у соціальних медіа можуть мати значний вплив на психологію учасників ринку, формуючи тренди та напрямки руху цін. У цьому контексті аналіз емоціональних оцінок у джерелах новин набуває важливого значення для прогнозування курсу біткоїну.

Підсумки

Отримані результати дають можливість порівняти ефективність різних методів прогнозування ціни закриття криптовалюти на поточний день за допомогою використання бібліотек аналізу настроїв криптовалют.

Використовуючи ці результати можна розробити сервіс для прогнозування курсу криптовалют на поточний день. Цей сервіс може бути корисний для трейдерів, фінансових аналітиків та інших учасників ринку криптовалют.

Також це дослідження може бути корисне для наукових дослідників, студентів, що вивчають теми прогнозування ціни фінансових активів.

Також результати цього дослідження можуть бути використані для подальшого дослідження інструментів аналізу настроїв та прогнозування ціни криптовалют.

ДОДАТОК Г

Звіт з результатів перевірки на унікальність тексту в базі ХНУРЕ



Ім'я користувача:
Кардаш Євген Вікторович каф.ПІ

ID перевірки:
1016334681

Дата перевірки:
08.06.2024 11:12:15 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
08.06.2024 11:16:03 EEST

ID користувача:
100013622

Назва документа: 2024_М_ПІ_ІПЗзди-22-1_Наумов_А_Б_скорочений

Кількість сторінок: 53 Кількість слів: 9875 Кількість символів: 72598 Розмір файлу: 2.03 MB ID файлу: 1016135139

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

4.85%
Схожість

Найбільша схожість: 0.84% з Інтернет-джерелом (<https://ela.kpi.ua/handle/123456789/31668>)

4.54% Джерела з Інтернету

53

Сторінка 55

1.51% Джерела з Бібліотеки

53

Сторінка 56

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи

13

Підозріле форматування

10
сторінок

ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ППЗздм-22-1
(група)

Наумов Антон Борисович

(прізвище, ім'я, по батькові)

Зауваження

| Пункт ДСТУ 3008-2015 | Зміст пункту | Сторінка кваліфікаційної роботи |
|-------------------------|---|---------------------------------------|
| 1 | 2 | 3 |
| | 7.1 Загальні положення | |
| | 7.3 Нумерація сторінок звіту | |
| | 7.4 Нумерація розділів, підрозділів, пунктів, підпунктів | |
| | 7.5 Рисунки | |
| | 7.6 Таблиці | |
| | 7.7 Переліки | |
| | 7.8 Примітки | |
| | 7.9 Виноски | |
| | 7.10 Формули та рівняння | |
| | 7.11 Посилання | |
| | 7.13 Список авторів | |
| | 7.14 Скорочення та умовні позначки | |
| | 7.15 Додатки | |

зауважень немає

Експерт

(підпис)

Олена ОЛІЙНИК
(прізвище, ініціали)

14.06.2024