

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
Факультет комп'ютерних наук, фізики та математики
Кафедра комп'ютерних наук та програмної інженерії

Розробка системи визначення впливу новин на зміну цін
фінансових інструментів

Кваліфікаційна робота (проект)
на здобуття ступеня вищої освіти «магістр»

Виконав: студент 2 курсу 261М групи

Спеціальності

126 «Інформаційні системи та
технології»

(шифр, назва)

Освітньо-професійної програми:
«Інформаційні системи та технології»

(назва)

Штанг Нікіта Денисович

Керівник: доктор економічних наук,
професор Кобець В.М.

Рецензент: Калініченко І.В.
Senior Softwareengineer
ІТ компанії Softserve

Івано Франківськ – 2024

ВСТУП.....	1
РОЗДІЛ 1.	3
ІСНУЮЧІ СИСТЕМИ ДЛЯ ПРОГНОЗУВАННЯ ФОНДОВОГО РИНКУ	4
1.1. Investing.com	4
1.2. MarketWatch	5
1.3. Yahoo Finance	7
1.4. Bloomberg Terminal	7
1.5. Аналіз обмежень існуючих систем	8
РОЗДІЛ 2.	8
ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ІНСТРУМЕНТІВ.....	9
2.1. Лінійні моделі для прогнозування цін акцій	9
2.1.1. Лінійна регресія	9
2.1.2. Модель авторегресії (AR)	10
2.1.3. Модель авторегресії ковзного середнього (ARMA)	11
2.1.4. Авторегресійна інтегрована модель ковзного середнього (ARIMA)	12
2.2. Нейронні мережі для прогнозування цін акцій.....	12
2.2.1. Багатошаровий перцептрон (MLP)	13
2.2.2. Рекурентні нейронні мережі (RNN)	14
2.2.3. Рекурентні мережі з довготривалою пам'яттю (LSTM)	14
2.2.4. Згорткові нейронні мережі (CNN) для обробки часових рядів	15
2.2.5. Комбінація LSTM та CNN для фінансових прогнозів	16
2.3. Аналіз настроїв ринку за допомогою ШІ	17
2.4. Управління інвестиційним портфелем за допомогою ШІ	18
2.5. Виявлення шахрайства та аномалій у фінансових даних	20
2.6. Використання штучного інтелекту для поліпшення точності прогнозів	22
2.7. Аналіз настроїв з традиційними фінансовими індикаторами	24
РОЗДІЛ 3	27
ВЛАСНА СИСТЕМА ДЛЯ ВИЗНАЧЕННЯ ВПЛИВУ НОВИН НА ФІНАНСОВІ ІНСТРУМЕНТИ	27
3.1.1. MongoDB	27
3.1.2. .NET	29
3.4. Результати.....	35

3.4.1. Вплив настроїв на зміну цін акцій	35
3.4.2. Порівняння результатів для моделей ARIMA та ARIMA + Sentiment	36
3.4.3. Статистична значущість результатів	36
3.4.4. Переваги та обмеження запропонованої моделі	36
ВИСНОВКИ	38
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	41

ВСТУП

Актуальність роботи. У сучасних умовах швидкого зростання обсягів інформації та застосування штучного інтелекту, ринок капіталу вимагає ефективних інструментів для аналізу. Значна кількість нових інвесторів і трейдерів використовують торгові платформи, такі як *Investing.com* та *MarketWatch*, для моніторингу фондового ринку та отримання новин. Проте, традиційні сервіси часто обмежуються відстеженням коливань цін без належного врахування впливу настроїв, що створює ризик прийняття неефективних рішень. Враховуючи значний обсяг даних і швидкоплинність ринкових змін, ручний аналіз новин є малоефективним. Це призводить до затримок у прийнятті рішень і втрати можливостей для інвесторів.

Автоматизовані системи, що інтегрують аналіз настроїв ринку в режимі реального часу, можуть підвищити швидкість і точність рішень на основі новин. У зв'язку з цим, актуальним є створення програмного сервісу, який дозволяє прогнозувати ціни акцій, комбінуючи кількісні (історичні дані про ціни) та якісні (аналіз настроїв) індикатори. Такий підхід сприятиме ухваленню більш обґрунтованих інвестиційних рішень та дозволить інвесторам краще орієнтуватися в динаміці фінансових ринків.

Об'єкт дослідження - інформаційні системи аналізу курсу фінансових інструментів.

Предмет дослідження - прогнозування цін акцій на основі аналізу новинних настроїв і авторегресійних моделей.

Мета дослідження - розробити програмний сервіс, який дозволяє прогнозувати напрямок змін цін акцій, враховуючи настрої інвесторів та історичні дані про ціни акцій.

Завдання:

- Провести огляд наукових досліджень у сфері прогнозування цін акцій із застосуванням методів аналізу настроїв та авторегресійних моделей.

- Дослідити архітектуру та компоненти систем для збору даних, включаючи API для отримання інформації про ціни акцій і новини.
- Описати методи та алгоритми, що використовуються для аналізу настроїв новин, включаючи FinBERT.
- Розробити модель прогнозування на основі ARIMA та оцінити її точність у поєднанні з аналізом настроїв для вибірки компаній.
- Визначити переваги та недоліки запропонованого сервісу в порівнянні з традиційними аналітичними інструментами.

Методи та технології дослідження - методи машинного навчання, зокрема аналіз настроїв за допомогою FinBERT, а також авторегресійна модель ARIMA для прогнозування цін. Для збору та зберігання даних застосовуються API та база даних MongoDB. Також використана мова програмування C#

Апробація.

Опублікована стаття у фаховому науковому журналі Herald of Advanced Information Technology (категорія Б):
<https://hait.od.ua/index.php/journal/article/view/229/247>

Структура роботи - робота складається зі вступу, трьох розділів, висновків та списку використаних джерел. У першому розділі аналізуються існуючі дослідження та методології у сфері прогнозування цін акцій на основі настроїв. Другий розділ присвячений опису архітектури програмного сервісу для збору даних і аналізу настроїв. Третій розділ описує розробку моделі прогнозування, експериментальне застосування сервісу та аналіз результатів для прийняття інвестиційних рішень.

РОЗДІЛ 1.

ІСНУЮЧІ СИСТЕМИ ДЛЯ ПРОГНОЗУВАННЯ ФОНДОВОГО РИНКУ

На сьогодні існує ряд платформ і сервісів для моніторингу фондового ринку, що забезпечують користувачів інструментами для відстеження змін на ринку, аналізу історичних даних та, іноді, новинних настроїв. Серед найбільш популярних платформ є *Investing.com*, *MarketWatch*, *Yahoo Finance*, та *Bloomberg Terminal*. Кожен з цих сервісів має свої переваги, однак вони мають і обмеження, які відкривають можливості для вдосконалення аналітичних підходів.

1.1. Investing.com

Investing.com є одним із найпопулярніших сервісів для фінансових аналітиків та інвесторів, що надає доступ до реальних котирувань, останніх ринкових новин, інтерактивних графіків та аналітичних інструментів для широкого спектра фінансових інструментів, зокрема акцій, облігацій, валютних курсів, товарів і криптовалют. Завдяки зручному інтерфейсу та безлічі доступних інструментів, користувачі можуть швидко знаходити інформацію, яка їх цікавить, і оперативно відслідковувати зміни на ринку. Платформа пропонує реальний час оновлення новин і котирувань, а також доступ до ринкового календаря, що містить дані про найважливіші економічні події, такі як вихід звітів, зміни у ставках, заяви центральних банків та інші події, що можуть суттєво вплинути на рух цін. Це робить *Investing.com* зручним інструментом для тих, хто прагне ухвалювати рішення на основі оперативної ринкової інформації.

Крім цього, *Investing.com* пропонує користувачам інструменти для технічного аналізу ринку, які допомагають у прийнятті інвестиційних рішень. Завдяки графікам та індикаторам, які включають такі популярні інструменти, як ковзні середні, індекс відносної сили (RSI), лінії підтримки і спротиву, інвестори можуть проводити аналіз тенденцій та оцінювати поточний стан активів на ринку.

Такий підхід до аналітики забезпечує можливість для розробки базових стратегій трейдингу.

Однак, попри багатий набір функцій, Investing.com (рис. 1.1) має суттєві обмеження, коли йдеться про аналіз настроїв і виявлення причинно-наслідкових зв'язків між новинами та рухом цін. Платформа не включає можливості для повноцінного аналізу новинного настрою, що знижує її аналітичну глибину. Наприклад, у випадках, коли на ринок впливають емоційні фактори, такі як оптимістичні або песимістичні настрої, Investing.com не здатен точно оцінити, який безпосередній вплив мають новини на зміни ціни. Це обмеження знижує можливості інвесторів для розуміння справжніх мотивів руху ринку та потенційного впливу останніх новин на цінові коливання. Відсутність аналітики настроїв робить платформу менш зручною для тих, хто прагне враховувати не лише історичні дані, але й психологічний аспект ринку для комплексного прогнозування тенденцій.



Рис. 1.1. Віджет новин Investing.com

1.2. MarketWatch

MarketWatch є популярним джерелом інформації про ринок капіталу та макроекономічні події, надаючи користувачам доступ до актуальних новин, аналітики та даних про котирування акцій, облігацій, товарів та валют. Окрім загальної ринкової інформації, платформа має функцію “referenced symbols feature”, яка дозволяє користувачам відстежувати вплив конкретних новин на котирування окремих акцій або інших фінансових інструментів. Ця функція є особливо корисною для інвесторів, які прагнуть оцінити загальну ринкову тенденцію та зрозуміти, як окремі події впливають на ринок у цілому. Вона дає можливість краще орієнтуватися в динаміці цін, що може бути корисним для ухвалення обґрунтованих рішень щодо купівлі або продажу активів.

MarketWatch також надає своїм користувачам доступ до широкого спектра фінансових інструментів для аналізу ринку, зокрема до технічних показників та макроекономічних оглядів, що дозволяє створювати більш цілісну картину економічної ситуації. Інформація про корпоративні звіти, зміни у ставках центральних банків та інші важливі новини може бути відфільтрована відповідно до певних символів, що полегшує виявлення важливих подій для конкретних активів. Це дає можливість інвесторам оперативно відстежувати зміни на ринку і розуміти, як загальний економічний контекст може впливати на їхні інвестиції.

Проте, подібно до платформи Investing.com, MarketWatch не надає функції аналізу настроїв, що обмежує її аналітичну ефективність у прогнозуванні. Відсутність інструментів для визначення настроїв у новинних повідомленнях робить MarketWatch менш корисним для розуміння того, чи саме конкретна новина стала основною причиною зміни ціни акцій. Відсутність причинно-наслідкового аналізу також ускладнює оцінку безпосереднього впливу окремих новин на ринкові рухи, що знижує можливість повного розуміння впливу новинного контексту на цінові коливання. Таким чином, MarketWatch надає

лише базову інформацію про ринок без можливості глибшого аналізу настроїв, що обмежує його потенціал як повноцінного аналітичного інструменту для прийняття інвестиційних рішень (рис. 1.2).

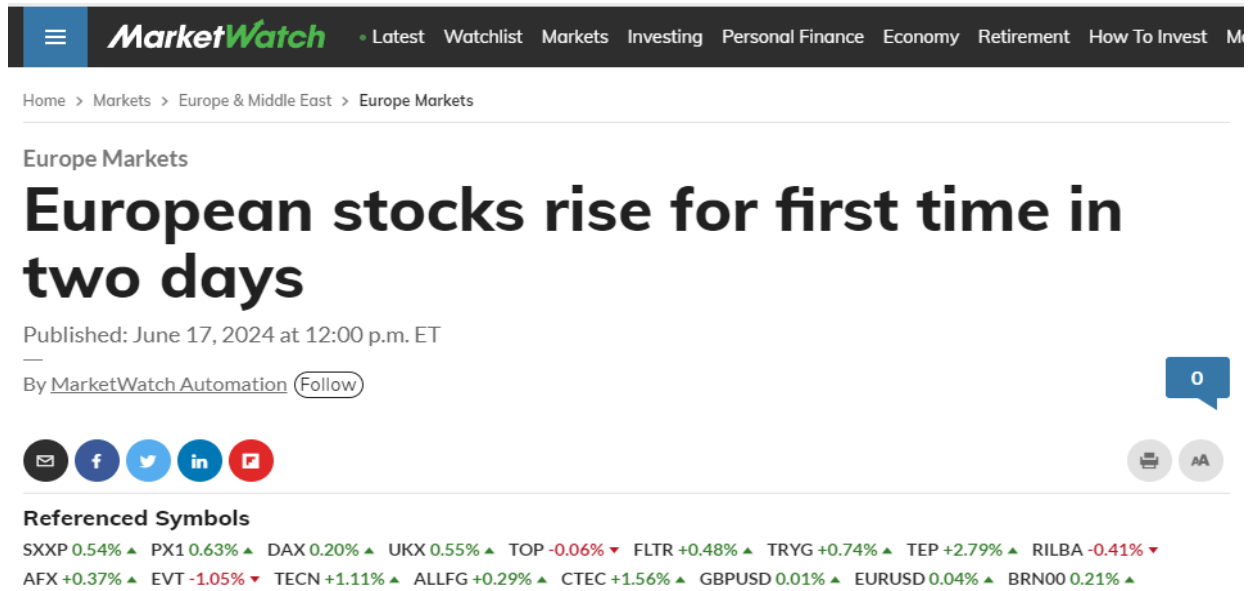


Рис. 1.2. Приклад новинної статті Market Watch

1.3. Yahoo Finance

Yahoo Finance забезпечує широкий доступ до ринкової інформації, включаючи історичні дані, індекси, графіки і фінансові новини. Ця платформа надає можливість використовувати базові інструменти для аналізу акцій і порівняння фінансових показників. Тим не менш, платформа не включає комплексних можливостей для аналізу настроїв, що є важливим фактором для короткострокового прогнозування цін акцій.

1.4. Bloomberg Terminal

Bloomberg Terminal є професійною системою для фінансових аналітиків, яка забезпечує широкий доступ до даних, новин, а також інструментів для фундаментального й технічного аналізу. Крім того, платформа використовує алгоритми машинного навчання для обробки

великих обсягів даних, що дозволяє користувачам отримувати більш глибокий аналітичний інсайт. Водночас вартість Bloomberg Terminal є дуже високою, що обмежує його доступність для середнього інвестора.

1.5. Аналіз обмежень існуючих систем

Попри широкий спектр функцій, існуючі системи мають спільні обмеження, зокрема:

1. **Відсутність аналізу настроїв новин у реальному часі:** Більшість платформ лише відображають новини, але не здійснюють повноцінного аналізу їхнього впливу на ринок, що обмежує здатність інвесторів прогнозувати зміни.
2. **Недостатність причинно-наслідкових зв'язків між новинами та цінами:** Системи не надають чіткого аналізу, який би показував прямий вплив настроїв новин на рух цін. Це робить прийняття рішень менш точним.
3. **Висока вартість (Bloomberg Terminal):** Доступ до платформ, що пропонують просунутий аналіз, часто є обмеженим через високу вартість передплати, що не дозволяє широкому колу користувачів скористатися їхніми можливостями.

Таким чином, існує потреба у сервісі, який міг би поєднувати функції аналізу настроїв новин у реальному часі з можливістю прогнозування змін цін акцій на основі історичних даних. Це дозволить інвесторам приймати більш обґрунтовані та оперативні рішення на фондовому ринку.

РОЗДІЛ 2.

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ІНСТРУМЕНТІВ

Штучний інтелект має значний вплив на розвиток фінансових технологій, зокрема, завдяки своїй здатності обробляти великі обсяги даних та виявляти приховані закономірності. У сфері прогнозування цін акцій використовуються різноманітні алгоритми, які допомагають аналізувати історичні дані, виявляти тренди і навіть враховувати емоційні аспекти ринку. Серед методів, що найчастіше застосовуються для прогнозування цін, є лінійні моделі, нейронні мережі та архітектури з довготривалою пам'яттю.

2.1. Лінійні моделі для прогнозування цін акцій

Лінійні моделі є одними з найпростіших та найзрозуміліших інструментів для прогнозування цінових змін на фінансових ринках. Вони дозволяють встановлювати залежності між змінними, що значно полегшує інтерпретацію результатів і розуміння трендів. Однак ці моделі мають обмеження, коли мова йде про складніші фінансові ринки з нелінійною поведінкою.

2.1.1. Лінійна регресія

Лінійна регресія є одним із найпростіших методів прогнозування, який передбачає, що залежна змінна (ціна акцій) є лінійною функцією незалежної змінної (наприклад, часу або деяких макроекономічних факторів). Основна форма рівняння лінійної регресії виглядає так:

$$y = \beta_0 + \beta_1 x + \epsilon$$

де: y – прогнозоване значення залежної змінної (наприклад, ціни акцій),

β_0 – вільний член (перетин із віссю ординат),

β_1 – коефіцієнт нахилу, що показує вплив незалежної змінної x на залежну y ,

ϵ – похибка моделі.

Лінійна регресія добре підходить для базових прогнозів і є ефективною, коли дані мають чітко виражений лінійний тренд. Проте у фінансових даних, які зазвичай демонструють більш складні, нелінійні тренди, лінійна регресія може показати неточні результати.

Переваги:

- Простота та швидкість обчислень.
- Легка інтерпретація коефіцієнтів, що дозволяє легко аналізувати вплив незалежних змінних.

Недоліки:

- Неможливість моделювання нелінійних залежностей.
- Чутливість до мультиколінеарності (коли незалежні змінні корельовані між собою).
- Обмежена здатність до обробки складних ринкових даних.

2.1.2. Модель авторегресії (AR)

Модель авторегресії (AR) використовується для прогнозування часових рядів, де значення змінної залежать від її попередніх значень. Наприклад, модель AR(1) виглядає так:

$$y_t = \phi y_{t-1} + \epsilon_t$$

де: y_t – значення змінної на момент часу t ,

y_{t-1} – значення змінної на попередньому кроці $t-1$,

ϕ – коефіцієнт авторегресії, який показує, наскільки сильно попереднє значення впливає на поточне,

ϵ_t – похибка моделі.

Модель AR є корисною для прогнозування цін акцій, які залежать від попередніх трендів. Вона добре працює з короткостроковими даними та дозволяє виявляти короткострокові кореляції. Моделі AR можуть бути розширені до вищих порядків, наприклад, AR(2) чи AR(p), де p – кількість попередніх значень, які враховуються в моделі.

Переваги:

- Легкість у налаштуванні та інтерпретації.
- Добре працює з даними, які мають автокореляцію.

Недоліки:

- Не враховує вплив зовнішніх факторів, окрім власного історичного значення.
 - Чутливість до вибору порядку моделі p, що може вимагати додаткових тестів для його оптимального визначення.

2.1.3. Модель авторегресії ковзного середнього (ARMA)

Модель ARMA об'єднує дві складові – авторегресію (AR) та ковзне середнє (MA). Це дозволяє моделювати не лише залежність від попередніх значень, але й враховувати вплив випадкових похибок:

$$y_t = \phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$$

де: θ – коефіцієнт для компонентів ковзного середнього, що враховує вплив попередніх похибок на поточне значення.

ARMA моделі часто застосовуються для стаціонарних часових рядів і можуть забезпечити більш точні прогнози порівняно з простою AR моделлю.

Переваги:

- Забезпечує точніший прогноз завдяки врахуванню як попередніх значень, так і попередніх похибок.
- Гнучкість у налаштуванні параметрів.

Недоліки:

- Вимагає, щоб дані були стаціонарними (тобто, середнє та дисперсія не змінювалися з часом), що обмежує застосування моделі на фінансових ринках, де ціни можуть мати тренди.

•

2.1.4. Авторегресійна інтегрована модель ковзного середнього (ARIMA)

Модель ARIMA є розширенням ARMA і застосовується для нестационарних часових рядів, що є більш характерним для фінансових даних. ARIMA додає етап диференціювання, щоб перетворити нестационарні ряди на стаціонарні:

$$y_t = \phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$$

де диференціювання дозволяє видалити тренд і коливання в середньому, що робить дані стаціонарними.

ARIMA моделі широко використовуються у фінансовому прогнозуванні, оскільки вони здатні адаптуватися до трендів, сезонності та високої волатильності цін акцій.

Переваги:

- Можливість роботи з нестационарними даними, що є критичним для фінансових ринків.
- Враховує короткострокові коливання та тренди, що дозволяє моделювати ціни з урахуванням коливань.

Недоліки:

- Складність у виборі параметрів (авторегресія, диференціювання, ковзне середнє), що вимагає додаткового налаштування.
- Чутливість до волатильності ринку, що може знижувати точність прогнозів в умовах раптових ринкових змін.

2.2. Нейронні мережі для прогнозування цін акцій

Нейронні мережі — це обчислювальні моделі, що імітують роботу людського мозку, дозволяючи обробляти великі обсяги даних і виявляти складні залежності між змінними. Вони особливо корисні для фінансового

прогнозування, оскільки здатні моделювати нелінійні взаємозв'язки, обробляти часові ряди та адаптуватися до динамічних змін на ринку.

2.2.1. Багатошаровий перцептрон (MLP)

Багатошаровий перцептрон (MLP) є базовою архітектурою нейронної мережі, що складається з вхідного шару, одного або кількох прихованих шарів та вихідного шару. Кожен нейрон у прихованих шарах використовує нелінійну активаційну функцію для трансформації вхідних даних, що дозволяє моделювати складні залежності.

Принцип роботи:

- Вхідний шар приймає числові дані, наприклад, ціни акцій або інші індикатори.
- Приховані шари здійснюють нелінійну обробку, що дозволяє моделювати складні зв'язки між вхідними значеннями.
- Вихідний шар генерує прогнозне значення (наприклад, майбутню ціну акції).

Переваги:

- Простота та ефективність при обробці невеликих обсягів даних.
- Може застосовуватись для моделювання нелінійних залежностей між змінними.

Недоліки:

- MLP не має можливості обробляти часові залежності, оскільки не зберігає інформацію про послідовність вхідних даних.
- Погано підходить для прогнозування часових рядів, де є тривалі залежності.

2.2.2. Рекурентні нейронні мережі (RNN)

Рекурентні нейронні мережі (RNN) використовуються для обробки послідовних даних і відрізняються від MLP здатністю зберігати інформацію про попередні стани. Завдяки рекурентному зворотному зв'язку між нейронами RNN можуть "пам'ятати" попередні значення, що є важливим для прогнозування часових рядів.

Принцип роботи:

- RNN обробляє послідовність вхідних даних покроково, на кожному кроці передаючи інформацію про попередній стан.
- На основі поточного і попереднього стану нейрони RNN генерують прогноз на кожному кроці.

Переваги:

- Можливість зберігати короточасну пам'ять, що робить їх ефективними для прогнозування на основі послідовних даних.
- Ідеально підходить для короткострокових часових рядів, де залежність від попередніх значень є ключовою.

Недоліки:

- RNN страждають від проблеми затухання або вибухання градієнта, через що втрачається інформація про далекі попередні значення.
- Обмежена ефективність при прогнозуванні довготривалих трендів або при аналізі великих обсягів даних.

2.2.3. Рекурентні мережі з довготривалою пам'яттю (LSTM)

LSTM (Long Short-Term Memory) є вдосконаленням RNN і розроблені для подолання проблеми затухання градієнта, дозволяючи зберігати інформацію про далекі минулі стани. Основною особливістю LSTM є використання блоків пам'яті з вхідним, забуваючим та вихідним шлюзами, що керують потоком інформації.

Принцип роботи:

- **Вхідний шлюз** визначає, яка інформація є важливою для запам'ятовування.
- **Забуваючий шлюз** контролює, яку частину попередньої інформації слід "забути".
- **Вихідний шлюз** дозволяє передавати інформацію до наступного стану, зберігаючи важливі залежності.

Переваги:

- Висока здатність до моделювання довготривалих залежностей, що робить LSTM ідеальними для прогнозування довгих часових рядів.
- Стійкість до затухання градієнта, що дозволяє обробляти як короткострокові, так і довгострокові тренди.

Недоліки:

- Висока обчислювальна складність, що може вимагати значних ресурсів для тренування.
- Чутливість до вибору гіперпараметрів, таких як кількість шарів та нейронів, що вимагає оптимізації для кожної задачі.

2.2.4. Згорткові нейронні мережі (CNN) для обробки часових рядів

Хоча CNN зазвичай використовуються для аналізу зображень, вони можуть бути застосовані і до фінансових даних, особливо для виявлення локальних патернів у часових рядах. CNN може виявляти короткострокові тренди та коливання в цінових даних.

Принцип роботи:

- CNN використовують згорткові шари, які виділяють локальні особливості даних (наприклад, короткострокові патерни у часових рядах).
- Пулинг (pooling) зменшує розмірність даних, виділяючи найбільш важливі особливості.

Переваги:

- Виявлення локальних патернів, що є корисним для високочастотного трейдингу.
- Можливість поєднання CNN з іншими типами мереж, такими як LSTM, для одночасної обробки локальних і глобальних трендів.

Недоліки:

- CNN обмежені у можливості запам'ятовування послідовностей і не підходять для моделювання довготривалих залежностей.
- Вимагають налаштування і оптимізації параметрів згорткових шарів.

2.2.5. Комбінація LSTM та CNN для фінансових прогнозів

Для досягнення високої точності прогнозування цін акцій часто використовуються комбіновані архітектури, такі як LSTM-CNN. Це дозволяє поєднати сильні сторони обох типів мереж:

- **CNN-LSTM:** CNN виділяє локальні особливості у часовому ряді, а LSTM враховує послідовні залежності та довготривалі тренди.
- **LSTM-CNN:** LSTM моделює довготривалі тренди, тоді як CNN використовуються для виділення основних характеристик у фінансових даних.

Комбіновані архітектури дозволяють моделювати як короткострокові, так і довгострокові залежності, що забезпечує точніші прогнози, особливо на динамічних ринках із різкими коливаннями.

2.3. Аналіз настроїв ринку за допомогою ШІ

Аналіз настроїв, відомий також як аналіз настроїв, став невід'ємною частиною фінансових прогнозів, особливо в контексті короткострокових

інвестицій. Використання моделей машинного навчання, таких як FinBERT, дозволяє визначати позитивні, негативні або нейтральні настрої, виражені в новинах, коментарях, соціальних мережах та інших текстових джерелах.

Цей підхід допомагає інвесторам оцінювати реакцію ринку на поточні події, що, у свою чергу, впливає на рішення щодо купівлі або продажу акцій.

Основний етап аналізу настроїв у фінансовій сфері розпочинається із збирання та обробки текстових даних. Фінансові новини, коментарі на форумах, соціальні мережі, такі як X (Twitter) або Reddit, стають ключовими джерелами, оскільки вони надають найбільш оперативну інформацію. Для якісного аналізу важливо провести фільтрацію отриманого контенту, щоб виключити нерелевантні повідомлення та дублі. Наприклад, видалення неінформативних символів, реклами та спам-повідомлень забезпечує чистоту даних і підвищує точність подальшої обробки.

Для перетворення текстових даних у числовий формат, придатний для обробки штучним інтелектом, застосовується токенізація та векторизація. Ці етапи допомагають подати текст у вигляді числових векторів, що враховують значення слів і їхній контекст. У фінансовому аналізі контекст має вирішальне значення, оскільки певні слова можуть набувати різного значення залежно від ситуації. Наприклад, слово "зростання" може мати позитивний або негативний відтінок, залежно від того, чи йдеться про зростання прибутків або витрат компанії. Модель FinBERT, створена на основі BERT, враховує контекст кожного слова, що дозволяє значно точніше визначати настрої у фінансових текстах.

FinBERT відмінно підходить для фінансового прогнозування, оскільки вона враховує як контекст, так і емоційне забарвлення тексту. З її допомогою можна автоматично ідентифікувати тональність новин про певну компанію або ринок у цілому, що забезпечує більш обґрунтований прогноз. Наприклад, якщо потік новин здебільшого позитивний, це може свідчити про можливий ріст ціни акцій.

Негативні новини, навпаки, часто вказують на ймовірність зниження вартості акцій.

Інтеграція даних про настрої у фінансові моделі має кілька аспектів. Окрім використання даних настроїв як додаткових змінних у традиційних моделях прогнозування, можна також створювати індикатори, що відображають середній настрій ринку або його динаміку. Це дозволяє інвесторам оцінювати загальні настрої і швидше реагувати на зміни. Дані про настрої також можуть використовуватися для фільтрації ризикових активів, якщо новини про компанію вказують на можливі труднощі, такі як негативні відгуки або юридичні проблеми. У такому разі аналіз настроїв стає додатковим інструментом для управління ризиками.

Попри переваги, аналіз настроїв має деякі виклики. Наприклад, полісемія, сарказм або іронія у фінансових новинах можуть призвести до помилок в інтерпретації. Деякі терміни або вирази можуть мати різне значення залежно від контексту, що підвищує вимоги до обробки природної мови. Крім того, частина новин може бути нерелевантною для короткострокового прогнозування, і тому модель повинна мати механізми відбору найбільш важливих повідомлень. Ще один важливий аспект – це регулярне оновлення моделі, оскільки лексика і контексти можуть змінюватися з часом, що впливає на точність аналізу.

2.4. Управління інвестиційним портфелем за допомогою ШІ

Сучасні методи штучного інтелекту (ШІ) активно використовуються для автоматизації управління інвестиційними портфелями, що дозволяє підвищувати ефективність прийняття рішень, зменшувати витрати на аналіз і знижувати ризики інвестування. Завдяки можливості обробляти значні обсяги даних, ШІ дає змогу адаптувати інвестиційні стратегії в режимі реального часу, враховуючи швидкі зміни на фінансових ринках і широке коло факторів, таких як ринкові настрої, макроекономічні показники та історичні тренди.

Одним з ключових застосувань ШІ в управлінні портфелем є рекомендаційні системи. Використовуючи алгоритми кластеризації, такі системи допомагають формувати портфелі на основі подібності компаній або фінансових інструментів. Це дозволяє виявляти акції з подібними характеристиками або високим потенціалом доходу, створюючи оптимальні портфелі на основі уподобань і цілей інвестора. Наприклад, якщо інвестор надає перевагу компаніям з високими темпами зростання, алгоритм рекомендацій зосередиться на виборі акцій, що мають потенціал до швидкого зростання, навіть якщо вони належать до різних галузей. Такий підхід дозволяє не тільки підвищити дохідність портфеля, але й забезпечити диверсифікацію ризиків.

Оптимізація портфелів за допомогою ШІ також надає інвесторам додаткові можливості для покращення структури інвестиційного портфеля з урахуванням очікуваного доходу та ризику. Такі методи, як алгоритм рою частинок (Particle Swarm Optimization, PSO) або генетичні алгоритми, використовуються для пошуку оптимальної структури портфеля, що враховує низку параметрів, зокрема ліквідність, волатильність, кореляцію між активами та поточну ситуацію на ринку. Вони здатні автоматично перебирати комбінації активів і знаходити найкращий варіант розподілу коштів, що максимізує дохід та мінімізує ризик. Наприклад, PSO може визначити найкращий портфель, що враховує як очікуваний дохід, так і толерантність до ризику інвестора, і водночас адаптуватися до змін на ринку. Завдяки цьому оптимізація стає не лише точнішою, але й швидшою, що особливо корисно на волатильних ринках, де затримка з ухваленням рішення може призвести до втрат.

Робо-адвайзери, засновані на ШІ, змінюють підхід до інвестиційного консультування, дозволяючи індивідуальним інвесторам отримувати персоналізовані поради без необхідності звертатися до фінансових консультантів. Використовуючи алгоритми ШІ, такі робо-адвайзери можуть пропонувати стратегії, що враховують індивідуальні цілі інвестора, його толерантність до ризику, поточну ситуацію на ринку та навіть аналіз настроїв.

Наприклад, якщо інвестор має середній рівень толерантності до ризику, робо-едвайзер може запропонувати портфель з більш безпечними активами під час періодів ринкової нестабільності. Окрім того, робо-едвайзери часто мають функції автоматичного ребалансування портфеля, що дозволяє підтримувати задану структуру навіть після значних змін у цінах активів.

Впровадження штучного інтелекту в управління інвестиційним портфелем має значний потенціал, але також пов'язане з певними викликами. Одним із головних ризиків є залежність від якості та обсягу даних, на основі яких здійснюється аналіз. Дані повинні бути максимально точними та актуальними, щоб уникнути помилкових прогнозів. Крім того, автоматизовані системи управління потребують регулярного оновлення алгоритмів і моделей, оскільки зміни в ринкових умовах і нестабільність можуть значно впливати на точність їхньої роботи.

Загалом, методи ШІ дозволяють значно підвищити точність управління портфелем, роблячи інвестиції більш доступними, персоналізованими та ефективними. Використання рекомендаційних систем, оптимізації портфеля та робо-адвайзерів відкриває інвесторам можливість адаптуватися до ринкових змін і отримувати дохід навіть за умов високої волатильності.

2.5. Виявлення шахрайства та аномалій у фінансових даних

Штучний інтелект відіграє значну роль у виявленні аномалій та можливого шахрайства на фінансових ринках, допомагаючи ідентифікувати підозрілі патерни та запобігати ризикам. Завдяки здатності аналізувати великі обсяги даних і виявляти відхилення від нормальної поведінки, ШІ дає змогу своєчасно реагувати на загрози, що можуть вплинути на ринкову стабільність і фінансові втрати інвесторів.

Для виявлення аномалій на ринку використовуються різноманітні алгоритми кластеризації та моделі, що спеціалізуються на аналізі відхилень, такі

як автоенкодерів. Алгоритми кластеризації дозволяють групувати дані за певними характеристиками та створювати "профілі" звичайної поведінки. Відтак, якщо певна транзакція або серія операцій відрізняється від встановлених шаблонів, система може позначити це як аномалію. Наприклад, у разі, коли виявляється різке збільшення кількості операцій з певними акціями або валютами, це може вказувати на маніпуляцію або інші незаконні дії.

Автоенкодерів, спеціальні нейронні мережі для виявлення аномалій, також активно застосовуються для ідентифікації незвичайних патернів у даних. Ці моделі навчаються на нормальних даних і здатні відтворювати звичні поведінкові шаблони з високою точністю. Коли ж система стикається з транзакціями або операціями, що не відповідають таким шаблонам, автоенкодер виявляє різницю і позначає її як потенційно аномальну. Так, наприклад, якщо на ринку відбувається серія операцій із великими обсягами або частотою, що не відповідають стандартній поведінці, автоенкодер може автоматично визначити це як ризиковий сигнал і позначити для подальшого аналізу.

Застосування моделей для виявлення аномалій має велике значення для запобігання шахрайству. На ринках з високою волатильністю, де кожна затримка у виявленні шахрайства може спричинити великі збитки, ШІ дозволяє реагувати на підозрілі дії у режимі реального часу. Це особливо важливо у випадках високочастотного трейдингу, де системи можуть відстежувати і виявляти аномалії за лічені секунди. Окрім захисту інвесторів від шахрайства, такі інструменти допомагають зберегти ринкову стабільність, оскільки завдяки їм можна запобігати паніці та великим ринковим коливанням, які можуть виникнути у разі неконтрольованих шахрайських дій.

Отже, штучний інтелект стає ефективним захисним інструментом для фінансових ринків, дозволяючи інвесторам уникати ризиків та оперативного реагувати на аномальні події, що можуть свідчити про можливі загрози. Завдяки використанню алгоритмів кластеризації та моделей на основі автоенкодерів, ШІ

здатен забезпечувати точний і надійний моніторинг ринкових процесів, допомагаючи зберігати безпеку і стабільність у фінансовій сфері.

2.6. Використання штучного інтелекту для поліпшення точності прогнозів

Штучний інтелект (ШІ) стає незамінним інструментом для інвесторів і фінансових аналітиків завдяки здатності обробляти величезні обсяги даних, виявляти приховані закономірності та робити точні прогнози. Однією з найважливіших переваг ШІ є його здатність аналізувати та інтерпретувати дані з різних джерел, таких як історичні ціни, фінансові показники компаній, новини, соціальні мережі та макроекономічні індикатори. Завдяки такій всебічній обробці інформації, моделі ШІ здатні пропонувати інвесторам більш точні прогнози та виявляти нові можливості для інвестування, які могли б залишитися непоміченими при традиційному підході.

Нейронні мережі є одними з найбільш потужних інструментів, що дозволяють ШІ виявляти складні, нелінійні залежності у фінансових даних. Наприклад, рекурентні нейронні мережі (RNN) і мережі з довготривалою пам'яттю (LSTM) здатні зберігати інформацію про попередні стани та будувати прогнози на основі послідовності даних, що є критичним для часових рядів у фінансах. Такі моделі здатні визначати як короткострокові, так і довгострокові тренди на ринку, що дозволяє інвесторам своєчасно реагувати на зміни. Використання LSTM, зокрема, є надзвичайно ефективним для фінансових ринків із високою волатильністю, оскільки ці моделі можуть враховувати складні тренди, які впливають на коливання цін протягом тривалого періоду.

Аналіз настроїв, або сентимент-аналіз, також відіграє важливу роль у покращенні прогнозів. Моделі, такі як FinBERT, дозволяють ШІ враховувати емоційні та психологічні фактори, що впливають на ринок. Наприклад, зростання позитивних новин про компанію може свідчити про потенційний ріст акцій, тоді як негативний новинний фон часто вказує на зниження інтересу

інвесторів. Інтеграція аналізу настроїв у прогностичні моделі дає змогу інвесторам отримувати більш обґрунтовані рекомендації та краще розуміти, як новинні події можуть вплинути на динаміку цін. Оскільки фінансовий ринок є надзвичайно чутливим до новин, інструменти, що враховують настрої, підвищують адаптивність прогнозів і дозволяють трейдерам ухвалювати більш обґрунтовані рішення.

Штучний інтелект також здатний ефективно обробляти великі обсяги даних, виявляючи приховані патерни, які можуть залишатися поза увагою аналітиків, що працюють із традиційними інструментами. Завдяки цьому ШІ дає змогу інвесторам оцінювати складні комбінації показників та відшукувати взаємозв'язки, що можуть сигналізувати про появу нових трендів. Наприклад, за допомогою кластеризації та інших методів машинного навчання ШІ може виявити групи акцій, що демонструють схожу поведінку на ринку, або виділити взаємозалежні індикатори, що допомагають точніше передбачати рух цін. Такі можливості сприяють не лише підвищенню точності прогнозів, але й дозволяють розширити інвестиційні можливості, які раніше не розглядалися.

Завдяки можливості обробляти великі обсяги даних у режимі реального часу та враховувати емоційні й психологічні фактори, ШІ допомагає інвесторам швидко реагувати на змінні ринкові умови, створюючи конкурентну перевагу на сучасному ринку. Однак точність прогнозів також залежить від якості даних, на яких базується модель, тому забезпечення актуальності та релевантності інформації залишається важливою умовою успіху у використанні ШІ для прогнозування фінансових показників.

2.7. Аналіз настроїв з традиційними фінансовими індикаторами

Дослідження впливу новинного настрою на фондові доходи значно просунулися вперед, охоплюючи різні аналітичні методи для підвищення точності прогнозів. Сучасна література з цієї теми містить широке різноманіття результатів, що підкреслює постійно розвивається характер застосування аналізу

сентименту на фінансових ринках. Основні теми, що виникли під час огляду літератури, включають наступні ключові ідеї:

Аналіз реального часу сентименту став важливим інструментом у прогнозуванні дохідності акцій з високою частотою. Використовуючи метод прогнозування зі змішаною частотно-ковзаючою декомпозицією, дослідження показали цінність своєчасного залучення даних сентименту у фінансові моделі. Такий підхід дозволяє вловлювати миттєвий та короткостроковий вплив настроїв інвесторів на ціни акцій, особливо в умовах, що характеризуються високочастотними торгівлями. Можливість інтегрувати аналіз реального часу у прогнозні моделі підвищує точність і чутливість прогнозів фондових ринків, забезпечуючи інвесторам більш динамічний інструмент для навігації у коливаннях ринку [1].

Дисперсія новинного сентименту, або варіативність настроїв серед різних джерел новин, була визначена як важливий чинник, що впливає на події на ринку, такі як злиття та поглинання (M&A). Дослідження показали, що відмінності у сентиментах різних ЗМІ можуть вплинути на сприйняття та результати транзакцій M&A. Згідно з висновками, більша дисперсія сентименту підвищує невизначеність на ринку, що, своєю чергою, впливає на процес прийняття рішень як інвесторами, так і компаніями. Цей висновок підкреслює нюансованість і складність інтерпретації новинного сентименту, де важливий не лише загальний настрій, але і послідовність чи мінливість цього настрою, що може суттєво впливати на поведінку ринку та впевненість інвесторів [2].

Інтеграція аналізу новинного сентименту у побудову інвестиційних портфелів значно покращила інвестиційні стратегії. Використовуючи аналіз сентименту, інвестори можуть приймати більш обґрунтовані рішення, що підвищує продуктивність портфеля. Дослідження в цій сфері вивчили, як поєднання даних сентименту з традиційними фінансовими індикаторами

дозволяє створювати більш інтелектуальні методології побудови портфеля. Такий підхід забезпечує більш цілісний аналіз ринкових умов, дозволяючи інвесторам прогнозувати ринкові тенденції та відповідно коригувати свої стратегії. У результаті, портфелі, що враховують аналіз настрою, часто краще готові скористатися ринковими можливостями і пом'якшити потенційні ризики [3, 4].

Застосування методів машинного навчання, таких як гіперналаштовані нейронні мережі з використанням оптимізації рою частинок (PSO), ще більше вдосконалили точність прогнозування цін акцій. Поєднання аналізу настрою з передовими моделями машинного навчання дозволило дослідникам оптимізувати продуктивність прогнозів, що призводить до більш надійних та стійких фондових прогнозів [5]. Крім того, моделі на основі глибокого навчання, які враховують як місцевий, так і глобальний настрій подій, були розроблені для ефективного прогнозування рухів на фондових біржах. Ці моделі демонструють потенціал інтеграції штучного інтелекту з аналізом настрою, надаючи потужний інструмент для фінансового прогнозування, який може адаптуватися до складного та динамічного характеру глобальних ринків [6].

Останнім часом застосування технік глибокого навчання і аналізу настрою для прогнозування цін акцій привернуло значну увагу. Дослідження показали, що моделі глибокого навчання, особливо нейронні мережі, відмінно прогнозують ціни акцій завдяки їхній здатності вловлювати нелінійні закономірності у фінансових даних. Порівняльні дослідження цих моделей підкреслюють їхню ефективність у вирішенні складності поведінки фінансових ринків.

Гібридні моделі, що інтегрують глибоке навчання з аналізом інвесторського настрою, також були вивчені. Ці моделі враховують як ринкові дані, так і настрої інвесторів, показуючи покращення точності прогнозів

завдяки врахуванню психологічних факторів, які визначають ринкові тренди. Цей підхід виявився цінним у покращенні продуктивності традиційних методів прогнозування [7].

Еволюція підходів, заснованих на даних, для прогнозування фінансових ринків стала темою численних досліджень, що підкреслюють зростаюче значення великих обсягів даних та передових обчислювальних методів. Ці моделі використовують великі набори даних і складні алгоритми для підвищення точності прогнозів, пропонуючи нові перспективи у визначенні руху цін акцій [8].

Інформативність технічних індикаторів також була предметом численних досліджень, що показують, що поєднання технічного аналізу з передовими прогнозними моделями може призвести до більш точних прогнозів. Включення кількох джерел даних, таких як технічні індикатори, значно покращує прогностичну силу цих моделей [9, 10].

Аналіз настрою відіграє важливу роль у покращенні моделей прогнозування цін акцій. Дослідження показали, що аналіз новин та ринкового настрою може надати цінну інформацію про реакції ринку, впливаючи на дохідність акцій. Методи глибокого навчання були застосовані до аналізу настрою, використовуючи масштабні дані для уточнення прогнозів, заснованих на настрої. Крім того, динамічний зв'язок між новинним настроєм і волатильністю дохідності акцій був досліджений за допомогою моделей перемикання режимів, ілюструючи, як настрій може впливати на коливання ринку [11, 12].

Подальші досягнення в аналізі настрою з використанням методів глибокого навчання показали важливість відображення ринкової поведінки з неструктурованих джерел даних, таких як новини та соціальні мережі. Ці дослідження в цілому підкреслюють критичну роль інтеграції глибокого

навчання, технічного аналізу та аналізу настрою для покращення точності та надійності моделей прогнозування цін акцій [13, 14, 15, 16].

РОЗДІЛ 3

ВЛАСНА СИСТЕМА ДЛЯ ВИЗНАЧЕННЯ ВПЛИВУ НОВИН НА ФІНАНСОВІ ІНСТРУМЕНТИ

3.1. Використанні технології

Сервіс прогнозування цін акцій вимагає стабільного, швидкодіючого та ефективного середовища для зберігання і обробки великих обсягів даних, які постійно оновлюються. Для цього було обрано комбінацію бази даних MongoDB та фреймворку .NET. Така інтеграція забезпечує високу продуктивність, масштабованість та гнучкість у роботі з фінансовою інформацією, де швидкість доступу та адаптивність до зміни структури даних є надзвичайно важливими.

3.1.1. MongoDB

MongoDB є сучасною NoSQL базою даних, яка оперує документами у форматі JSON-подібних об'єктів, що дозволяє легко інтегрувати та зберігати різноманітні типи даних. Це робить MongoDB ідеальною для даного проєкту, оскільки фінансові дані, такі як ціни акцій або новинні тексти з налаштуваннями, часто мають різну структуру і потребують гнучкого підходу до зберігання.

Основними перевагами MongoDB для цього сервісу є її масштабованість, гнучкість у схемах даних та висока продуктивність для запитів у реальному часі. Завдяки можливості горизонтального масштабування MongoDB може обробляти значні обсяги даних без зниження продуктивності, що є вирішальним для сервісу, де щоденно накопичується нова інформація про акції та настрої на ринку. Система може автоматично розподіляти дані між кількома серверами, що підвищує її надійність і дозволяє масштабувати зберігання та обробку даних, не знижуючи швидкість запитів.

MongoDB також пропонує надзвичайну гнучкість у визначенні схем, що дозволяє зберігати документи різних форматів без необхідності чіткої структури.

Це стає особливо корисним у ситуаціях, коли дані можуть відрізнятися залежно від типу активів чи інформаційних джерел. Наприклад, для фінансових інструментів може зберігатися інформація про дату, час, ціну до і після новинного впливу, тоді як дані про настрої новин включають оцінку настрою, джерело новин та інші метадані. Така гнучкість дозволяє зберігати структуровані та неструктуровані дані в одному середовищі, що полегшує доступ до різномірної інформації та забезпечує більшу ефективність у проведенні аналітики.

Важливою перевагою MongoDB є її висока продуктивність для запитів у реальному часі, що дозволяє сервісу оперативно обробляти інформацію і формувати прогнози на основі актуальних даних. Завдяки оптимізованим запитам, які використовують індекси, MongoDB забезпечує швидкий доступ до великих обсягів даних, що є критичним у ситуаціях, коли прогноз потрібно створити негайно, враховуючи останні зміни на ринку. Швидкість обробки запитів допомагає сервісу генерувати прогнози у режимі реального часу, надаючи користувачам можливість ухвалювати рішення на основі останніх даних.

MongoDB також виконує роль буфера для тимчасового зберігання оброблених даних перед їхньою інтеграцією у фінальні аналітичні моделі. Завдяки цьому система накопичує нову інформацію, яка згодом використовується для прогнозування цін акцій. Буферизація даних забезпечує узгодженість і дає можливість оперативно відображати результати, навіть коли нова інформація ще обробляється та інтегрується. Це значно підвищує продуктивність системи, оскільки MongoDB зберігає дані безпосередньо в тому вигляді, в якому вони будуть використовуватися, що мінімізує затримки у прогнозуванні та покращує узгодженість інформації в сервісі.

Отже, використання MongoDB для зберігання і обробки даних надає сервісу гнучкість, надійність і високу швидкість роботи, що робить цю базу даних

ідеальним вибором для задач прогнозування цін акцій на основі комплексного аналізу ринкових даних і новинного фону.

3.1.2. .NET

Фреймворк .NET було обрано для розробки бізнес-логіки сервісу завдяки його високій продуктивності, надійності та гнучкості. Однією з ключових переваг .NET є його кросплатформенність, яка дозволяє розробникам створювати універсальні додатки, що можуть працювати на різних операційних системах, таких як Windows, macOS та Linux. Це забезпечує широку доступність сервісу і зручність у його подальшій інтеграції в різні корпоративні середовища.

Асинхронна обробка запитів у .NET дозволяє сервісу ефективно взаємодіяти з базою даних MongoDB без затримок, навіть за умови великих обсягів даних і високої частоти запитів. Завдяки асинхронному програмуванню, підтримуваному у .NET, можна обробляти численні запити одночасно, не блокуючи основний потік. Це критично важливо для задач у реальному часі, оскільки інвестори потребують оперативної інформації про ринкові зміни. Наприклад, при збиранні нових даних про ціни акцій або настрої новин, асинхронна обробка запитів забезпечує стабільність і швидкість роботи системи, що дозволяє негайно оновлювати прогнози.

.NET також забезпечує просту та ефективну інтеграцію з MongoDB завдяки бібліотеці MongoDB.Driver, яка надає всі необхідні інструменти для взаємодії з базою даних. Це дозволяє бізнес-логіці сервісу зчитувати, оновлювати та зберігати документи, що містять інформацію про ціни акцій, новинний фон і настрої. За допомогою MongoDB.Driver розробники можуть легко створювати запити до бази даних, обробляти результати та оновлювати записи в реальному часі, що значно спрощує підтримку та оновлення коду. Ця бібліотека підтримує всі ключові функції MongoDB, включаючи операції CRUD (створення, читання,

оновлення, видалення) та індексацію, що забезпечує ефективний доступ до даних навіть при значних навантаженнях.

Для структурування бізнес-логіки всі етапи обробки даних у сервісі були розроблені у вигляді окремих модулів у .NET, що підвищує модульність та зручність управління кодом. Наприклад, модулі для збору інформації з API, аналізу новинного настрою та передавання даних до MongoDB були розроблені як незалежні компоненти, які можуть працювати автономно або інтегруватися у спільний робочий процес. Це дозволяє розробникам швидко оновлювати окремі частини системи без необхідності вносити зміни в інші компоненти. Така модульна структура коду також полегшує масштабування сервісу, оскільки кожен модуль може бути незалежно оптимізований для обробки зростаючих обсягів даних.

Фреймворк .NET надає підтримку для розгортання додатків у хмарних середовищах, що є важливою перевагою для сервісу з великою кількістю користувачів. Сервіс можна легко інтегрувати в різні хмарні платформи, такі як Microsoft Azure, Amazon Web Services (AWS) або Google Cloud Platform (GCP). Це дозволяє масштабувати проєкт відповідно до зростання обсягу даних або кількості користувачів. Наприклад, під час пікових навантажень система може автоматично збільшувати ресурси для забезпечення стабільної роботи, а в періоди низького навантаження — оптимізувати витрати, зменшуючи використання обчислювальних ресурсів. Завдяки хмарним рішенням, .NET дає змогу реалізувати гнучку та адаптивну інфраструктуру, яка забезпечує стабільність і продуктивність сервісу навіть при значних коливаннях трафіку.

3.2. Опис системи

1. Система отримує дані з трьох API: Stock price, News та Sentiment analysis
2. Data Controller агрегує ці дані та передає їх до Data Service

3. Data Service обробляє отримані дані та відправляє їх у Data Repository або повертає помилку у Data Controller

4. Репозиторій даних керує даними перед тим, як вони остаточно зберігаються у базі даних MongoDB, або може повернути помилку Сервісу даних.

Цей опис візуалізовано на діаграмі нижче: (Рис. 3.1).

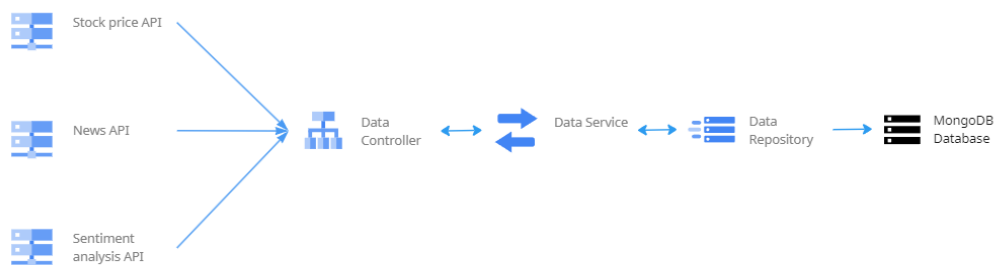


Рис. 3.1 Високорівнева діаграма архітектури системи

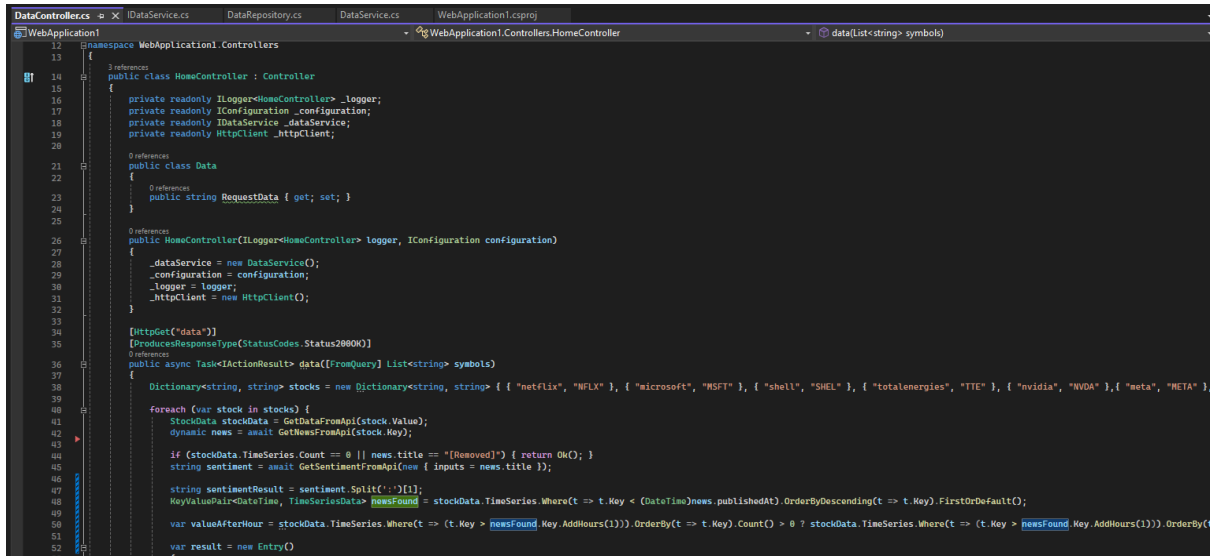
3.3. Компоненти системи:

3.3.1. API:

- **Stock Price API:** Надає інформацію про реальні ціни акцій у режимі реального часу для обраних акцій.
- **News API:** Постачає новинні статті з Google News, налаштовані для вибору найпопулярнішої статті наприкінці дня (UTC +0) для обраних акцій.
- **Sentiment Analysis API:** Аналізує заголовки новин для визначення настрою, надаючи оцінку (наприклад, позитивний, негативний, нейтральний).

3.3.2. Контролер даних (Data Controller):

Цей компонент збирає дані з трьох різних API. Він діє як агрегатор, що забезпечує уніфікацію даних із різних джерел та правильне форматування для подальшої обробки.



```
12 namespace WebApplication1.Controllers
13 {
14     public class HomeController : Controller
15     {
16         private readonly ILogger<HomeController> _logger;
17         private readonly IConfiguration _configuration;
18         private readonly IDataService _dataService;
19         private readonly HttpClient _httpClient;
20
21         0 references
22         public class Data
23         {
24             0 references
25             public string RequestData { get; set; }
26         }
27
28         0 references
29         public HomeController(ILogger<HomeController> logger, IConfiguration configuration)
30         {
31             _dataService = new DataService();
32             _configuration = configuration;
33             _logger = logger;
34             _httpClient = new HttpClient();
35         }
36
37         [HttpGet("data")]
38         [ProducesResponseType(StatusCodes.Status200OK)]
39         public async Task<ActionResult> data([FromQuery] List<string> symbols)
40         {
41             Dictionary<string, string> stocks = new Dictionary<string, string> { { "netflix", "NFLX" }, { "microsoft", "MSFT" }, { "shell", "SHEL" }, { "totalenergies", "TTE" }, { "nvidia", "NVDA" }, { "meta", "META" }
42
43             foreach (var stock in stocks) {
44                 StockData stockData = GetDataFromApi(stock.Value);
45                 dynamic news = await GetNewsFromApi(stock.Key);
46
47                 if (stockData.TimeSeries.Count == 0 || news.title == "[Removed]") { return Ok(); }
48                 string sentiment = await GetSentimentFromApi(news { inputs = news.title });
49
50                 string sentimentResult = sentiment.Split(" ");
51                 KeyValuePair<DateTime, TimeSeriesData> newsFound = stockData.TimeSeries.Where(t => t.Key < (DateTime)news.publishedAt).OrderByDescending(t => t.Key).FirstOrDefault();
52                 var valueAfterHour = stockData.TimeSeries.Where(t => (t.Key > newsFound.Key.AddHours(1))).OrderBy(t => t.Key).Count() > 0 ? stockData.TimeSeries.Where(t => (t.Key > newsFound.Key.AddHours(1))).OrderBy(
```

Рис. 3.2. Розробка компоненту Data Controller

3.3.3. Сервіс даних (Data Service):

Сервіс даних відповідає за бізнес-логіку. Він обробляє зібрані дані, що може включати фільтрацію, трансформацію та збагачення даних перед їх збереженням.

Сервіс може взаємодіяти з Контролером даних для отримання інформації та з Репозиторієм даних для збереження та вилучення оброблених даних (рис. 3.3).

```

1 using Microsoft.Extensions.Configuration;
2 using System.Diagnostics;
3 using WebApplication1.Controllers;
4 using WebApplication1.Models;
5 using WebApplication1.Repositories;
6
7 namespace WebApplication1.Services
8 {
9     public class DataService : IDataService
10    {
11        private readonly IDataRepository _dataRepository;
12
13        public DataService()
14        {
15            _dataRepository = new DataRepository();
16        }
17
18        public void SaveData(Entry entry)
19        {
20            _dataRepository.SaveEntry(entry);
21        }
22
23        public string GetData(string id)
24        {
25            return "data";
26        }
27    }
28 }

```

Рис. 3.3. Розробка сервісу даних

3.3.4. Репозиторій даних (Data Repository):

Це проміжний шар зберігання, де оброблені дані зберігаються тимчасово. Він виступає буфером між Сервісом даних та базою даних MongoDB.

Репозиторій даних забезпечує узгодженість даних та надає ефективний доступ до даних як для зберігання, так і для вилучення.

```

4
5 namespace WebApplication1.Repositories
6 {
7     public class DataRepository : IDataRepository
8     {
9         const string connectionString = "mongodb+srv://[redacted]?retryWrites=true&majority";
10
11         IMongoClient _client;
12
13         IMongoCollection<Entry> entries;
14
15         public void SaveEntry(Entry entry) {
16             try
17             {
18                 _client = new MongoClient(connectionString);
19             }
20             catch (Exception e)
21             {
22                 Console.WriteLine("There was a problem connecting to your " +
23                 "Atlas cluster. Check that the URI includes a valid " +
24                 "username and password, and that your IP address is " +
25                 "$*in the Access List. Message: {e.Message}");
26                 return;
27             }
28
29             try
30             {
31                 var collection = _client.GetDatabase("testDB").GetCollection<Entry>("entries");
32                 collection.InsertOne(entry);
33             }
34             catch (Exception e)
35             {
36                 Console.WriteLine($"Something went wrong trying to insert the new documents." +
37                 "$* Message: {e.Message}");
38                 return;
39             }
40         }
41     }
42 }
43
44

```

Рис. 3.4. Розробка репозиторію даних

3.3.5. База даних MongoDB:

Фінальне місце зберігання даних: MongoDB – це NoSQL база даних, відома своєю масштабованістю та гнучкістю в обробці різних типів даних.

Вона зберігає оброблені та структуровані дані для довгострокового зберігання та подальшого вилучення (рис. 3.5).

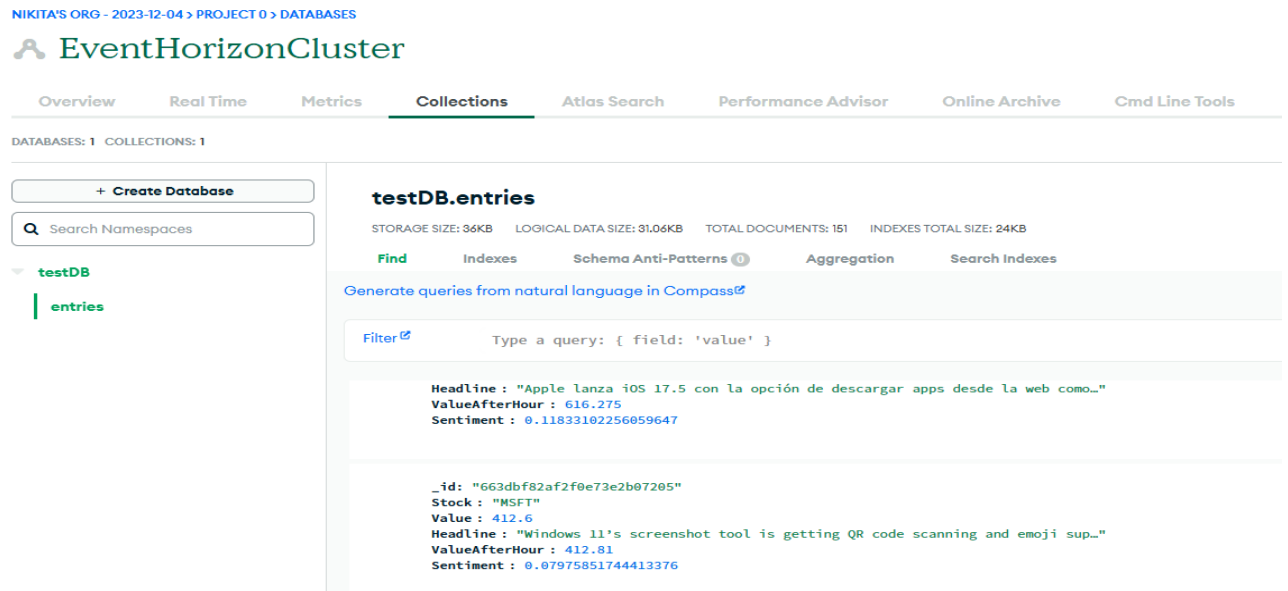


Рис. 3.5. NoSQL база даних MongoDB

3.4. Результати

У ході експерименту було зібрано дані про ціни акцій та настрої новин для вибірки з 11 великих компаній, включаючи *Meta*, *Nvidia*, *Microsoft*, *Amazon* та інших. Дані збиралися з використанням API для цін на акції, новин, а також аналізу настроїв за допомогою *FinBERT*. Загалом було зібрано 141 спостереження, що включають значення акцій перед і після виходу новин, а також відповідний показник настрою новин (позитивний, негативний або нейтральний).

3.4.1. Вплив настроїв на зміну цін акцій

Дані, отримані з API, представлені у Таблиці 3.1.

Таблиця 3.1

Дані, отримані з API

Stock	Value	Headline	Value After Hour	Sentiment	Date:
MSFT	422.92	Xbox Cloud Gaming finally supports keyboard and mouse inputs on web browsers	422.965	0.28985899686813354	\$2024-05-15T16:52:15Z
SHEL	73.3	Tamagotchi collectors rejoice: Bandai is finally rereleasing a beloved model from 2004	73.41	- 0.0826990008354187	\$2024-05-15T16:46:45Z
AMD	159.199	10 Things You Should Be Doing Daily to Protect Your Eye Health - CNET	158.942	0.41071946918964386	\$2024-05-15T14:05:41Z
INTC	31.195	Intel's New Thunderbolt Share Provides File and Screen Sharing Without Hurting Network Performance	31.21	0.0814552903175354	\$2024-05-15T16:42:00Z
NKE	91.6	I have spent entirely too much time thinking about Mark Zuckerberg's outfit at his birthday party	91.571	- 0.1332276463508606	\$2024-05-15T17:58:47Z
NFLX	638.33	Netflix releases first look at new Witcher after Henry Cavill left for Warhammer 40K	638.33	0.33275464177131653	\$2024-05-22T20:14:13Z
NVDA	1007.0	Nvidia just keeps hitting it out of the park	1007.0	- 0.1962529420852661	\$2024-05-22T22:33:16Z
GOOG	177.42	Snap brings its AR lenses to Chrome through an extension	177.59	0.1168382465839386	\$2024-05-22T19:28:05Z
TTE	70.5	Patrick Pouyanni affirme que TotalEnergies restera en France et rassure Macron	70.5	0.30006177723407745	\$2024-05-23T19:32:13Z

Після попереднього очищення та підготовки даних ми можемо отримати наступну Таблицю 3.2, де

$$pt = \text{ValueAfterHour}_t = \text{ValueAfterHour}_t$$

$$pt-1 = \text{Value}_{t-1} = \text{Value}_{t-1}$$

$$st-1 = \text{Sentiment of investors}_{t-1} = \text{Sentiment of investors}_{t-1}$$

(Настрій інвесторів).

Ціни фінансових інструментів до і після виходу новин представлені у таблиці 3.2

Таблиця 3.2

Ціни фінансових інструментів до і після виходу новин

Акція	Ціна	Ціна через годину	Сентимент	Дата	Час
MSFT	422,92	422,97	0,289858997	15.05.2024	16:52:15
SHEL	73,3	73,41	-0,082699001	15.05.2024	16:46:45
TTE	73,5	73,48	0,25498122	15.05.2024	18:00:02
AMD	159,199	158,94	0,410719469	15.05.2024	14:05:41
INTC	31,195	31,21	0,08145529	15.05.2024	16:42:00
NKE	91,6	91,57	-0,133227646	15.05.2024	17:58:47
NKE	92,44	92,44	0,353558853	22.05.2024	22:05:10
NFLX	638,33	638,33	0,332754642	22.05.2024	20:14:13
NVDA	1007	1007	-0,196252942	22.05.2024	22:33:16
GOOG	177,42	177,59	0,116838247	22.05.2024	19:28:05
NKE	92,44	92,44	0,353558853	22.05.2024	22:05:10
NKE	91,39	91,47	0,416316167	23.05.2024	18:22:14
NFLX	634,45	634,45	0,028437734	23.05.2024	22:00:33
SHEL	70,33	70,19	0,407320112	23.05.2024	13:58:25
TTE	70,5	70,5	0,300061777	23.05.2024	19:32:13
NVDA	1056,4	1037,9	0,397178553	23.05.2024	13:38:21

У першій моделі AR(1), як авторегресійної моделі першого порядку, що може бути описана як ARIMA(1,0,0), для прогнозування ціни акцій використовуються два предиктори: ціна акцій перед новинами (Value) та настрої інвесторів (Sentiment) у формі множинної регресії:

$$p_t = a_0 + a_1 \cdot p_{t-1} + a_2 \cdot s_{t-1} + u_t \quad (1)$$

де

p_t – залежна змінна або значення ціни через годину після впливу настрою,

p_{t-1} – пояснювальна змінна або значення ціни до впливу настрою;

s_{t-1} – пояснювальна змінна або значення впливу настрою (позитивний, нейтральний чи негативний);

\hat{u}_t – залишковий член.

Друга модель побудована наступним чином:

$$dpt = b_0 + b_1 \cdot st-1 + wt \quad (2)$$

де

dpt – залежна змінна, яка дорівнює 1, якщо $pt > pt-1$, $p_t > p_{t-1}$, -1, якщо $pt < pt-1$, $p_t < p_{t-1}$, і дорівнює 0, якщо $pt = pt-1$, $p_t = p_{t-1}$;

$st-1$ – пояснювальна змінна або значення впливу настрою (позитивний, нейтральний чи негативний);

wt – залишковий член.

Після оцінки обох моделей методом найменших квадратів і застосування t-критерію для перевірки статистичної значущості параметрів ми отримуємо наступні результати (Таблиця 3.3).

Таблиця 3.3.

Параметри моделей прогнозування ціни

Акція	a_1	a_2	R^2	b_1	R^2
TTE	0.9986*	0.2199	99.58%	0.8870	5.69%
SHEL	1.0432*	0.1254	99.21%	0.0407	0.01%
NVDA	0.9986*	-5.4765	99.99%	1.1046	9.00%
NKE	0.9990*	0.0726	99.97%	0.8422	3.92%
NFLX	0.9979*	-3.8008*	99.95%	-1.8080	20.60%
MSFT	1.0046*	0.2195	99.98%	2.6682	26.06%
META	0.9956*	1.4641*	99.99%	3.4737*	70.00%
INTC	0.9104*	-0.2202	90.04%	-0.5167	3.04%
GOOG	1.0097*	-0.5636	99.88%	-1.5804	11.04%
AMZN	0.8740*	-1.4521	99.84%	3.0305	23.83%
AMD	1.0369*	-0.1624	99.95%	1.0345	11.52%

* означає статистичну значущість параметрів

Результати експерименту продемонстрували значний вплив настроїв на зміни цін акцій для деяких компаній. Зокрема:

- **Meta:** Інвестиційні настрої мали визначальний вплив на ціну акцій Meta. Коефіцієнт детермінації склав 70%, що означає, що 70% змін ціни акцій можна пояснити зміною настроїв ринку. Це підтверджує високий рівень впливу новинних настроїв на компанію, що активно обговорюється в медіа.

- **Nvidia:** Негативні новинні настрої мали суттєвий вплив на ціну акцій Nvidia. Для цієї компанії було зафіксовано зниження цін після публікацій новин з негативним настроєм, що свідчить про вразливість акцій до новинного фону.

3.4.2. Порівняння результатів для моделей ARIMA та ARIMA + Sentiment

Порівняння результатів авторегресійної моделі (ARIMA) з моделлю, що включала фактор настроїв, показало, що інтеграція настроїв новин підвищує точність прогнозування. Основні висновки включають:

- **Точність прогнозів:** Модель ARIMA, доповнена настроями новин, забезпечила більш точні прогнози для акцій Meta та Nvidia у порівнянні з моделлю, яка враховувала лише історичні ціни. Це вказує на важливість врахування якісних факторів при прогнозуванні.

- **Значення настроїв як фактора:** Для інших компаній (наприклад, Microsoft, AMD, Amazon) вплив настроїв був меншим, але все ж позитивно вплинув на точність прогнозування. Це підкреслює, що настрої можуть мати різний рівень значущості для різних компаній залежно від їхньої медійної активності та сприйняття ринку.

3.4.3. Статистична значущість результатів

Для оцінки статистичної значущості результатів було застосовано t-критерій для параметрів моделей. Було виявлено, що для Meta та Nvidia параметри, що відповідають за настрої, мали високу статистичну значущість ($p < 0,05$), тоді як для інших компаній цей показник варіювався залежно від ринкових умов.

3.4.4. Переваги та обмеження запропонованої моделі

Результати експерименту показали, що запропонована модель має суттєві переваги, особливо у підвищенні точності прогнозування. Інтеграція аналізу настроїв новин, зокрема використання спеціалізованих моделей для обробки тексту, дозволила значно підвищити точність прогнозів для окремих компаній. Завдяки тому, що модель враховує емоційне забарвлення новинного фону, прогнози краще відображають ринкову ситуацію та стають більш точними і чутливими до короткострокових змін. Врахування настроїв стало важливим індикатором, який дозволяє краще передбачати реакцію ринку на певні новини або події, що надає інвесторам можливість вчасно реагувати на зміни.

Крім того, автоматизоване збирання та аналіз новинного фону забезпечують високу швидкість обробки даних. Система в режимі реального часу збирає новини, аналізує їх і оперативно адаптує прогнози відповідно до ринкових умов. Це забезпечує значну перевагу для інвесторів, які мають доступ до актуальних прогнозів і можуть ухвалювати рішення на основі свіжих даних. Швидкість обробки є критично важливою на сучасному ринку, де затримка в ухваленні рішень може призвести до втрати можливостей.

Однак, попри значні переваги, модель має і певні обмеження, які впливають на надійність прогнозів. Одним із таких обмежень є залежність від якості вхідних даних. Надійність аналізу настроїв залежить від точності та повноти текстових даних, які отримуються з новинних джерел. Якщо дані є неточними або неповними, це може вплинути на результати прогнозів, оскільки модель може отримувати спотворену картину настроїв ринку. Крім того, ефективність моделі може знижуватися за умови недостатньої кількості спостережень, що особливо важливо для компаній з меншою

кількістю новинних згадок. У таких випадках прогнози можуть бути менш надійними, оскільки модель має обмежену кількість інформації для аналізу настроїв і побудови точного прогнозу.

Таким чином, хоча запропонована модель значно підвищує точність і швидкість прогнозування для окремих компаній, її ефективність залежить від якості та кількості вхідних даних. Удосконалення цих аспектів у майбутньому може додатково підвищити надійність і точність системи, забезпечуючи інвесторам ще більшу гнучкість та обґрунтованість у прийнятті рішень.

ВИСНОВКИ

Проведений експеримент продемонстрував переваги використання сучасних моделей для прогнозування ринкових змін, зокрема через врахування інвесторських настроїв, визначених за допомогою аналізу новин. Було розроблено новий сервіс, який інтегрує аналіз настроїв у реальному часі з прогнозуванням фондового ринку, що дозволяє подолати обмеження існуючих інструментів. Сервіс автоматизує збір і обробку даних через API, надаючи трейдерам та інвесторам своєчасну та релевантну інформацію, що економить час і дозволяє зосередитися на стратегічному плануванні, позбавляючи від рутинної ручної роботи.

Результати експерименту показали, що поєднання історичних даних із кількісним аналізом новинних настроїв підвищує точність прогнозування, забезпечуючи більш обґрунтовані та своєчасні рішення для інвесторів. Розроблена модель перевершує існуючі сервіси завдяки здатності виявляти причинно-наслідкові, а не лише кореляційні зв'язки, та перетворювати інвесторські настрої з новин у кількісні значення. Це дозволяє точніше передбачати зміни цін на акції, враховуючи як історичні дані, так і настрої ринку.

Під час дослідження було виявлено перспективні напрями для подальшої роботи, зокрема вдосконалення моделей з урахуванням специфіки різних ринкових сегментів і типів активів. Додатковим кроком може стати аналіз впливу конкретних індикаторів настроїв, таких як соціальні мережі або новинні джерела, на поведінку інвесторів. Це дозволить створити моделі, здатні швидко реагувати на динамічні зміни ринку та знижувати ризики.

Загалом, інтеграція аналізу настроїв у прогностичні моделі дозволяє створити новий рівень аналітики, який підвищує точність і своєчасність рішень. Врахування як кількісних даних, так і настроїв ринку сприяє

розробці більш обґрунтованих стратегій інвестування, що має значний потенціал у сучасному фінансовому середовищі. З огляду на подальше вдосконалення цих технологій очікується, що вони відіграватимуть дедалі важливішу роль у фінансовому секторі, допомагаючи інвесторам адаптуватися до ринкових умов та виявляти нові можливості для прибутку.

REFERENCES

1. MongoDB - Wikipedia – [Электронный ресурс]. – Режим доступа: <https://en.wikipedia.org/wiki/MongoDB>
2. .NET Framework - Wikidpedia – [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/.NET_Framework
3. Cai, Y., Tang, Z., Chen, Y. “Can real-time investor sentiment help predict the high-frequency stock returns? Evidence from a mixed-frequency-rolling decomposition forecasting method”. *The North American Journal of Economics and Finance* 2024; 72: 102147. DOI: 10.1016/j.najef.2024.102147.
4. Chen, Y., Lu, J., Ma, W., Kumar, S., Shahab, Y. “Dispersion in News Sentiment and M&As Outcomes”. *Research in International Business and Finance* 2024; 71: 102415. DOI: 10.1016/j.ribaf.2024.102415.
5. Hung, M. C., Hsia, P. H., Kuang, X. J., Lin, S. K. (). “Intelligent portfolio construction via news sentiment analysis”. *International Review of Economics & Finance* 2024; 89: 605–617. DOI: 10.1016/j.iref.2023.07.103.
6. Shahi, T. B., Shrestha, A., Neupane, A., Guo, W. “Stock Price Forecasting with Deep Learning: A Comparative Study”. *Mathematics* 2020; 8(9): 1441. DOI: 10.3390/math8091441.
7. Chauhan, A., SJ, S., H, S., Quadir, A., Venkataraman, N. “Stock price forecasting using PSO hypertuned neural nets and ensembling”. *Applied Soft Computing* 2023; 147: 110835. DOI: 10.1016/j.asoc.2023.110835.
8. Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., Muhammad, K. “A local and global event sentiment based efficient stock exchange forecasting using deep learning”. *International Journal of Information Management* 2020; 50: 432–451. DOI: 10.1016/j.ijinfomgt.2019.07.011.
9. Jing, N., Wu, Z., Wang, H. “A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction”. *Expert Systems with Applications* 2021; 178: 115019. DOI: 10.1016/j.eswa.2021.115019.

10. Dai, T., Hu, J., Ding, X. “Progress and prospects of data-driven forecasting models for financial markets”. *International Journal of Cognitive Computing in Engineering* 2023; 19(2): 145-168. DOI: 10.1016/j.ijcin.2023.102415.
11. Shi, Y., Ho, K-Y., Liu, W-M. “Public Information Arrival and Stock Return Volatility: Evidence from News Sentiment and a Markov Regime-Switching Approach”. *International Review of Economics & Finance* 2016; 42: 291–312. DOI: 10.1016/j.iref.2015.10.011.
12. Zhou, Y., Tang, J., & Wang, Q. “Investigating the informativeness of technical indicators: A comprehensive study”. *Knowledge-Based Systems* 2022; 236: 107492. DOI: 10.1016/j.knosys.2022.107492.
13. Li, X., Xie, H., Chen, L., Wang, J., Deng, X. “News Impact on Stock Price Return via Sentiment Analysis”. *Knowledge-Based Systems* 2014; 69: 14–23. DOI: 10.1016/j.knosys.2014.07.011.
14. Souma, W., Vodenska, I., Aoyama, H. “Enhanced News Sentiment Analysis Using Deep Learning Methods”. *Journal of Computational Social Science* 2019; 2(1): 33–46. DOI: 10.1007/s42001-019-00040-6.
15. Shahi, T. B., Shrestha, A., Neupane, A., Guo, W. “Stock Price Forecasting with Deep Learning: A Comparative Study”. *Mathematics* 2020; 8(9): 1441. DOI: 10.3390/math8091441.
16. Sohangir, S., Wang, D., Pomeranets, A., Khoshgoftaar, T. M. “Big Data: Deep Learning for Financial Sentiment Analysis”. *Journal of Big Data* 2018; 5(1): 1–25. DOI: 10.1186/s40537-017-0111-6.
17. Ivanov, O., Kobets, V. “Data Analysis for Predicting Stock Prices Using Financial Indicators Based on Business Reports. *Proceedings of the Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2023*. CCIS, 2023, vol.1980, pp. 227-239. Springer, Cham, DOI 10.1007/978-3-031-48325-7_17.

18. Kobets, V., Yatsenko, V., Mazur, A., Zubrii, M. “Data analysis of personalized investment decision making using robo-advisers”. *Science and Innovation* 2020; 16(2): 80–93. DOI: 10.15407/scine16.02.080.