

Рисунок 4 - Методи зниження рівня ризиків ПС

Отже, безпекові ризики є домінантними лише в областях, на території яких безпосередньо ведуться бойові дії та на Дніпропетровщині. Водночас, менеджмент підприємств у центральних, південних і західних областях найбільше хвилюють саме операційні ризики (нестача кадрів та перебої в електропостачанні), а безпекові ризики сприймаються на тому ж рівні, що і економіко-політичні (зокрема, законодавчі зміни). За галузевим поділом різниця у сприйнятті ризиків між промисловими та ІТ-компаніями пов'язана із форматом роботи працівників та безпеки даних. Найбільш поширеними методами зниження рівня ризиків є створення плану дій та інвестиції в рішення, до яких більш схильні компанії областей, на території яких ведуться бойові дії.

ПРОГНОЗУВАННЯ ДИНАМІКИ РИНКУ КРИПТОВАЛЮТ НА ОСНОВІ ДАНИХ СОЦІАЛЬНИХ МЕРЕЖ

Гур'янова Лідія, доктор економічних наук, професор, професор кафедри економічної кібернетики та прикладної економіки

Харківського національного університету імені В.Н. Каразіна

м. Харків, Україна

ORCID ID 0000-0002-2009-1451

e-mail: guryanovalidiya@gmail.com

Луценко Ростислав, аспірант кафедри економічної кібернетики та прикладної економіки

Харківського національного університету імені В.Н. Каразіна

м. Харків, Україна

ORCID ID 0000-0003-0737-3902

e-mail: roxanisen@gmail.com

У сучасній економіці дослідження на основі даних із соціальних мереж набувають все більшої популярності. Інтерактивні платформи, такі як тікток, фейсбук, інстаграм, ікс (твіттер), лінкедин та інші, генерують велику кількість даних, які можуть бути використані для аналізу поведінки користувачів. Обробка великих обсягів інформації вимагає розробки ефективних інструментів їх збору та аналізу. Прикладний програмний інтерфейс (API) є одним з ключових методів, який забезпечує доступ до даних у реальному часі. Прикладний програмний інтерфейс дозволяє науковцям автоматизувати процес збору інформації, налаштовувати запити за різними критеріями та отримувати структуровані дані для подальшого аналізу.

Науковці активно досліджують використання API соціальних мереж для збору та аналізу даних, прогнозування динаміки ринку криптовалют за допомогою методів інтелектуального аналізу даних. Так, Poudel, Weninger аналізували проблеми пов'язані з доступом до соціальних мереж [1]. Розробка та керування прикладним програмним інтерфейсом привертає значну увагу розробників програмного забезпечення [2]. Публікація “Advances in Intelligent Data Analysis and Its Applications” Zhang, C., Li, W., Zhang, H., & Zhan, T. містить огляд теоретичних основ методів інтелектуального аналізу даних і пропонує практичні реко-

мендації для підвищення точності та швидкості аналізу [3]. У дослідженні Mudassir, M., Bennbaia "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach" автори прогнозують ціну на біткоїн за допомогою моделей машинного навчання, визначають точність прогнозування для короткострокового та середньострокового періоду упередження [4].

Метою дослідження є проведення комплексного аналізу моделей для прогнозування динаміки ринку криптовалют за даними соціальних мереж. Відповідно до мети дослідження основними завданнями є:

- зібрати базу даних про активність користувачів у соціальних мережах;
- виконати обробку, агрегацію даних;
- побудувати моделі прогнозування на основі поліноміальної регресії, а також методів машинного навчання (Random Forest, Decision Tree) та Prophet;
- оцінити точність та ефективність моделей за допомогою метрик Mean Squared Error та Mean Absolute Error.

Збір інформації з соціальних мереж включав формування тематичних запитів. Використання структурованих запитів дозволило ефективно вилучати дані, релевантні до досліджуваної теми, чим забезпечило репрезентативність вибірки для аналізу.

Парсинг включав перетворення необробленої інформації на структуровані формати, придатні для подальшого аналізу. Коректність обробки даних забезпечується застосуванням сучасних бібліотек Python, таких як requests та json.

Код для обробки включав адаптацію алгоритмів збору інформації до специфіки кожної соціальної мережі. Наступний етап - формування термінологічних баз із ключових слів, фраз та синонімів. Розробка програмного забезпечення для автоматизації створення запитів до API здійснювалась із застосуванням HTTP-методу POST. Розробка скриптів для перевірки статусу та результатів виконання запитів, зроблених до API, здійснювалась з використанням HTTP-методу GET. Зроблено автоматизацію перенесення даних із хмарної бази PostgreSQL (через PGAdmin) до локальних файлів формату CSV для подальшої обробки. Далі - оптимізація циклічного виконання запитів із врахуванням обмежень швидкості API та розробка алгоритмів для регулювання частоти запитів із використанням таймерів (time.sleep) або черг завдань для забезпечення відповідності правилам використання API. Реалізація скриптів для обробки, візуалізації та аналізу даних кожної соціальної мережі окремо включає методи частотного аналізу та аналізу настроїв із використанням бібліотек Python (NLTK, scikit-learn, matplotlib).

Система моніторингу збору даних дозволяє інтегрувати отримані дані в аналітичні та дослідницькі процеси. Така універсальність даної системи є перспективною для подальшого використання в бізнесі, науці та освіті. Досліджено варіанти інтеграції отриманих даних для описового аналізу, кількісного аналізу, лінгвістичного аналізу, аналізу часових рядів, мережевого аналізу, семантичного аналізу. Слід також зазначити, що отримані таким чином дані дозволяють здійснити детальний контент-аналіз на кшталт визначення найпопулярніших тегів, топ-постів по окремим соціальним мережам, топ найпопулярніших емодзі. Ефективним є їх використання для категоризації та візуалізації медіа-контенту.

Серед моделей прогнозування динаміки криптовалют саме модель дерева рішень продемонструвала хорошу точність. MSE становить 2,532,048.20, що свідчить про порівняно низький рівень розбіжностей між прогнозованими та фактичними значеннями. Коефіцієнт детермінації R^2 підтверджує, що 79.3% варіації ціни біткоїна пояснюється побудованою моделлю, що свідчить про її адекватність моделі та доцільність врахування даних соціальних мереж під час прогнозування ринку криптовалют.

Побудована модель дерева рішень продемонструвала певний успіх у прогнозуванні цін на біткоїн, проте зросли відхилення у прогнозах з часом. Тому в ході дослідження зроблено модель випадкового лісу для врахування більш складних взаємозв'язків між змінними (рис. 1).

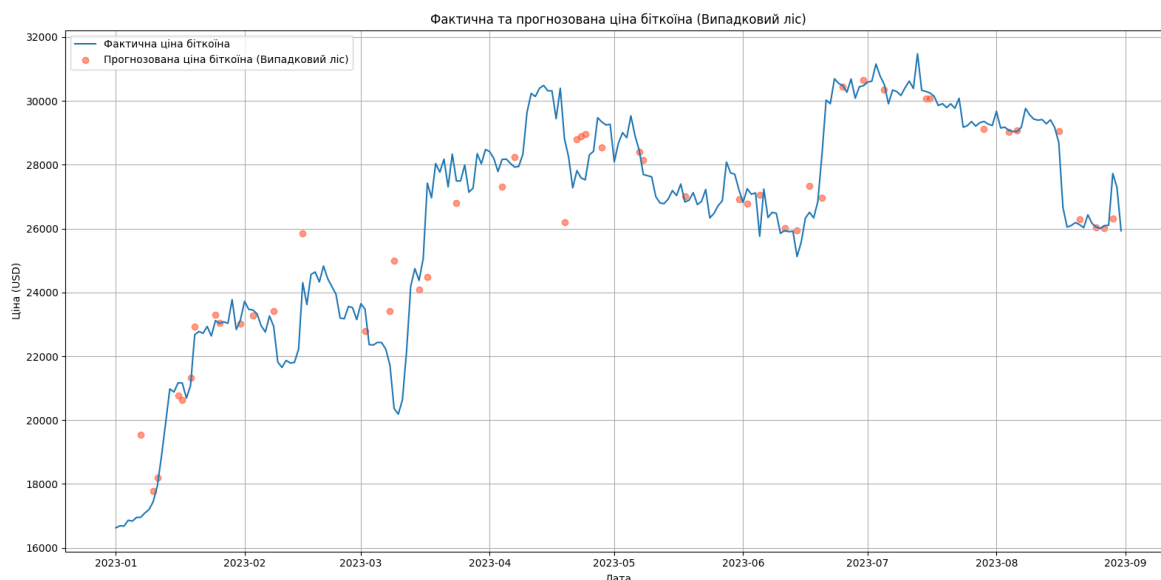


Рисунок 1 – Випадковий ліс для аналізу динаміки ціни біткоїна

Графік прогнозування (рис. 1) ілюструє фактичну ціну біткоїна (синя лінія) та прогнозовані значення, отримані за допомогою моделі випадкового лісу (помаранчеві точки). Оцінка моделі показує, що MSE становить 1,327,488.83, що свідчить про відносно невелике середнє відхилення прогнозованих значень від фактичних. R^2 демонструє здатність моделі пояснювати 89.1% варіації цін на біткоїн. Результати моделювання підтверджують ефективність випадкового лісу як методу для прогнозування цін на біткоїн, враховуючи активність у соціальних мережах.

Аналіз якості моделі прогнозування Prophet цін біткоїна на основі даних з п'яти соціальних мереж демонструє здатність відтворювати загальний тренд, однак на деяких відрізках спостерігаються відхилення між прогнозованими та фактичними цінами. На графіку (рис. 2) представлено три основні лінії: фактичні ціни біткоїна (помаранчева), прогнозовані ціни моделі (синя) та фактичні ціни за жовтень (червона).

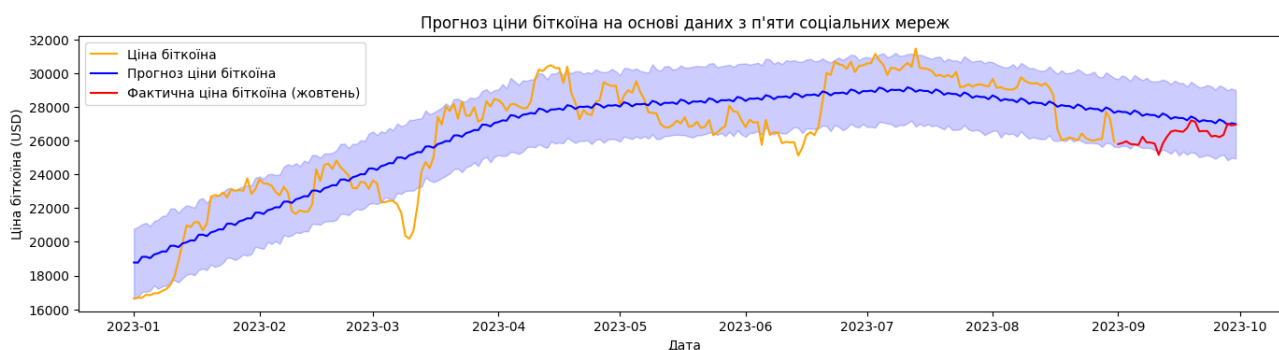


Рисунок 2 – Модель Prophet для аналізу динаміки ціни біткоїна

MAE вказує на те, що ціни відхиляються від фактичних на 1078.96 доларів, що є прийнятним результатом. MSE - 1,595,962.71, свідчить про наявність значних помилок у деяких прогнозах, що є результатом впливу різких змін на ринку біткоїна. RMSE - 1263.31. Значення RMSE на рівні понад тисячу доларів вказує на розподіл помилок, що свідчить про суттєві відхилення від фактичних значень у певних періодах.

Таким чином, у результаті дослідження зібрано та агреговано базу даних соціальних мереж, побудовані моделі прогнозування: поліноміальна регресія, Random Forest, Decision Tree та модель Prophet. Проведений аналіз підтвердив вплив активності користувачів у соціальних мережах на динаміку ринку криптовалют. Сплески обговорень часто передують

змінам цін на криптовалюти. Проведено оцінку точності та ефективності кожної моделі за допомогою метрик MSE та MAE. Результати показали, що модель Prophet продемонструвала найвищу точність прогнозування, з найнижчими значеннями MSE та MAE. У подальших дослідженнях будуть враховані поведінкові фактори для підвищення точності прогнозів.

Література:

1. Poudel, A., & Weninger, T. (2024). Navigating the post-API dilemma: Search engine results pages present a biased view of social media data. *arXiv*. <https://doi.org/10.48550/arXiv.2401.15479>
2. De, B. (2023). *API management: An architect's guide to developing and managing APIs for your organization* (2nd ed.). Apress. <https://doi.org/10.1007/979-8-8688-0054-2>
3. Zhang, C., Li, W., Zhang, H., & Zhan, T. (2024). Recent advances in intelligent data analysis and its applications. *Electronics*, 13(1), 226. <https://doi.org/10.3390/electronics13010226>
4. Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: A machine learning approach. *Neural Computing and Applications*, 32(20), 15869–15888. <https://doi.org/10.1007/s00521-020-05129-6>

ЗАСТОСУВАННЯ МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ОЦІНКИ РИЗИКІВ ПІДПРИЄМНИЦЬКИХ СТРУКТУР

Лідія Гур'янова, д.е.н., проф.
Харківський національний університет ім. В.Н. Каразіна
ORCID ID <https://orcid.org/0000-0002-2009-1451>
e-mail: guryanovalidiya@gmail.com

Чала Ангеліна, магістр
Харківський національний економічний університет ім. С. Кузнеця
e-mail: angelina02chalaya@gmail.com

Сучасні умови функціонування українських підприємницьких структур характеризується значним ступенем ризику, пов'язаним з низьким рівнем стійкості енергетичної системи, інфляційними процесами, зниженням платоспроможного попиту та ін. Фактори ризику оказують асиметричний вплив на діяльність підприємницьких структур (ПС) різних регіонів, масштабу бізнесу, сфер діяльності. Тому цікавим науково-дослідним завданням є оцінка такого асиметричного впливу ризиків на ПС різного типу, що дозволяє визначити найбільш вразливі сфери діяльності та застосувати превентивні механізми фінансової стабілізації галузі.

Для розробки моделі класифікації використовувались методи ієрархічної агломеративної кластеризації (метод Уорда) та ітеративної кластеризації (метод k-means). Інформаційною базою дослідження є фінансові дані українських підприємств за такими показниками, як прибуток, рентабельність активів, коефіцієнт поточної ліквідності, рівень заборгованості та ін., а також класифікаційні ознаки ПС: масштаб бізнесу, регіональне розташування та відстань до зони бойових дій та ін. Під час побудови моделі кластерного аналізу дані були стандартизовані.

Дендрограма кластеризації за методом Уорда (рис. 1), графік «ліктя» та інші критерії, що дозволяють досліджувати зміну функціоналів якості класифікації в залежності від кількості кластерів, яка враховується в моделі класифікації, дозволили зробити висновок про наявність 7 груп в досліджуваній сукупності.