

Володимир Дорош¹, Роман Ваврик², Олена Станкевич³

¹Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери 12, Львів, Україна, E-mail: volodymyr.y.dorosh@lpnu.ua, ORCID 0009-0001-1782-2110²

²Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, Львів, Україна, E-mail: roman.r.vavryk@lpnu.ua@lpnu.ua, ORCID 0009-0009-0490-7519

³Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, Львів, Україна, E-mail: olena.m.stankevych@lpnu.ua, ORCID 0000-0002-5977-6351

РОЗРОБЛЕННЯ СЕНТИМЕНТ АНАЛІЗАТОРА З ВИКОРИСТАННЯМ CHATGPT ДЛЯ ФОНДОВОГО РИНКУ

Отримано: березень 12, 2024 / Переглянуто: березень 28, 2024 / Прийнято: квітень 01, 2024

© Дорош В., Ваврик Р., Станкевич О., 2024

<https://doi.org/>

Анотація. На сьогодні важлива проблема фінансового успіху полягає у пошуку ефективних підходів до торгівлі, які б могли адаптуватися до швидкозмінних умов ринку та забезпечувати високу дохідність інвестицій. На основі проведеного аналізу літературних джерел ChatGPT визначено як перспективну технологію, яка ефективніша за FinBert у використанні як компоненти для проведення сентиментального аналізу акцій. Також дослідження показують задовільну ефективність та продуктивність роботи ChatGPT. У відомих джерелах немає детального розкриття теми автоматизації процесу сентимент аналізу та тестування моделі ChatGPT на великих даних. Метою виконаних досліджень є розроблення автоматизованої системи для аналізу сентименту на основі ChatGPT з інтегрованим агрегатором новин для збору та аналізу фінансових даних. У дослідженні деталізовано створення комплексного ескізу системи. Наведено план, який охоплює весь спектр запропонованої системи. Розроблено попередню архітектуру додатку, що забезпечує візуальне та структурне представлення того, як різні компоненти системи взаємодіють та функціонують узгоджено. Цей архітектурний план слугує дорожньою картою для впровадження та розгортання автоматичного сентимент аналізатора, забезпечуючи чіткість і точність його розробки. Побудовано первинні діаграми зв'язків між сутностями у системі та запропоновано алгоритм її роботи. Подальші дослідження передбачають створення мінімальної робочої системи сентимент аналізатора та оцінювання її ефективності та якості роботи.

Ключові слова: штучний інтелект, ChatGPT, фінансовий ринок, сентимент аналіз, агрегатор новин

Вступ

Сучасний розвиток машинного навчання (МН, англ. Machine Learning, ML) та штучного інтелекту (ШІ, англ. Artificial Intelligence, AI) відкриває нові можливості для оптимізації торгових стратегій, підвищення ефективності прийняття рішень та зниження ризиків на фінансових ринках. Інноваційним інструментом для їх реалізації є система ChatGPT, яка ґрунтується на передових алгоритмах обробки природної мови (англ. Natural Language Processing, NLP) та МН. Значну кількість досліджень щодо можливостей та напрямків використання цієї технології опубліковано компанією OpenAI.

З огляду на швидкі зміни на ринку та велику кількість необхідної для аналізу інформації, важливо визначити, як штучний інтелект на основі аналізу даних, новин та трендів може допомогти у прийнятті обґрунтованих інвестиційних рішень. Розглядаючи ChatGPT як інструмент для аналізу та виконання торгових стратегій, важливо вивчити його здатність адаптуватися до динаміки ринку, здійснювати прогнозування та оптимізувати торгові алгоритми з метою максимізації прибутку та мінімізації ризиків.

Постановка проблеми

У сучасному світі інвестування та торгівлі на фінансовому ринку існує два основних підходи до аналізу акцій: фундаментальний та технічний. Фундаментальний аналіз передбачає вивчення економічних, фінансових та інших квалітативних та кількісних чинників, які можуть вплинути на вартість цінних паперів. Цей підхід передбачає глибоке розуміння того, як різноманітні економічні показники, фінансові звіти компаній, новини про галузі та макроекономічні тенденції можуть впливати на інвестиційну привабливість та майбутню прибутковість акцій. Натомість, технічний аналіз займається вивченням цінових графіків та торгових об'ємів з метою визначення майбутніх тенденцій ринку на основі історичних даних.

Водночас, фундаментальний аналіз стикається з низкою проблем та викликів, серед яких до головних відносять обсяг і складність необхідних для аналізу даних. Інвесторам та аналітикам потрібно враховувати велику кількість інформації, що вимагає значних зусиль та часу для збору, опрацювання та інтерпретації даних. Крім того, суб'єктивна інтерпретація фінансових звітів та економічних показників може призвести до різних висновків, що ускладнює прийняття обґрунтованих інвестиційних рішень.

Традиційно фінансові аналітики покладаються на фундаментальний і технічний аналіз для прогнозування руху ринку. Однак ці методи часто не враховують психологічні та емоційні чинники, які можуть суттєво впливати на вартість активів. Сентиментальний аналіз заповнює цю прогалину, надаючи уявлення про емоційний пульс ринку, пропонуючи більш цілісний погляд на потенційні напрямки руху ринку. Поєднання можливостей сентиментального аналізу та ChatGPT може допомогти у вирішенні низки зазначених вище проблем та водночас сприяти підвищенню ефективності різних напрямків фінансової діяльності.

Огляд сучасних джерел інформації за тематикою публікації

На сьогодні сентиментальний аналіз, який часто називають “аналізом думок”, став ключовим інструментом у фінансовому секторі, що дає змогу аналітикам, біржовим торговцям та інвестиційним фірмам оцінювати ринкові настрої та приймати обґрунтовані рішення. Основна передумова аналізу настроїв у фінансовій сфері полягає в тому, що колективний настрій або настрої учасників ринку – отримані зі статей новин, соціальних мереж, фінансових звітів та інших текстових даних – можуть суттєво впливати на ціни на акції, обсяги торгів і ринкові тенденції.

Поява технологій ШІ і МН зробила революцію в сфері сентиментального аналізу, особливо у великій і волатильній сфері фінансових ринків. Інструменти штучного інтелекту, засновані на складних алгоритмах і можливостях NLP, можуть аналізувати величезні обсяги даних із безпрецедентною швидкістю, виявляючи закономірності, настрої і прогнозні сигнали, невидимі для людського ока.

Спочатку аналіз настроїв у фінансовій сфері здійснювали вручну. Трудомісткий та за своєю суттю упереджений процес покладався на спрощені методи класифікації слів як позитивних, негативних або нейтральних без контексту. ШІ та МН дали змогу подолати ці обмеження через використання моделей глибокого навчання, які розуміють контекст, нюанси і навіть сарказм, забезпечуючи точніший і дієвіший аналіз настроїв.

Передові моделі ШІ, такі як ChatGPT, значно розширили можливості аналізу настроїв у фінансовому контексті. ChatGPT може розуміти та інтерпретувати тонкощі людської мови, що дає йому можливість аналізувати заголовки новин, фінансові звіти та повідомлення в соціальних мережах з надзвичайною точністю. Це дає змогу аналізувати настрої в реальному часі, що має вирішальне значення на швидкоплинних фінансових ринках, де умови можуть змінюватися за лічені секунди.

Інструменти ШІ не обмежуються лише аналізом текстів. Вони здатні інтегрувати кількісні дані з якісним аналізом настроїв, пропонуючи комплексне уявлення про ринкові настрої. Така інтеграція дає змогу біржовим торговцям і аналітикам приймати більш обґрунтовані рішення, поєднуючи традиційні фінансові показники з даними про настрої для більш точного прогнозування ринкових рухів.

Розроблення сентимент аналізатора з використанням ChatGPT для фондового ринку

Окрім інструменту ChatGPT відомі й інші моделі ШІ, зокрема, модель FinBERT. FinBERT це NLP-модель, яка натренована на фінансових комунікаційних даних: корпоративних та аналітичних звітах, стенограмах дзвінків фінансистів. Оптимізована для аналізу фінансового сектору модель FinBERT ефективно впоралася з класифікацією текстових даних на основі їх сентименту. Однак її застосування обмежене статичністю навчального набору даних, що не дає можливості адекватно реагувати на нові, раніше невідомі ринкові тенденції чи специфічний жаргон у контексті новин. Водночас модель ChatGPT через використання широкого спектру даних для навчання, що дає змогу краще розуміти контекст і нюанси фінансових новин, відрізняється гнучкістю та адаптивністю. Ще одна її перевага полягає у здатності до постійного навчання, що може забезпечити більшу точність аналізу сентименту, особливо у випадках, коли ринкові умови швидко змінюються.

У праці [1] подано результати порівняння продуктивності та ефективності двох моделей GPT-3.5 і FinBERT. Вибрали шість наборів ключових інструкцій для сентимент аналізу фінансових настроїв на валютному ринку за допомогою ChatGPT та FinBERT. Для визначення ефективності моделей класифікації настроїв розробили систему оцінювання на багатьох показниках: правильність (accuracy), точність (precision), запам'ятовування (recall), оцінка F1 (F1-Score). Ці показники характеризують ступінь передбачення моделями класів настрою, визначених в анотованому наборі даних. Ключовим показником у вибраному наборі інструментів оцінювання слугує середнє значення настрою – абсолютна похибка (S-MAE). За результатами досліджень встановлено, що моделі GPT, зокрема GPT-P2 і GPT-P4, мають вищі значення параметрів оцінювання та найнижчий показник S-MAE, порівняно з моделлю FinBERT. Це свідчить про те, що моделі GPT є ефективнішими в аналізі фінансових настроїв, що потенційно може призвести до створення кращих моделей прогнозування для торгівлі на ринку валютного ринку. Дослідження підкреслює важливість вибору правильної моделі для конкретних валютних пар для оптимізації торгових стратегій на основі аналізу настроїв.

У праці [2] подано результати порівняння різних моделей GPT-1, GPT-2, GPT-3.5 та найновішої моделі GPT-4 для використання ChatGPT у передбаченні вартості цінних паперів. Різні мовні моделі досліджували на основі заголовків новин, порівнюючи їхню ефективність із традиційними методами аналізу настроїв. Виявили значну позитивну кореляцію між оцінками, згенерованими ChatGPT, і подальшою прибутковістю акцій, що підкреслює перевагу першої над більш ранніми моделями, такими як GPT-1, GPT-2 і BERT, і традиційними оцінками настроїв. Показано, що складні моделі, зокрема, ChatGPT-4, забезпечують вищі коефіцієнти Шарпа. Отже, вдосконалені мовні моделі можуть покращити інвестиційні стратегії шляхом точної інтерпретації новинних настроїв. Запропоновано новий метод оцінки аргументації моделей, який завдяки включенню новітніх мовних моделей в інвестиційні рішення може дати точніші прогнози і поліпшити кількісні показники торгових стратегій. Такий підхід поєднує глибокий лінгвістичний і контекстуальний аналіз заголовків фінансових новин з кількісними даними фондового ринку. В основі методу лежить вилучення нюансів сентимент аналізу і тематичних інсайтів, які корелюються з прибутковістю фондового ринку для оцінки їхньої прогностичної цінності. Загалом запропонований інноваційний підхід дає змогу точніше прогнозувати рух фондових ринків на основі настроїв новин, що є значним досягненням у сфері кількісних фінансів.

Здатність ChatGPT прогнозувати тенденції на фондовому ринку на основі сентимент аналізу настроїв у Twitter вивчали у праці [3]. Зосередившись на Microsoft і Google, автори виявили позитивну кореляцію між оцінками настроїв ChatGPT і динамікою акцій обох компаній на наступний день, досягнувши точності 70% для Microsoft і приблизно 64% для Google. Дослідження підкреслило зростаючу важливість ШІ у формуванні прогнозів фінансового ринку.

Використання ChatGPT для аналізу настроїв як перспективний підхід представлено у різних дослідженнях. Зокрема, у праці [4] вивчали можливості моделі ChatGPT як універсального аналізатора настроїв, оцінюючи його використання для розуміння думок, настроїв та емоцій, які містяться в тексті. Провели оцінювання 7 репрезентативних завдань аналізу настрою, що охоплюють 17 еталонних наборів даних. На цих наборах порівняли ChatGPT з точно

налаштованими моделями BERT і SOTA. Встановили, що модель ChatGPT може ефективніше вирішувати проблему зміни полярності під час аналізу настроїв та показує високу продуктивність у сценаріях відкритого домену.

Практичне застосування ChatGPT для аналізу настроїв, особливо в дослідженнях користувача та UX-дизайні представлено у праці [5]. У ній підкреслюється здатність моделі ефективно обробляти великі обсяги тексту для аналізу настроїв, використовуючи функцію інформаційного маркування для подолання обмеження в 4096 символів. Це дозволяє аналізувати великі набори даних, розбиваючи їх на менші, керовані частини, які можна позначати, об'єднувати та аналізувати. У статті також надається інформація про вибір правильної мовної моделі для аналізу настроїв на основі таких критеріїв, як розмір набору даних, метрики продуктивності, обчислювальні ресурси та складність завдання.

На основі аналізу літературних джерел можна зробити висновок, що ChatGPT має потенціал використання як основи для аналізатора настроїв. Це стимулює подальші спроби модифікації алгоритмів чи розробки рішень, які б поєднувались із агрегаторами новин. Такий підхід може підвищити точність і ефективність аналізу настроїв в фінансових даних, відкриваючи нові можливості для інвесторів і біржових торговців на ринку.

Цілі та проблеми дослідження

Метою виконаних досліджень є розроблення автоматизованої системи для аналізу настроїв на основі ChatGPT з інтегрованим агрегатором новин для збору та аналізу фінансових даних. Для її реалізації необхідно визначити оптимальні способи модифікації алгоритму ChatGPT для підвищення його ефективності у сфері фінансових ринків, а також розробити механізми для ефективної інтеграції з агрегаторами новин для автоматизованого збору та аналізу відповідного контенту.

Виклад основного матеріалу

Здатність моделі ChatGPT розуміти нюанси мови та надавати детальний аналіз настроїв навіть у складних сценаріях підкреслює її потенціал як цінного ресурсу в різних сферах, включаючи маркетингові дослідження, аналіз відгуків клієнтів та моніторинг соціальних мереж. Отже, ChatGPT можна використати як основний механізм для настроїв аналізатора. Загальне рішення можна модифікувати завдяки агрегатору новин.

Агрегатори новин для аналізу фондових настроїв передбачають, насамперед, збір великої кількості оновлених даних, пов'язаних з конкретними акціями або фондовим ринком в цілому. Далі ці дані можна проаналізувати, щоб визначити загальний настрій щодо цих акцій, який може бути позитивним, негативним або нейтральним. Процес часто складається з кількох етапів:

Збір даних: агрегатори збирають новини з різних джерел, включаючи веб-сайти фінансових новин, блоги і форуми, на яких обговорюються тенденції на фондовому ринку.

Опрацювання даних: зібрані дані очищають і опрацьовують для видалення релевантної інформації, такої як заголовки, дати публікацій і текст, з акцентом на ключові слова, пов'язані з фондовими біржами.

Аналіз настроїв: передові алгоритми, часто на основі технологій ШІ та обробки природної мови, такі як GPT-моделі OpenAI, аналізують текст, щоб оцінити настрої, висловлені щодо згаданих акцій.

Агрегація та розрахунок балів: оцінки настроїв агрегуються за певний період, щоб розрахувати загальну оцінку настроїв для кожної акції, яку можна використовувати для прогнозування руху ринку.

Використання агрегаторів новин для аналізу настроїв на фондовому ринку є перспективним напрямком для покращення прогнозів та інвестиційних стратегій. Цей підхід поєднує в собі технологічні досягнення в галузі ШІ та обробки природної мови з широким висвітленням новин агрегаторами, забезпечуючи широкий набір даних для аналізу.

Результати та обговорення

Для побудови аналітики сентимент аналізатора запропоновано алгоритм, який охоплює три основні кроки: 1) щоденна оцінка настроїв: агрегування оцінки настроїв усіх статей для акції за кожен день; передбачається використання усередненої оцінки або середньозваженого значення, заснованого на довірі до джерела статті або її впливу; 2) аналіз тенденцій: обчислення ковзних середніх оцінок настроїв за різні періоди часу (7-денний, 30-денний) для виявлення тенденцій; 3) аналіз волатильності: вимірювання мінливості оцінок настроїв як показника невизначеності ринкових настроїв.

На Рис. 1 подано розроблену діаграму функціональних компонент сентимент аналізатора.

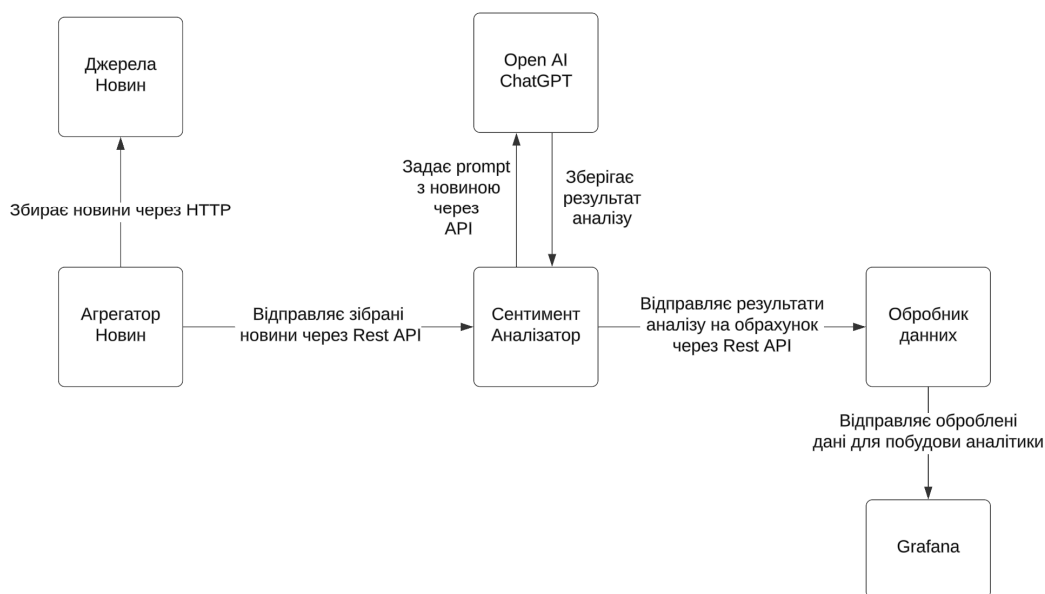


Рис. 1. Функціональні компоненти автоматизованого сентимент аналізатора.

Комунікацію між компонентами автоматизованого аналізатора забезпечуватимуть HTTP запити, та API, спроектовані за конвенцією REST. Послідовність роботи системи така: джерела новин надходять до Агрегатора Новин; Агрегатор Новин надсилає статті до API OpenAI для аналізу настроїв; результати аналізу настроїв зберігаються в базі даних разом зі статтями та інформацією про акції; обробник даних обчислює аналітику на основі інформації з бази даних; Grafana візуалізує аналітику для кінцевих користувачів.

Перейдемо до проектування компоненти сентимент аналізатора. Проаналізуємо потенційні сутності, які можна використати у нашій системі. Для цього спроектуємо діаграму зв'язків між сутностями бази даних (EntityRelationshipDatabase). ERD для цього проекту включатиме такі сутності: статті (Articles) (містить інформація про статті з новинами: назва, джерело, URL і дата публікації); акції (Stocks) (містить символи акцій, назви компаній і, можливо, класифікацію секторів чи галузей); аналіз настроїв (SentimentAnalysisResult) (зберігає результати сентимент-аналізу, посилаючись на статті та акції, може містити поля для оцінки настроїв, рівня довіри і, можливо, витягнуті ключові слова); аналітика (StocksAnalytics) (це може бути не таблиця, а подання або набір процедур, які обчислюють сукупні оцінки настроїв, тенденції та інші показники в часі для візуалізації). Діаграму запропонованих сутностей зображено на Рис. 2. Взаємозалежності у сутностях не вказано, оскільки у подальшій реалізації доцільніше використовувати нереляційну базу даних.

Для збору статей про акції слід обрати авторитетні джерела новин, серед них можуть бути наступні: Bloomberg, Reuters, CNBC, FinancialTimes, TheWallStreetJournal, Investing.com, YahooFinance (пропонує API, який може бути особливо корисним), AlphaVantage (надає API для цін на акції і в майбутньому може включати новини).

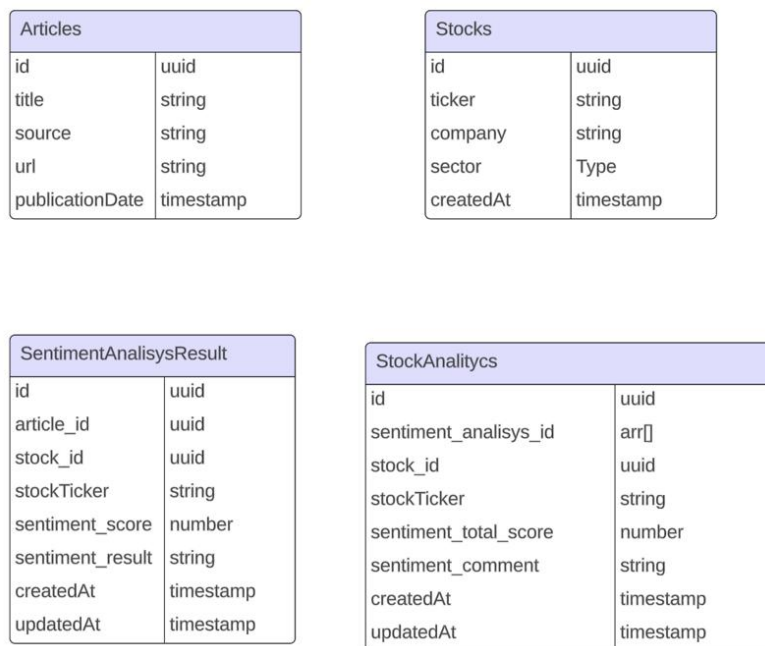


Рис. 2. Первинний ескіз діаграми сутностей sentiment аналізатора.

Перейдемо до більш детальної побудови концепції компоненти агрегатора новин, структури бази даних, ескізу базового алгоритму агрегації новин та методів опрацювання і збору даних. Для спрощення процесу агрегації новин та проведення інтеграцій пропонуємо обрати єдину платформу, на якій представлені більшість авторитетних джерел новин. Однією з таких платформ є Telegram, який має функціональність публічних каналів для поширення новин. На Telegram представлені такі видавництва, як Bloomberg, CNBC тощо. Створення алгоритму агрегування платформи Telegram є простішим у порівнянні з інтеграцією окремих ресурсів, які мають різну структуру API та різні структури сайту. Перейдемо до проектування сутностей, які оперуватимуться в агрегаторі та БД.

Для початкової вибірки сутностей для Telegram пропонується використати сервіс telegramer та tgstat. Написано дві функції, які за вибраними параметрами – теми та країни, отримують перелік сутностей та вносять інформацію про них в базу даних. Для збереження даних про канали та групи створено таблицю ‘channels’ із атрибутами для збереження даних про унікальний текстовий ідентифікатор, ім’я, опис та дату останнього опрацювання. Опрацювання каналів відбуватиметься за допомогою бібліотеки requests на Linux серверах, оскільки ця операція не вимагає від користувача авторизації. Розгортання буде виконано за допомогою Docker. Для наочності буде використано демонстрацію можливостей на основі браузера Chrome. Отримавши завдання на опрацювання, метод буде надсилати HTTP-запит за посиланням ‘https://t.me/s/{channel_slug}’, де ‘channel_slug’ – унікальний текстовий ідентифікатор каналу. Результатом виконання запиту є генерування HTML-сторінки із детальною інформацією про канал. Приклад такої сторінки зображено на Рис. 3. На сторінці можна знайти інформацію про канал, яка зберігається в елементі зі значенням XPath ‘//div[@class="tgme_channel_info"]’. Із цього тегу можна вибрати назву каналу та його детальний опис.

Наступний крок – це знаходження на сторінці останньої публікації. Батьківським елементом для усіх публікацій є тег ‘section’ зі значенням ‘tgme_channel_historyjs-message_history’ параметру ‘class’. Ідентифікатори публікації знаходяться в атрибуті ‘href’ дочірнього тегу ‘a’ зі 72 значенням ‘tgme_widget_message_video_playernot_supportedjs-message_video_player’ параметру ‘class’. Результатом пошуку буде список публікацій на каналі. Для подальшого опрацювання потрібно лише значення ідентифікатора останнього. Дані про останню публікацію та сторінку буде передано на наступний етап – опрацювання публікацій.

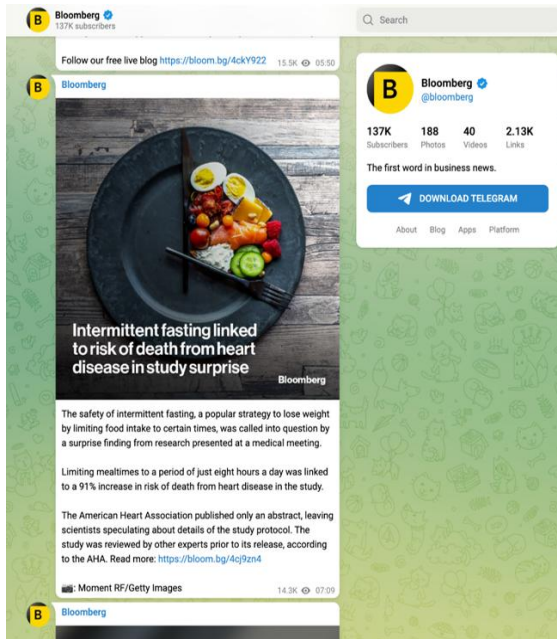


Рис. 3. Вигляд сторінки Telegram каналу.

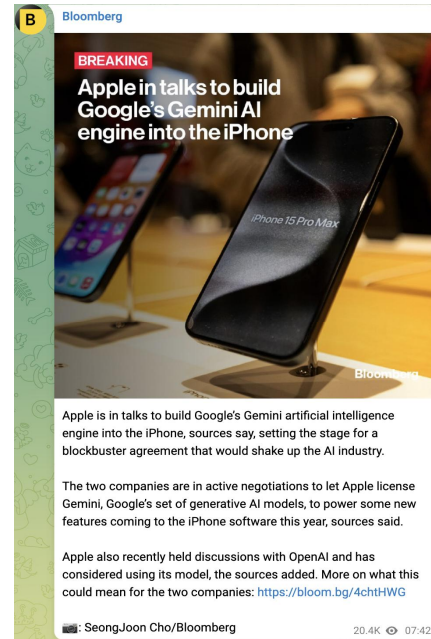


Рис. 4. Вигляд публікації в Telegram.

Для опрацювання публікацій буде виконано цикл HTTP-запитів за посиланням 'https://t.me/{channel_slug}/{post_id}?embed=1', де 'channel_slug' – ідентифікатор каналу, 'post_id' – цифровий ідентифікатор публікації. Також встановлено фільтр часу, за яким публікації, які старіші за часовий ліміт, не опрацьовуватимуться. На кожній ітерації циклу, значення 'post_id' буде зменшуватися на одиницю. Таким чином унеможливиться ситуація, коли на етапі виявлення публікацій для опрацювання на попередньому етапі буде загублено певну їх кількість. Сторінку публікації зображено на Рис. 4. Зі сторінки публікації можна отримати наступні дані, на основі яких створено таблицю 'posts' із відповідними атрибутами: ідентифікатор публікації; детальний опис; дата публікації; кількість переглядів; посилання на публікацію; посилання на медіа файл.

Також важливою інформацією (за наявності) слугують дані про першоджерело публікації. Якщо таке існує, його буде внесено в таблицю 'channels'. Таким чином відбуватиметься автоматичне розширення мережі каналів та груп для опрацювання. Дана реалізація зберігає значну кількість ресурсів, які мали б бути витрачені на постійний пошук релевантних матеріалів, як у випадку із TikTok. За наявності коментарів їх опрацювання буде виконано за допомогою Telegram API [6], для використання якого необхідно створити акаунт розробника за посиланням <https://my.telegram.org>. Акаунт розробника зображено на Рис. 5. Параметри 'Appar_id' та 'Appar_hash' використовуються для авторизації користувача бібліотекою Telethon [7], яка розроблена для простої взаємодії користувача із Telegram API. Із переліку функцій бібліотеки буде використано метод `iter_messages()`, який приймає два параметри – текстовий ідентифікатор каналу та ідентифікатор публікації, для якої необхідно отримати коментарі. Бібліотека також надає змогу опрацьовувати публікації та канали окремо, але використання її можливостей для цього не є раціональним рішенням, оскільки після певної кількості запитів, API блокує можливість надсилати нові, тому при можливості варто використовувати інші методи для отримання необхідних даних.

Для збереження коментарів створено таблицю 'comments' із атрибутами для збереження тексту повідомлення, ідентифікатора повідомлення для якого була зроблена відповідь, цифровий ідентифікатор повідомлення, текстовий ідентифікатор користувача. Загальну структуру бази даних для месенджеру Telegram зображено на Рис. 6.

Система для збору даних із Telegram працюватиме за схемою. Вузли системи працюватимуть на різних серверах, стаючи в чергу для отримання завдання за допомогою параметра 'last_searched'. Під час виникнення критичної помилки система припинить свою роботу та внесе інформацію про

помилку в таблицю 'logs'. Розширення мережі каналів відбуватиметься за допомогою того ж методу, що опрацює сутності та їхні публікації. Система працюватиме на Linux серверах із розгортанням за допомогою Docker контейнерів. Блок-схему методу для збирання даних із Telegram зображено на Рис. 7.

App configuration

App api_id:

App api_hash:

App title:

Short name:

alphanumeric, 5-32 characters

Рис. 5. Аккаунт розробника Telegram.

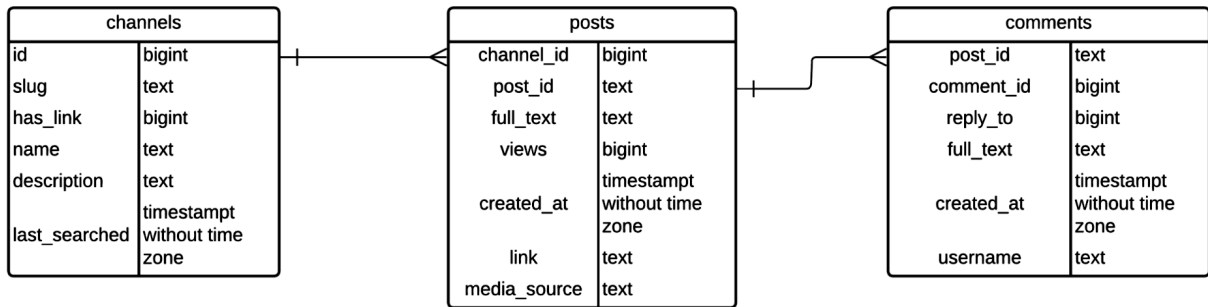


Рис. 6. Структура бази даних для збереження даних Telegram.

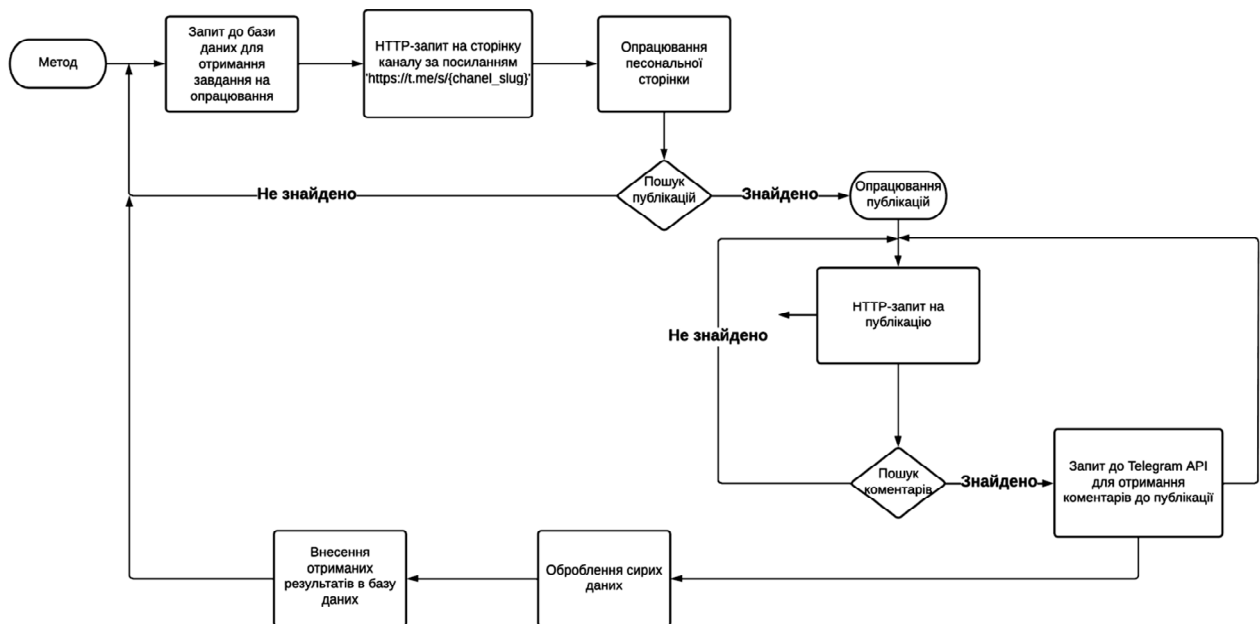


Рис. 7. Блок-схема для опрацювання сутностей в Telegram.

Розроблення сентимент аналізатора з використанням ChatGPT для фондового ринку

На основі розглянутих та спроектованих компонентів перейдемо до технологічного стеку. Попередня ідея технічного стеку така:

Серверна частина: Python (Flask або Django для управління веб-сервером, Requests або BeautifulSoup для скрапінгу, SQLAlchemy для ORM бази даних).

База даних: надаємо перевагу нереляційній базі даних MongoDB, враховуючи контекст невідомих викликів, які очікуватимуться під час реалізації, гнучкість нереляційної бази даних та структура ключ-значення відіграє переважуючу роль на відміну від реляційної бази.

Клієнтська частина: для клієнтської інформаційної панелі за межами Grafana бібліотека React у парі зі строго типізованою мовою Typescript чудово підходять для проектів невеликого масштабу та побудови мінімального робочого рішення.

Візуалізація даних: Grafana для інформаційних панелей, підключений до бази даних.

API для аналізу настроїв та комунікації між компонентами: OpenAI GPT API для аналізу тексту, Rest API для комунікації між компонентами. Перевагою Rest API є простота, поширеність та простота налагодження. Під час виникнення труднощів із швидкістю між внутрішніми компонентами, можна буде перейти на протокол gRPC.

Хостинг/інфраструктура: AWS або GoogleCloudPlatform для хостингу додатку, бази даних, а також для масштабованих обчислювальних ресурсів за потреби. GoogleCloudPlatform у пріоритеті, оскільки порівняно з іншими провайдер є дешевшим та надає хороші пільгові умови для початку роботи.

CI/CD: GitHubActions або GitLab CI для безперервної інтеграції та розгортання, що полегшує автоматичне оновлення та тестування.

Описаний технологічний стек охоплює загальну картину того, що необхідно для створення автоматизованого сентимент аналізатора акцій, орієнтованого на фондовий ринок. Кожну з технологій можна розширити до більш детальних планів і специфікацій упродовж просування проектної реалізації.

Подальші дослідження передбачають створення та тестування мінімального життєздатного продукту, який об'єднає ChatGPT з агрегатором новин для аналізу настроїв. Для оцінки його продуктивності, зокрема, його точності та надійності в реальних умовах буде використано великі набори даних новин. Цей етап тестування має вирішальне значення для вдосконалення автоматизованої системи та забезпечення її відповідності динамічним потребам аналізу фінансових ринків. Завдяки цьому процесу ми прагнемо створити надійний інструмент на основі штучного інтелекту, який покращить процес прийняття рішень на фінансових ринках.

Висновки

За результатами вивчення застосування ChatGPT у сфері торгових стратегій на фінансових ринках запропоновано вдосконалений підхід до використання штучного інтелекту для підвищення точності та ефективності аналізу ринкових настроїв. Це нове рішення інтегрує ChatGPT як основний двигун для аналізу настроїв з компонентом агрегатора новин, створюючи синергію, яка використовує сильні сторони ШІ в аналізі і розумінні нюансів фінансових новин і ринкових настроїв. Така інтеграція не тільки розширює можливості аналізу настроїв на фінансових ринках, а й виводить застосування ШІ у прийнятті фінансових рішень на нові рівні.

У дослідженні деталізовано створення комплексного ескізу автоматизованого аналізатора. Розроблено попередню архітектуру додатку, що забезпечує візуальне та структурне представлення того, як різні компоненти системи взаємодіють та функціонують узгоджено. Побудовано діаграми сутностей як для агрегатора новин, так і для компонентів аналізатора настроїв, що дає детальне уявлення про структуру бази даних, взаємозв'язки та взаємодію між різними сутностями. Наведено алгоритми для агрегування новин та проведення автоматизованого аналізу.

Розроблений архітектурний план слугує дорожньою картою для впровадження та розгортання автоматичного сентимент аналізатора, забезпечуючи чіткість і точність його розробки.

Перелік використаних джерел

- [1] G.Fatouros, J.Soldatos, K.Kouroumali, G.Makridis, D.Kyriazis, “Transforming Sentiment Analysis in the Financial Domain with ChatGPT”, Preprint submitted to Machine Learning with Applications, 2023, <https://arXiv.org/pdf/2308.07935.pdf>
- [2] A.Lopez-Lira, Y.Tang, “Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models”, 2023, Available at SSRN: <https://doi.org/10.2139/ssrn.4412788>
- [3] U.Mumtaz, S.Mumtaz, “Potential of ChatGPT in predicting stock market trends based on Twitter Sentiment Analysis”, 2023, <https://doi.org/10.48550/arXiv.2311.06273>
- [4] Z.Wang, Q.Xie, Y.Feng, Z.Ding, Z.Yang, R.Xia, “Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study”
- [5] D.Guaglianone, “Sentiment analysis using ChatGPT: Boost your research using AI”, 2024, <https://www.kiquix.com/how-to/sentiment-analysis-using-chatgpt-boost-your-research-using-ai/>
- [6] Telegram API [Електронний ресурс] – режим доступу: <https://core.telegram.org/>
- [7] Telethon’s Documentation [Електронний ресурс] – режим доступу: <https://docs.telethon.dev/en/stable/>

Volodymyr Dorosh¹, Roman Vavryk², Olena Stankevych³

¹ Computer Aided Design Department, Lviv Polytechnic National University S. Bandery st. 12, Lviv, Ukraine,
E-mail: volodymyr.y.dorosh@lpnu.ua, ORCID 0009-0001-1782-2110²

² Computer Aided Design Department, Lviv Polytechnic National University S. Bandery st. 12, Lviv, Ukraine
E-mail: roman.r.vavryk@lpnu.ua, ORCID 0009-0009-0490-7519

³ Computer Aided Design Department, Lviv Polytechnic National University S. Bandery st. 12, Lviv, Ukraine,
E-mail: olena.m.stankevych@lpnu.ua, ORCID 0000-0002-5977-6351

DEVELOPING A SENTIMENT ANALYZER USING CHATGPT FOR A STOCK MARKET

Received: March 12, 2024 / Revised: March 28, 2024 / Accepted: April 01, 2024

© Dorosh V., Vavryk R., Stankevych O., 2024

Abstract. Today, an important problem of financial success is to find effective trading approaches that can adapt to rapidly changing market conditions and ensure high investment returns. Based on the literature analysis, ChatGPT is identified as a promising technology that is more effective than FinBERT in being used as a component for conducting sentiment analysis of stocks. The research also shows satisfactory efficiency and productivity of ChatGPT. Existing sources do not provide a detailed description of the automation of the sentiment analysis process and testing of the ChatGPT model on big data. The purpose of the performed research is to develop an automated system for sentiment analysis based on ChatGPT with an integrated news aggregator for collecting and analyzing financial data. The study details the creation of a comprehensive solution sketch. A plan is presented that covers the entire range of the proposed system. A preliminary application architecture has been developed that provides a visual and structural representation of how the various components of the solution interact and function in a coordinated manner. This architectural plan serves as a roadmap for the implementation and deployment of the automated sentiment analyzer, ensuring clarity and accuracy in its design. Initial diagrams of the relationships between the entities in the system have been developed and an algorithm for the system has been proposed. Further research will focus on creating a minimum working system for the sentiment analyzer and evaluating its efficiency and quality of work.

Keywords: Artificial Intelligence, ChatGPT, finance market, sentiment analysis, news aggregator