

## ПРОГНОЗУВАННЯ КОРОТКОЧАСНИХ ДАНИХ ДЛЯ КРИПТОРИНКУ

*Анотація:* Ефективне проведення високочастотної торгівлі можливе за умови швидкого виконання операцій, яке може забезпечити автоматизація процесу здійснення транзакцій. Розглянуто застосування методів експоненційного згладжування для короткочасного прогнозування ціни на криптобіржах. Для оцінки прогнозу ціни на криптовалюту проаналізовано три методи: точність прогнозування тренду, RMSE та MAPE. Дослідження проведені на даних криптобіржі Binance для символу BTCUSDT. Методи експоненційного згладжування перевірені на різних діапазонах даних розміру вікна та коефіцієнту  $\alpha$ . Здійснено порівняння статистичних даних короткочасного прогнозування ціни на криптобіржі, отриманих застосуванням кожного з трьох методів. За результатами дослідження встановлено, що прогноз ціни методом простого експоненційного згладжування дає найкращий результат.

*Ключові слова:* прогнозування тренду, експоненційне згладжування, Simple Exponential Smoothing, Holt's Linear Method, Additive Damped Trend Method, точність прогнозування тренду, RMSE, MAPE, криптобіржі, криптовалюти, прогнозування ціни.

### Вступ

Популярність та повсюдне використання різноманітних криптовалют призвели до створення відповідної інфраструктури для її існування. З'явилися криптобіржі, відповідна законодавча база та інші інституції, необхідні для ефективного функціонування такого типу валют, існування яких відбувається в інформаційному середовищі. Активний розвиток криптовалют та криптобірж викликає необхідність розроблення різноманітних інформаційних технологій, необхідних для створення інструментарію аналізу, торгівлі, а також для автоматизації інших бізнес-процесів, що виконуються у фінансових установах. Інструменти аналізу дозволяють користувачам досліджувати ринок, виявляти його тренди, відстежувати динаміку цін та інше. Традиційно функціонал засобів аналізу використовує різноманітні методи, які дозволяють будувати графіки динаміки цін, здійснювати технічний аналіз, створювати індикатори та отримувати інші результати аналізу даних щодо криптовалют.

Інструментарії аналізу та торгівлі дозволяють користувачам ефективно працювати на криптобіржах та забезпечують їм можливість отримати прибуток від торгівлі криптовалютами та зменшити ризики, пов'язані з купівлею та продажем криптовалют.

Однією з багатьох фінансових операцій, пов'язаних з криптовалютою, в яких може брати участь велика кількість компаній та приватних осіб, є купівля та продаж криптовалюти. Для ефективної торгівлі важливо постійно відстежувати курс криптовалюти та швидко приймати рішення щодо її продажу або купівлі. Один з основних підходів до торгівлі заснований на використанні алгоритму високочастотної торгівлі (high-frequency trading або HFT) [1]. Алгоритм, побудований на прогнозуванні цін на криптовалютній біржі, здійснює прогноз на декілька секунд уперед та використовує результати прогнозу для швидкого відкриття або закриття позицій. Для ефективного проведення HFT транзакцій важлива висока швидкість виконання операцій, якої можна досягнути тільки за умов автоматизації процесу здійснення транзакцій. У такому випадку автоматизована система, яка використовує цей алгоритм, прогнозує ціну певної криптовалюти і, якщо ціна змінюється протягом декількох секунд, то система може самостійно купити цю криптовалюту за низькою ціною та продати її, коли ціна зросте.

Для автоматизації HFT необхідно виконати кілька етапів обробки інформації про ціну криптовалюти. Спочатку треба здійснити фільтрацію даних короткострокового прогнозу ціни. Для цього доцільно використовувати експоненційне згладжування. Експоненційне згладжування належить до методів статистичного аналізу часових рядів та дозволяє не лише згладжувати шуми та випадкові флуктуації в даних, а й виявляти тренди та періодичності, які існують у змінах цін криптовалют. При використанні експоненційного згладжування для кожного моменту часу визначається експоненційно зважена середня величина значень всіх попередніх точок отримання даних. Причому, чим далі дані у часі від даних точки аналізу, тим менша їх вага у середньому. Тобто нові дані мають більшу вагу (в залежності від коефіцієнту  $\alpha$ ), ніж старі, але всі попередні дані також мають певний вплив на кінцевий результат згладжування.

### Методи оцінки результатів прогнозування

Для оцінки прогнозу розглянемо три методи: точність прогнозування тренду (ТПТ) [2], RMSE (Root Mean Square Error) [4] та MAPE (Mean Absolute Percentage Error).

Метод, заснований на розрахунку ТПТ, визначає точність прогнозу за допомогою порівняння подій, які прогнозовані правильно відносно усіх подій. За ознаку порівняння взято тренд  $T_{i+1}$  між попереднім  $P_i$  та наступним  $P_{i+1}$  значенням ціни. Для розрахунку  $T_{i+1}$  використовується таке рівняння:

$$T_{i+1} = \text{sig}(P_{i+1} - P_i),$$

тому тренд може приймати одне з трьох значень:

$$T_{i+1} = \{-1, 0, 1\}.$$

У методі ТПТ відбувається підрахунок кількості позитивно прогнозованих подій. Позитивно прогнозовані події  $S_{i+1}^+$  – це події, у яких передбачений тренд  $\hat{T}_{i+1}$  такий, як і фактичний  $T_{i+1}$ .

Точність  $\Delta_{ТПТ}$  підходу ТПТ визначається у такий спосіб:

$$\Delta_{ТПТ} = \frac{\sum S_{i+1}^+}{\sum S_{i+1}}, \forall i.$$

У методі RMSE розраховується корінь середньоквадратичної помилки. Обчислення здійснюються шляхом взяття квадратного кореня середнього значення квадратів різниць між передбаченими та фактичними значеннями за такою формулою:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}.$$

У методі MAPE визначається відсоток середньої абсолютної похибки прогнозу від фактичного значення.

Спочатку обчислюється абсолютна похибка для кожного прогнозу  $|Y_i - \hat{Y}_i|$ , як модуль різниці прогнозованого  $\hat{Y}_i$  та фактичного значення  $Y_i$ .

Результатом роботи MAPE є середнє значення абсолютних похибок:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i}.$$

Чим менше значення MAPE, тим точніше відбувається прогнозування [3].

### Методи досліджування

Для дослідження обрано секундні дані про ціну символу BTCUSDT [5] криптобіржі Binance. Графік зміни ціни за три хвилини (180 с) наведено на рис. 1.

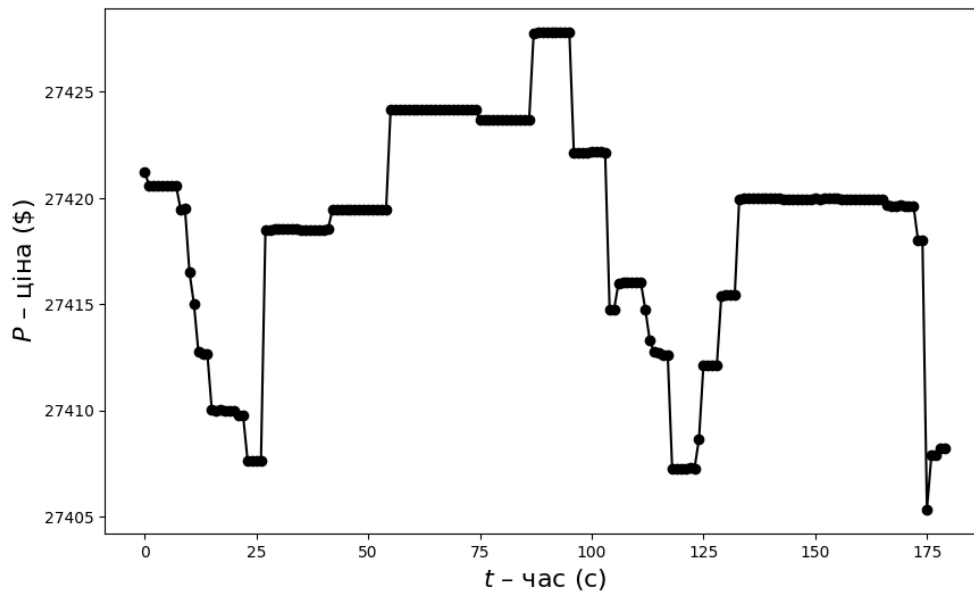


Рисунок 1. Зміна ціни на криптовалюту BTCUSDT протягом 180 с

Для прогнозування формується початкова вибірка даних, або вікно, яке зсонується на одиницю для розрахунків у кожній наступній точці.

Розглянемо три методи прогнозування [6], у яких використовується експоненційне згладжування.

1. Просте експоненційне згладжування (Simple Exponential Smoothing) – метод, що використовується для прогнозування часових рядів, що мають експоненційно зменшувану залежність. Цей метод використовується для того, щоб зменшити вплив шуму і випадкових змін на прогноз. Обчислення відбуваються у такий спосіб:

$$\hat{y}_{t+h|t} = l_t,$$

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1},$$

де  $h$  – крок прогнозування,  $l_t$  – згладжене (прогнозоване) значення в час  $t$ ,  $\alpha$  – параметр згладжування.

2. Метод Хольта (Holt's Linear Method) [8] – це розширений метод експоненційного згладжування, який містить в собі дві компоненти – рівень та тенденцію. Метод Хольта доцільно використовувати для прогнозування часових рядів, що мають лінійну тенденцію. Обчислення відбуваються за такими виразами:

$$\hat{y}_{t+h|t} = l_t + hb_t,$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta^* (l_t - l_{t-1}) + (1 - \beta^*)b_{t-1},$$

де  $b_t$  – значення тренду у час  $t$ ,  $\beta^*$  – параметр згладжування тренду.

3. Адитивне згладжування тренду зі згасанням (Additive Damped Trend Method) [9] – це метод, що використовується для прогнозування часових рядів, які мають тренд та сезонні зміни. Метод передбачає, що тренд буде згасати до деякого фіксованого значення в майбутньому. Розрахунок здійснюється за рівняннями:

$$\hat{y}_{t+h|t} = l_t + \phi_h b_t,$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}),$$

$$b_t = \beta^* (l_t - l_{t-1}) + (1 - \beta^*) \phi b_{t-1}.$$

де  $\phi$  – параметр згладжування [6].

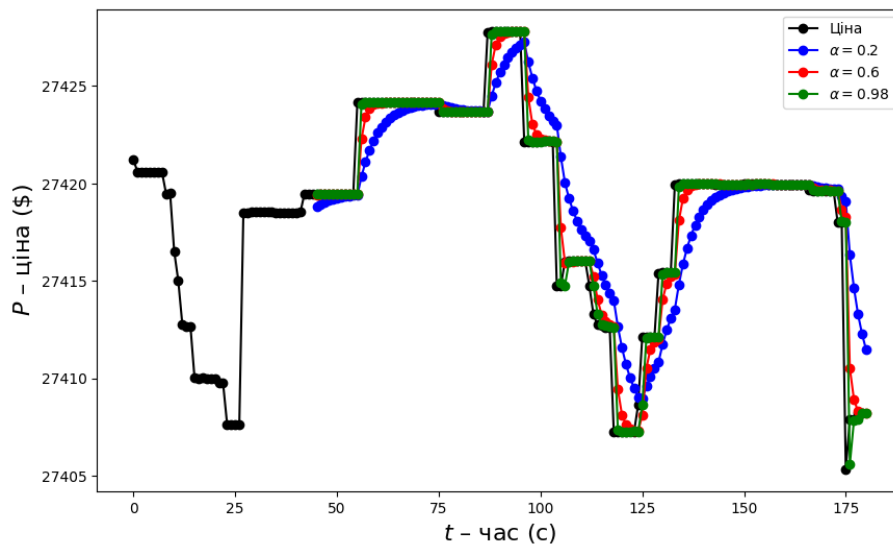


Рисунок 2. Результати використання методу простого експоненційного згладжування з розміром вікна в 45 с при різних значеннях  $\alpha$

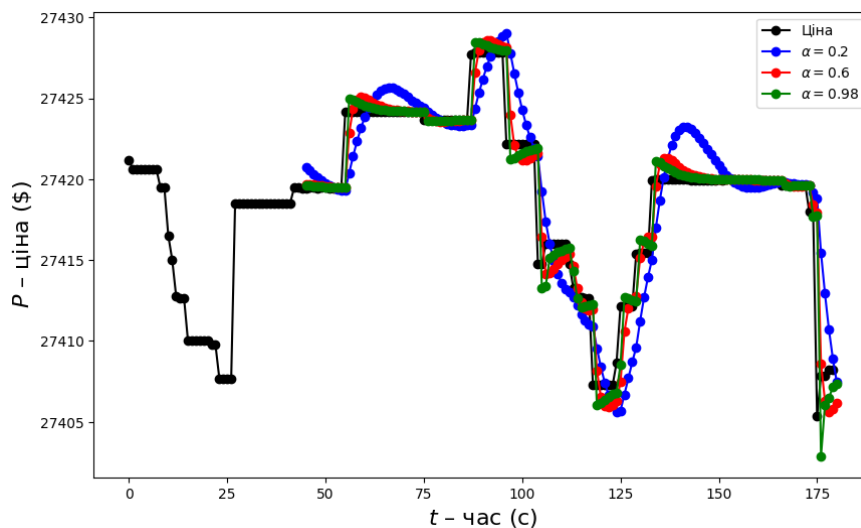


Рисунок 3. Результати використання методу Хольта з розміром вікна в 45 с при різних значеннях  $\alpha$

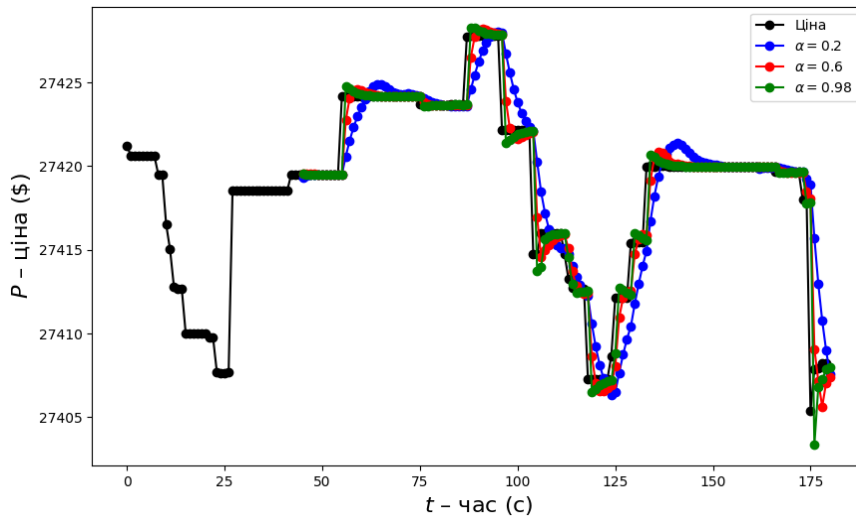


Рисунок 4. Результати використання методу адитивного згладжування тренду зі згасанням з розміром вікна в 45 с при різних значеннях  $\alpha$

### Результати досліджень

Для проведення експериментів використана бібліотека мови Python – statsmodels [7]. Для порівняння усіх можливих діапазонів результатів для кожного тесту використовувались однаковий вектор вхідних параметрів: розмір вікна – з 5 с до 65 с з кроком 5 с та  $\alpha$  – з 0,04 с до 0,98 с з кроком 0,04 с. Усі результати наведені у вигляді графіків: для простого експоненційного згладжування – рис. 5-7, для методу Хольта – рис. 8-10, для методу адитивного згладжування тренду – рис. 11-13.

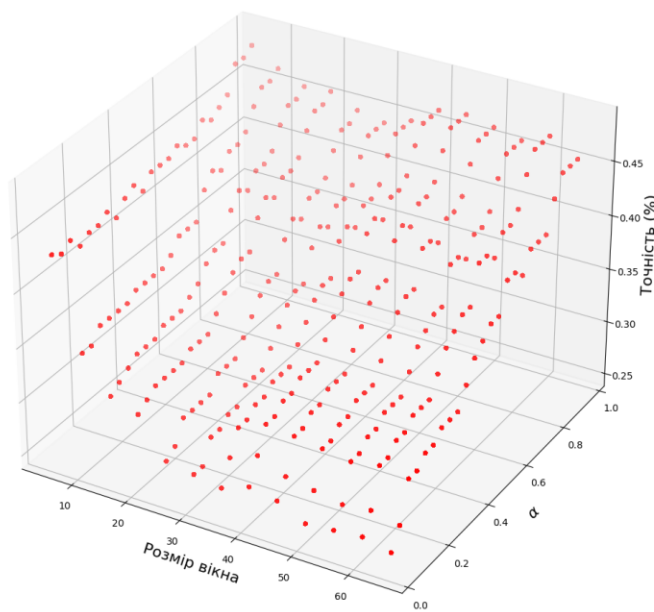


Рисунок 5. Точність методу простого експоненційного згладжування

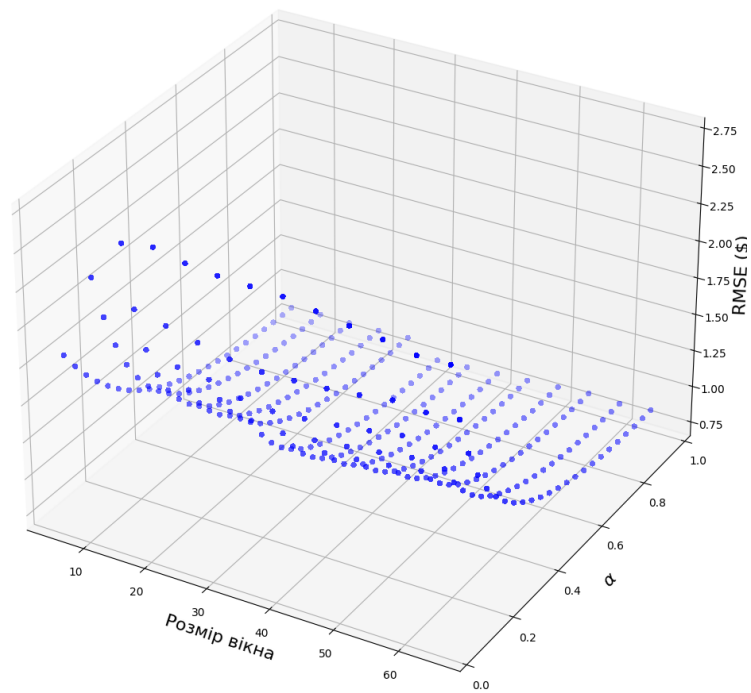


Рисунок 6. RMSE методу простого експоненційного згладжування

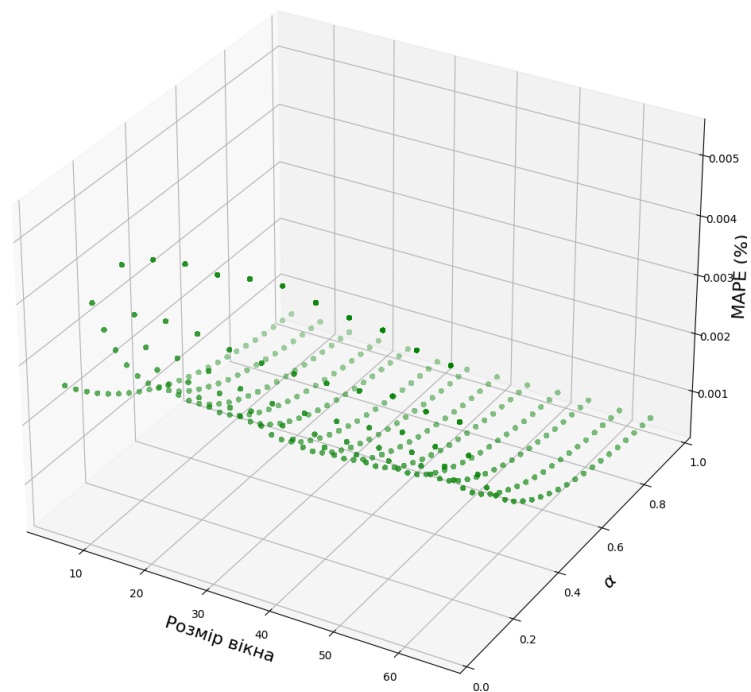


Рисунок 7. MAPE методу простого експоненційного згладжування

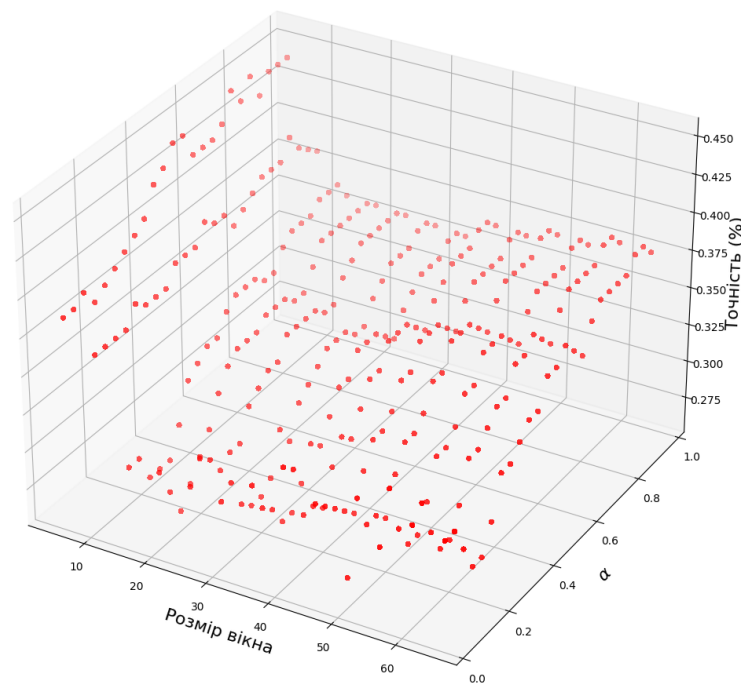


Рисунок 8. Точність методу Хольта

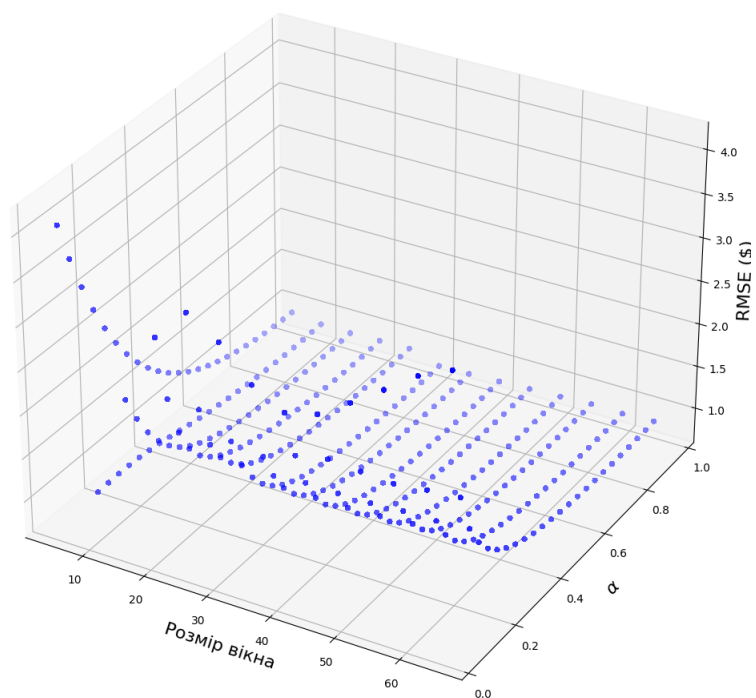


Рисунок 9. RMSE методу Хольта



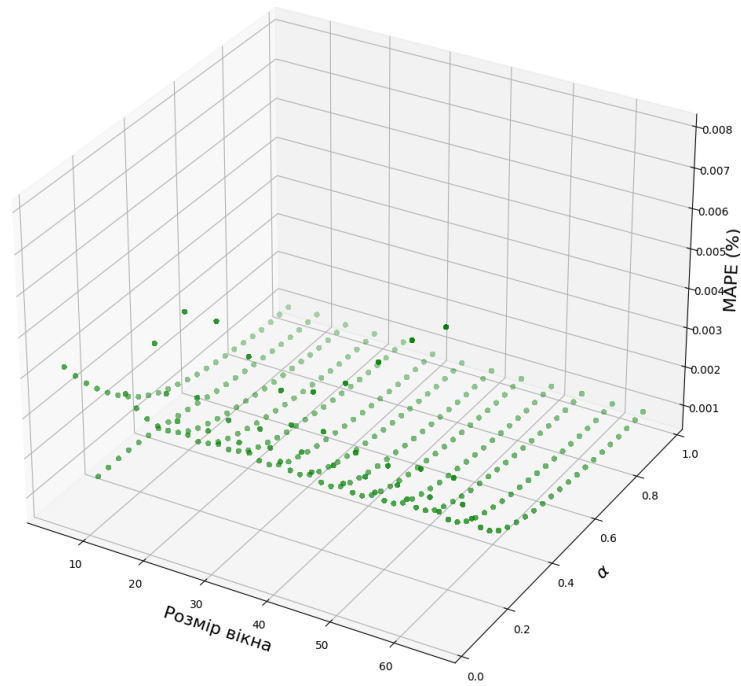


Рисунок 10. MAPE методу Хольта

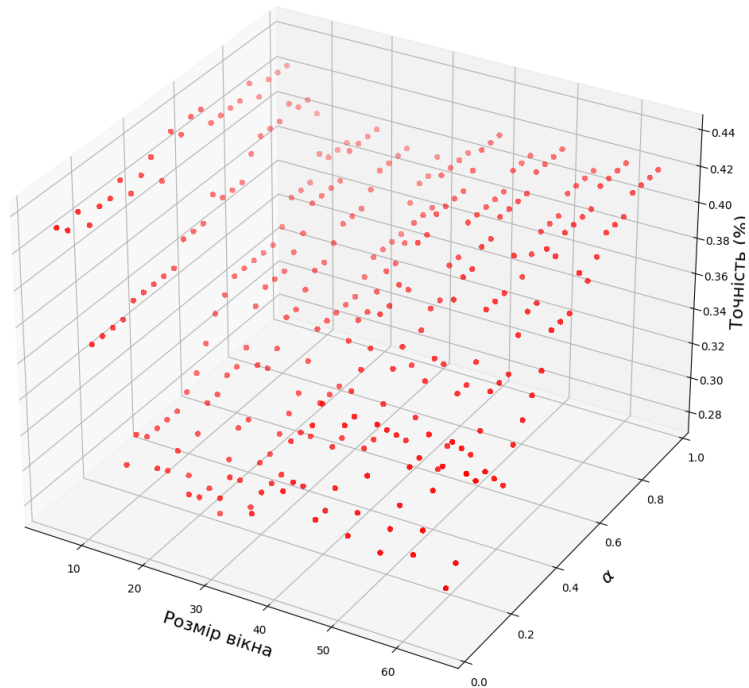


Рисунок 11. Точність методу адитивного згладжування тренду зі згасанням

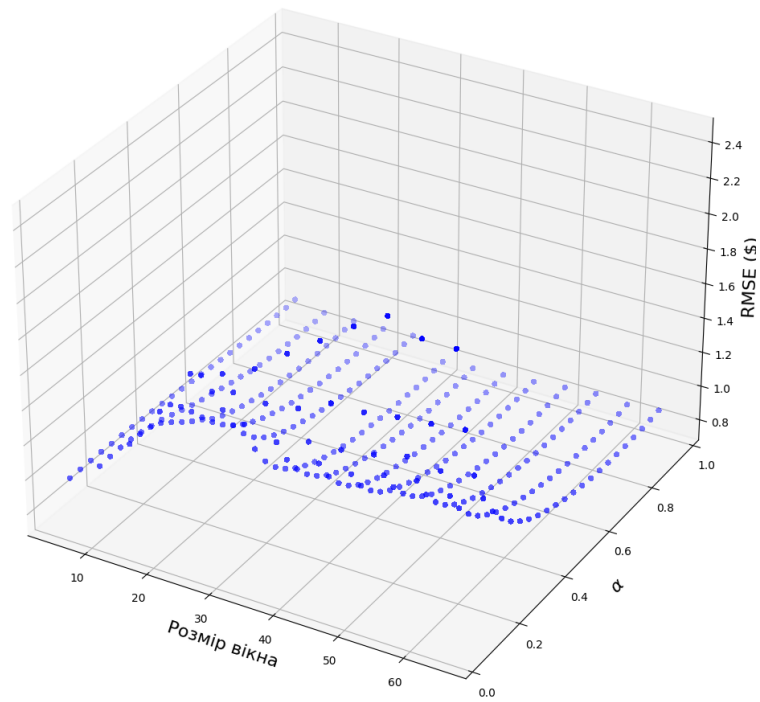


Рисунок 12. RMSE методу адитивного згладжування тренду зі згасанням

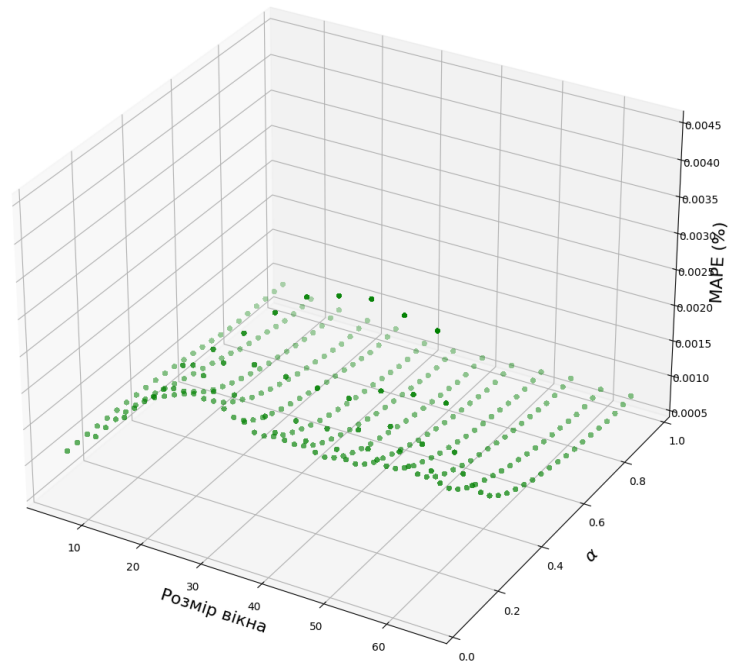


Рисунок 13. MAPE методу адитивного згладжування тренду зі згасанням

Статистичні результати дослідження наведено у таблиці.

Таблиця 1.

### Результати експериментів

Метод	Метод простого експоненційного згладжування	Метод Хольта	Метод адитивного згладжування тренду зі згасанням
Сер. точність	37,71 %	35,23 %	37,45 %
Макс. точність	48,27 %	44,82 %	43,67 %
Мін. точність	25,43 %	26,41 %	27,9 %
Сер. RMSE	\$0,97	\$0,97	\$0,93
Макс. RMSE	\$2,68	\$4,07	\$2,41
Мін. RMSE	\$0,78	\$0,82	\$0,79
Сер. MAPE	0,001 %	0,0012 %	0,001 %
Макс. MAPE	0,005 %	0,0078 %	0,0043 %
Мін. MAPE	0,0005 %	0,00087 %	0,00067 %

За результатами, наведеними у таблиці, видно, що метод Хольта гірший з трьох досліджуваних за характеристикою MAPE, яка відрізняється більш ніж на 20%. Метод простого експоненційного згладжування кращий за максимальним та середнім значенням точності прогнозування тренду, але є найгіршим за мінімальним, що вказує на те, що цей метод найбільше залежить від правильності підібраних параметрів  $\alpha$  та розміру вікна. Водночас метод адитивного згладжування тренду зі згасанням не так залежить від цих параметрів, заданих оператором, оскільки найбільші значення ТПТ та MAPE менші, а отже, кращі, ніж у попереднього методу.

### Висновки

Ефективне проведення високочастотної торгівлі можливе за умов швидкого виконання операцій, яке може забезпечити автоматизація процесу здійснення транзакцій. Розглянуто застосування методів експоненційного згладжування для короткочасного прогнозування ціни на криптобіржах. Для оцінки прогнозу ціни на криптовалюту проаналізовано три методи: точність прогнозування тренду, RMSE та MAPE.

За результатами дослідження встановлено, що прогноз ціни методом простого експоненційного згладжування дає найкращий результат. В нього максимальна точність прогнозу тренду серед усіх – 48,27% та найменше значення MAPE – 0.001. Крім того, з усіх методів він найшвидший, що є перевагою для високочастотної торгівлі, а для його роботи потрібно менше оперативної пам'яті, бо потрібно зберігати лише попереднє значення.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Clapham, B., Haferkorn, M. & Zimmermann, K. The Impact of High-Frequency Trading on Modern Securities Markets. *Bus Inf Syst Eng* 65, 7–24 (2023). <https://doi.org/10.1007/s12599-022-00768-6>.
2. Trevethan R (2017) Sensitivity, Specificity, and Predictive Values: Foundations, Plabilities, and Pitfalls in Research and Practice. *Front. Public Health* 5:307. <https://doi.org/10.3389/fpubh.2017.00307>.
3. A. Supriatna, D. Susanti, E. Hertini. Application of Holt exponential smoothing and ARIMA method for data population in West Java. *IOP Conf. Series: Materials Science and Engineering* 166 (2017) 012034 doi:10.1088/1757-899X/166/1/012034.
4. Chai, T. and Draxler, R. R.: Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, *Geosci. Model Dev.*, 7, 1247–1250, <https://doi.org/10.5194/gmd-7-1247-2014>, 2014.
5. Kline/Candlestick Data – Binance API Documentation. <https://binance-docs.github.io/apidocs/spot/en/#kline-candlestick-data>. Accessed 1 Apr. 2023.
6. Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2014.
7. Statistical models, hypothesis tests, and data exploration. <https://www.statsmodels.org/stable/index.html>. Accessed 1 Apr. 2023.
8. Charles C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, *International Journal of Forecasting*, Volume 20, Issue 1, 2004, Pages 5-10, ISSN 0169-2070, <https://doi.org/10.1016/j.ijforecast.2003.09.015>.
9. Giacomo Sbrana, Andrea Silvestrini, Forecasting with the damped trend model using the structural approach, *International Journal of Production Economics*, Volume 226, 2020, 107654, ISSN 0925-5273, <https://doi.org/10.1016/j.ijpe.2020.107654>.