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

Застосування контент аналізу для підвищення продуктивності бізнесу IT продукту

Кваліфікаційна робота (проєкт)

на здобуття ступеня вищої освіти «бакалавр»

Виконав: здобувач 4 курсу,12- 461 групи
Спеціальності
126 Інформаційні системи та технології
Освітньо-професійної програми
першого (бакалаврського) рівня вищої освіти
Тіток Ігор Михайлович
Керівник: доктор економічних наук, професор
Кобець Віталій Миколайович

Рецензент: senior developer, team-lead, tech-lead, IT компанія DataArt
Яцюта Владислав Олександрович

Херсон – Івано-Франківськ – 2024

ЗМІСТ

ВСТУП..... 3

РОЗДІЛ 1. ТЕХНОЛОГІЇ ОБРОБКИ ПРИРОДНОЇ МОВИ ДЛЯ СЕМАНТИЧНОГО АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ В УПРАВЛІННІ ІННОВАЦІЯМИ.……4

1.1. Роль відкритих інновацій та інноваційного посередництва в епоху технологій……..4

1.2. Моделі для семантичного аналізу тексту ……..7

РОЗДІЛ 2. ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ В БІЗНЕС-АНАЛІТИЦІ: МОДЕЛІ, ІНСТРУМЕНТИ ТА ПЕРЕВАГИ..... 9

2.1. Бізнес аналітика в прийнятті рішень 9

2.2. Система бізнес-аналітики на основі машинного навчання 11

РОЗДІЛ 3. АНАЛІЗ НАСТРОЇВ ЗА ДОПОМОГОЮ ІНСТРУМЕНТУ FinBERT..... 17

3.1. Поняття і методи аналізу настроїв у бізнесі 17

3.2. Моделі аналізу текстів BERT і FinBERT ……...20

3.3. Архітектура та методи навчання FinBERT ……...22

3.4. Аналіз настроїв на основі заголовків біржових новин ……...26

ВИСНОВКИ……...30

СПИСОК ВИКОРИСТАННИХ ДЖЕРЕЛ……..31

ВСТУП

Актуальність. Аналіз інформації в бізнес-аналітиці – це кропітка та довга робота, якщо нею займається людина без ІТ-інструментарію і автоматизованих методів обробки вхідних даних, і щоб модифікувати роботу та заощадити час, було вирішено проаналізувати предметну область і на основі проведеного дослідження застосувати інструменти, які допоможуть з мінімальним людським втручанням аналізувати фінансові настрої тексту для прийняття управлінських рішень.

Об'єкт дослідження – контент аналіз.

Предмет дослідження - аналіз настроїв у бізнес аналітиці для обґрунтування фінансових рішень.

Мета дослідження – автоматизувати процес аналізу настроїв фінансових даних для обґрунтування фінансових рішень.

Завдання дослідження

- Проаналізувати інформацію в бізнес-аналітиці для обґрунтування фінансових рішень.
- Дослідити методологію розбору концепції аналізу настроїв та його застосування у різних галузях, включаючи фінанси.
- На основі підготовлених даних у вигляді 3000 заголовків біржових новин за допомогою навченої для аналізу природної моделі FinBERT виконати аналіз їх настроїв для обґрунтування фінансових рішень

Методи і технології дослідження: мова програмування Python, попередньо навчена модель NLP для аналізу настрою фінансового тексту FinBERT, W&B Tables.

Структура роботи: вступ, чотири розділи, висновки, список використаних джерел.

РОЗДІЛ 1
ТЕХНОЛОГІЇ ОБРОБКИ ПРИРОДНОЇ МОВИ ДЛЯ СЕМАНТИЧНОГО АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ В УПРАВЛІННІ ІННОВАЦІЯМИ

1.1 Роль відкритих інновацій та інноваційного посередництва в епоху технологій

Щоб бути в тренді та адаптуватися до постійних змін в середовищі, організації постійно шукають нові рішення для відповіді на інноваційні виклики. Впровадження практики відкритих інновацій дозволяє організаціям отримувати доступ до інформації з різних джерел і зменшує витрати на дослідження можливостей та потенційних рішень інноваційних проблем. Як результат, відкриті інновації стали основною парадигмою для дослідження та були успішно прийняті численними фірмами та громадськими організаціями. У той же час виникла концепція інноваційного посередництва, яка описує процес сканування, збирання, об'єднання та обміну інформацією, пов'язаною з інноваціями, які здійснюють організаційні агенти, так звані інноваційні посередники. Ці посередники можуть виконувати важливі дії на ранніх етапах інноваційного процесу, такі як прогнозування, формулювання потреб та вимог, сканування та фільтрування інформації, або генерація та поширення знань.

Перехід до більш відкритих стратегій пошуку та обміну знаннями був спричинений появою нових технологій, таких як Інтернет, великі дані, API та хмарні обчислення. Ці інновації розширили спектр доступних джерел для інновацій. Технології дозволили створити нові типи посередників без участі людей, наприклад, патентні бази даних, онлайн-спільноти та соціальні мережі, що призвело до появи нових методів пошуку інновацій. [1]

Сучасні організації можуть легко накопичувати велику кількість контенту, пов'язаного з рішеннями, з великих офіційних баз даних патентів

або внутрішніх продуктів. Вони також можуть збирати ідеї з різних внутрішніх та зовнішніх джерел знань або сканувати Інтернет у пошуках цінних оглядів продуктів або онлайн-документів.

Люди швидко досягають своїх когнітивних меж у розумінні інформації та можуть обмежувати свою увагу, посилюючи упередження або слідуючи загальноприйнятим ідеям, тому доступ до великого обсягу інформації без засобів її ефективної обробки не створює переваг для організацій. Технології штучного інтелекту, такі як машинне навчання, глибоке навчання та обробка природної мови, дозволяють організаціям покращити обробку інформації, автоматизувати пошук та переосмислити інноваційні процеси. Ці технології допомагають аналізувати текст, відділяти важливе від неважливого та навіть створювати нові знання з мінімальною участю людини. Інтеграція цих можливостей у зручні інтерфейси дозволяє менеджерам здійснювати краще посередництво між знаннями у процесі інновацій. [2]

Протягом останніх десятиліть динаміка обміну інформацією та створення нових продуктів і послуг зазнала істотних змін. Якщо раніше цей процес переважно відбувався в межах відділів досліджень і розробок та маркетингу, то зараз знання стали більш доступними і широко поширеними. Це означає, що неможливо знаходити всю необхідну інформацію в одній організації чи в одному відділі.

Відкриті інновації відкривають перед організаціями нові можливості. Вони дозволяють отримувати доступ до знань із різних джерел і водночас знижують витрати на виявлення різноманітних бізнес-можливостей та потенційних рішень інноваційних проблем. Зокрема, стала популярною діяльність практик відкритих інновацій, які збирають інформацію, пов'язану з інноваціями, і витягують цінні знання з різних джерел.

У сучасному світі інноваційні посередники відіграють важливу роль у сприянні обміну знаннями між учасниками, що шукають інновації. Вони допомагають розширити доступ до інформації про потреби, можливих

партнерів, технології і продукти. Останні роки відзначаються появою нових типів інноваційних посередників, таких як онлайн-спільноти, патентні бази даних та соціальні мережі. Ці цифрові посередники відкрили нові можливості для пошуку інновацій та розширили простір рішень для організацій у цифрову еру.

У сучасному світі, де технології швидко розвиваються, організації стикаються з безпрецедентним потоком інформації та викликами, пов'язаними з управлінням інноваціями. Нові технології, такі як штучний інтелект та аналіз даних, привносять нові можливості, але разом з цим і нові завдання. Зміна підходу до пошуку та обробки знань стає важливою для забезпечення конкурентоспроможності.[3]

У мережі платформ та сервісів, де доступ до інформації відкритий, важливо знати, як ефективно здійснювати пошук та аналіз знань. Відкриті інновації відкривають нові можливості для організацій, дозволяючи їм отримувати доступ до широкого спектру інформації з різних джерел. Це може включати обмін знаннями з партнерами, клієнтами або навіть конкурентами через спеціалізовані платформи та мережі.

Аналогічно розвиток штучного інтелекту і аналізу даних відкриває нові перспективи в обробці та аналізі інформації. Зокрема, обробка природної мови стає ключовим інструментом у виявленні нових можливостей та трендів у масиві даних. Це дозволяє організаціям здійснювати більш точний та ефективний пошук, а також прогнозувати майбутні тенденції на основі аналізу текстової інформації.

Таким чином, використання новітніх технологій у пошуку та аналізі знань може виявитися вирішальним для успіху сучасних організацій у конкурентному середовищі.

1.2 Моделі для семантичного аналізу тексту

В останні роки наука про розробку мовних моделей для семантичного аналізу тексту розвивається швидкими темпами. З'являється множина різних підходів і методик, які переважно ґрунтуються на штучних нейронних мережах [4]. Важливим досягненням стало створення моделей, які можуть автоматично аналізувати та розуміти зміст текстової інформації. Це відкриває широкі можливості для застосування обробки природної мови у різних сферах, включаючи аналіз настроїв, класифікацію текстів, машинний переклад, тематичне моделювання та багато інших.

Наприклад, раніше популярні моделі "мішка слів" підраховували кількість входжень окремих слів у текст і перетворювали їх на числові представлення, проте ці моделі не враховували порядок слів у тексті [5]. Сучасні ж нейронні мережі, такі як Word2Vec або GloVe, дозволяють розуміти асоціації між словами, використовуючи великі набори текстових даних для навчання [6]. Це дозволяє створювати більш точні та розумні моделі, які здатні враховувати значення слів у різних контекстах.

Крім того, останні тенденції показують, що контекстні моделі, такі як BERT або GPT-3.5, стають все популярнішими завдяки своїм можливостям адаптуватися до конкретного контексту у змісті тексту. Це означає, що кожне слово може мати різне значення в залежності від контексту, в якому воно використовується. Ці нові моделі використовують глибоке навчання для створення універсальних представлень тексту, що робить їх вельми потужними інструментами для розуміння та аналізу природної мови [7].

Хоча огляди надають цінне уявлення про сучасний стан NLP в управлінні інноваціями і технологіями, пропонують методологічні рекомендації для майбутнього застосування, вони не вбудовують пошук інновацій за допомогою NLP в існуючі інноваційні теорії і концепції. Аналіз практик пошуку за допомогою NLP з погляду інноваційного посередництва є плідним способом системного розуміння діяльності та функцій NLP як

посередника без участі людини в інноваціях, що може допомогти виявити і розвитати нову галузь досліджень. [8]

РОЗДІЛ 2.

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ В БІЗНЕС-АНАЛІТИЦІ: МОДЕЛІ, ІНСТРУМЕНТИ ТА ПЕРЕВАГИ

2.1 Бізнес аналітика в прийнятті рішень

Ганс Петер Лун вперше представив концепцію системи бізнес-аналітики для компаній у 1958 р. За останні кілька десятиліть бізнес-аналітика стала широко використовуватися в різних галузях, від малого бізнесу до великих корпорацій зі списку Fortune 500. (рис. 2.1)

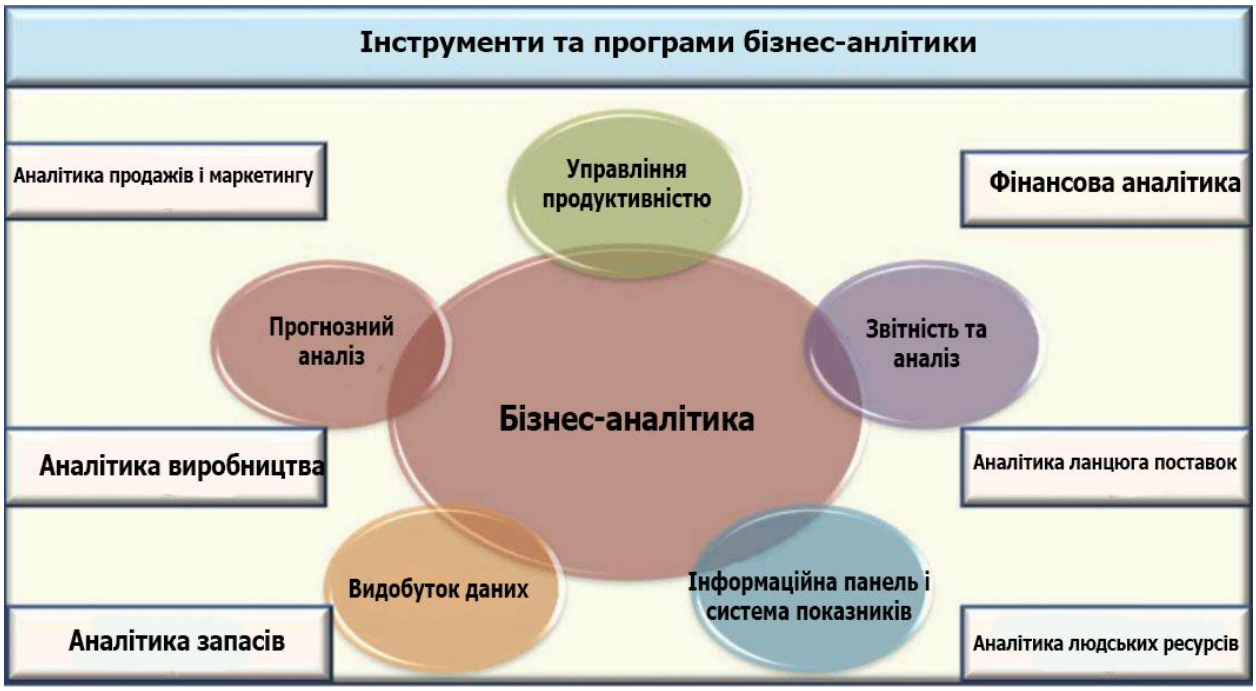


Рис. 2.1 Інструменти та додатки бізнес-аналітики

Основна мета програмного забезпечення для бізнес-аналітики (BI) - допомогти користувачам швидко знаходити та аналізувати дані, важливі для прийняття обґрунтованих бізнес-рішень в режимі реального часу. Сьогодні бізнес-аналітика включає в себе рішення на різних рівнях управління, від операційного до тактичного, що поєднують різні дані з різних джерел. Це дозволяє бізнесу приймати більш обґрунтовані рішення, оскільки бізнес-аналітика дозволяє працювати зі структурованими та неструктурованими даними з різних джерел, таких як бази даних, електронні таблиці, Інтернет тощо. [9]

Аналізуючи відомості про постачальників, клієнтів, партнерів і товари компанії, інструменти бізнес-аналітики можуть надати та візуалізувати корисну інформацію для прийняття рішень. Бізнес-функції, такі як продажі та маркетинг, фінанси, управління персоналом, виробництво та ланцюг поставок, використовують інструменти BI для звітності, аналізу, управління продуктивністю, прогнозного аналізу та прийняття рішень.

Ефективні бізнес-операції у XXI ст. потребують використання BI в операційному середовищі. Сучасний бізнес потребує ефективних засобів прийняття рішень. Такі інструменти можна отримати, поєднавши аналітичні та операційні системи, що дозволить розширити можливості аналізу до операційного рівня.[10]

Однак завдяки розвитку апаратного та програмного забезпечення, таких як мультиядерні 64-бітні процесори, віртуалізація, зберігання за рядками та стовпцями, алгоритми резидентної бази даних у основній пам'яті та аналітика у пам'яті, прогалина між операційними та аналітичними системами зменшується. Це стимулює оцінку наступного покоління BI, відомого під різними назвами: операційна бізнес-аналітика, операційна аналітика, аналітика в реальному часі, динамічна BI та гібридна BI. Такі системи, що є поточними, динамічними, однорідними, зменшують час дії, доступні для великої кількості користувачів та надають інформацію для прийняття рішень на операційному, тактичному та стратегічному рівнях, є відмінними рисами таких передових систем операційного BI. [11]

Одним із аспектів BI, що найшвидше розвиваються, є операційний BI. Операційні BI надає інформацію про відповідні показники в реальному часі, що відрізняє його від звичайних BI. Останнім часом інструменти прийняття рішень на низькому рівні набувають все більшого значення в компаніях. Операційний BI стає ідеальним інструментом для бізнесу, оскільки він допомагає приймати рішення як на низькому рівні (для щоденних операцій), так і на високому рівні (для стратегічного планування). [12]

Оцінити цінність програмного забезпечення BI можна за допомогою моделей машинного навчання (ML). Програмне забезпечення бізнес-аналітики створене для обробки та аналізу великих обсягів даних, щоб робити корисні висновки та здійснювати стратегічне планування завдяки автоматизації та оптимізації функцій інструментів BI.[13]

За допомогою ML-моделей можна оптимізувати та автоматизувати внутрішні операції BI-інструменту. Алгоритми машинного навчання можна використовувати для покращення методів агрегації та узагальнення даних, автоматизації генерації звітів та визначення найкращих візуалізацій для певного набору даних. Ці зміни роблять BI-інструмент більш ефективним, заощаджують час користувачів і підвищують його цінність для прийняття рішень.

Загалом, ML-моделі можуть значно покращити функції предиктивної аналітики, виявлення аномалій, обробки природної мови, рекомендаційних систем, оцінки якості даних та оптимізації BI-інструментів, роблячи їх більш корисними для цілей оцінки. Інтегруючи машинне навчання в BI-платформи, компанії можуть отримати нові знання, вдосконалити процеси прийняття рішень і отримати більше інсайдів зі своїх даних [14].

2.2 Система бізнес-аналітики на основі машинного навчання

Компанії впроваджують інструменти бізнес-аналітики (BI), щоб краще аналізувати і представляти великі обсяги даних. Прогнозуючи та аналізуючи вплив цих систем на ключові показники ефективності та бізнес-результати, моделі машинного навчання можна використовувати для оцінки цінності рішень бізнес-аналітики (BI). [15]

Нижче наведено приклади алгоритмів машинного навчання, які можна використовувати для вимірювання рентабельності інвестицій в інструменти BI. Прогнозування впливу BI-систем на ключові показники ефективності,

такі як дохід, задоволеність клієнтів і продуктивність співробітників, можна зробити за допомогою регресійних моделей. Використовуючи ці моделі, компанії можуть розрахувати вартість своїх інвестицій у програмне забезпечення (ПЗ) для бізнес-аналітики (BI) і визначити, де вони можуть

досягти найбільшого прогресу. (рис.2.2) [16]

Рис.2.2 Типи моделей бізнес-аналітики на основі машинного навчання

Регресійні моделі

- Прогнозування продажів: Прогнози продажів можна робити за допомогою регресійних моделей, які враховують минулі дані, а також зовнішні змінні, такі як рекламні акції, економічні умови та сезонність. Компанії можуть використовувати ці моделі для покращення управління запасами, планування виробництва та реклами.
- Довічна цінність клієнта (LTV): Цінність клієнта впродовж життя можна оцінити за допомогою регресійних моделей, які враховують демографічні дані клієнта, історію покупок та інші характеристики. Використовуючи ці дані, компанії можуть краще спрямовувати свої маркетингові зусилля на залучення та утримання найприбутковіших клієнтів.
- Продуктивність працівників: На продуктивність працівників може впливати низка змінних, зокрема освіта, досвід і задоволеність роботою. Використовуючи ці дані, компанії можуть точно налаштувати свої стратегії управління людськими ресурсами, щоб підвищити моральний дух і продуктивність.
- Задоволеність клієнтів: Якість продукції, ціна та сервіс - це лише деякі зі змінних, які можна врахувати в регресійній моделі, щоб визначити їхній вплив на загальну задоволеність клієнтів. Компанії можуть використовувати ці дані для покращення своїх пропозицій та завоювання лояльності клієнтів.

За допомогою алгоритмів класифікації можна виявити та класифікувати патерни даних. Їх можна використовувати для прогнозування ймовірності того, що споживач зробить покупку або працівник звільниться. Розпізнаючи ці тенденції, бізнес може покращити свої судження та підвищити ефективність, а також виявити тенденції в даних і відсортувати їх за категоріями, якими можна керувати. На основі наявних даних ці моделі можна використовувати для прогнозування ймовірності подій або для

класифікації даних за категоріями. Приклади типових застосувань моделі категоризації в BI-системах включають наступне:

Моделі кластеризації

- Сегментація клієнтів: За допомогою моделей категоризації клієнтів можна розділити на підгрупи відповідно до їх демографічних характеристик, купівельної поведінки та особистих смаків. Компанії можуть використовувати ці дані для вдосконалення своїх маркетингових стратегій і варіантів продуктів у відповідь на запити споживачів.
- Виявлення шахрайства: Моделі класифікації можна використовувати для виявлення підозрілих фінансових транзакцій шляхом виявлення відхилень і повторюваних шаблонів у даних. Організації можуть використовувати ці дані для кращого захисту своїх систем і зменшення потенційних фінансових втрат.
- Прогнозування відтоку клієнтів: Прогнозування того, які споживачі припинять або скасують підписку на основі попередньої поведінки та демографічних даних, можна зробити за допомогою моделей класифікації. Використовуючи ці дані, компанії можуть краще передбачити потреби клієнтів, що, в свою чергу, підвищує рівень їх утримання та довіри.
- Аналіз настроїв: Залежно від тональності та змісту відгуків клієнтів, моделі класифікації дозволяють класифікувати їх на позитивні, негативні та нейтральні. Маючи ці дані, бізнес може краще реагувати на потреби та запити клієнтів.

Ці методи аналізу даних стають дедалі популярнішими в сучасному бізнесі, оскільки допомагають організаціям краще розуміти своїх клієнтів, оптимізувати процеси та приймати обґрунтовані стратегічні рішення.

Елементи даних можна класифікувати відповідно до їхніх спільних характеристик за допомогою моделей кластеризації. Наприклад, компанії

можуть використовувати ці моделі для поділу клієнтів на підгрупи відповідно до їхніх демографічних характеристик, моделей покупок та особистих уподобань. Знання цих підгруп дозволяє компаніям краще таргетувати свою рекламу та розробляти продукти, які будуть цікавими для певної демографічної групи.

За допомогою цих моделей можна робити вибір на основі даних, що допоможе компаніям виявити закономірності та тенденції у своїй інформації. Серед поширених застосувань алгоритмів кластеризації в BI-системах можна виділити наступні:

З моменту появи BI-інструментів минуло понад 20 років, але дизайн і корисність "стандартних" BI-інструментів значно змінилися. Тепер кожен постачальник надає аналітикам не просто статичні звіти, а кастомізовані звіти та дашборди для спільної роботи. Крім того, самообслуговування BI стає нормою для рутинних бізнес-задач, що дає власникам малого бізнесу можливість заощаджувати на аналітиці. Останні розробки в області бізнес-аналітики включають хмарні платформи і мобільні BI-звіти, які відображають більш широкі тенденції в IT. Тож розуміючи загальноприйняті практики і технології в цій сфері, можна розробити власну індивідуальну систему бізнес-аналітики або вибрати вже існуючу. [17]

Використання бізнес-даних більше не обмежується лише вищими рівнями управління компанією. Доступ до даних для прийняття рішень з різним правами має бути на кожному рівні прийняття рішень.

Кожен інтернет-магазин відстежує дії та покупки клієнтів і генерує багато інформації, а машинне навчання відіграє вирішальну роль, аналізуючи ці дані і виявляє приховані патерни поведінки клієнтів, закономірності, статистику, інформацію та історії, що містяться в них.

Гнучкість машинного навчання є однією з його головних переваг над більш традиційними аналітичними методами наприклад, статистичні моделі або експертні системи. Гнучкість машинного навчання проявляється в його

здатності адаптуватися до різних типів даних і завдань без необхідності явного програмування. Воно може автоматично виявляти складні взаємозв'язки та закономірності у великих обсягах даних, що робить його ефективним інструментом для вирішення різноманітних завдань у різних галузях, від прогнозування до систем класифікації та рекомендацій.. Чим краща аналітика та прогнози алгоритму ML, тим більше даних йому потрібно обробити. Машинне навчання швидко поширюється в усіх секторах економіки: від сільського господарства та медицини до фондового ринку та відстеження трафіку. [18]

Впровадження машинного навчання для будь-яких бізнес-цілей тепер можливе завдяки таким онлайн-сервісам, як Azure Machine Learning та Amazon SageMaker. Бізнес-статистика та інтелект надають широкий спектр ключових переваг бізнес-даних для прийняття обґрунтованих стратегічних рішень:

- Визначення тенденції та закономірності
- Підвищення продуктивності та доходів
- Підвищення операційної ефективності
- Вдосконалення за допомогою прогнозів
- Розумніша та швидша звітність

РОЗДІЛ 3

АНАЛІЗ НАСТРОЇВ ЗА ДОПОМОГОЮ ІНСТРУМЕНТУ FinBERT

3.1. Поняття і методи аналізу настроїв у бізнесі

Аналіз настроїв - це процес визначення емоційного тону тексту, який може бути позитивним, негативним або нейтральним. Цей підхід відіграє важливу роль у різних галузях, включаючи фінанси. У фінансовому аналізі та прогнозуванні ринків аналіз сентиментів дозволяє виявляти та враховувати настрої ринку, що може бути корисним при прийнятті фінансових рішень. Позитивні або негативні емоції, виражені у фінансових новинах, соціальних медіа або фінансових звітах, можуть впливати на ціни акцій, курси валют та інші ринкові показники.[19]

Аналіз настроїв має застосування в маркетингу та брендингу, допомагає компаніям вимірювати ефективність своїх маркетингових кампаній і реагувати на відгуки споживачів. Вчасна реакція на негативні настрої може поліпшити сприйняття бренду та продуктів.

Загалом, аналіз настроїв стає все більш важливим інструментом для прийняття рішень у різних галузях, оскільки він дозволяє отримувати інсайди з великої кількості даних текстуального характеру та використовувати їх для прийняття обґрунтованих рішень.[20]

Визначення позитивних, негативних та нейтральних настроїв у текстах полягає у розпізнаванні емоційного тону, вираженого в тексті, та класифікації його як позитивний, негативний або нейтральний. Ось коротке уявлення про кожен з цих типів настроїв:

Позитивний настрій:

- Текст виражає радість, задоволення, оптимізм або підтримку щодо обговорюваної теми.
- Приклади: "Прекрасна новина про зростання прибутків компанії", "Клієнти захоплені новим продуктом".

Негативний настрій:

- Текст виражає незадоволеність, обурення, розчарування або негативний відгук щодо обговорюваної теми.
- Приклади: "Великий втрата на біржі призвела до падіння акційних цін", "Погані відгуки клієнтів про якість обслуговування".

Нейтральний настрій:

- Текст не виражає сильних емоцій або відчуттів, він може бути об'єктивним або інформативним.
- Приклади: "Презентовано звіт про фінансові показники компанії за останній квартал", "Сповіщення про розклад роботи на вихідних".

Для визначення настрою в тексті можуть застосовуватися різні методи, включаючи лексичні аналізатори, машинне навчання та глибоке навчання. Класифікація настроїв важлива для різних завдань, включаючи аналіз відгуків користувачів, вимірювання громадської думки та прийняття рішень у бізнесі.[21]

Методи аналізу настроїв включають традиційні методи (лексичні, машинне та глибоке навчання), кожен з яких має переваги та недоліки.

- **Лексичні методи** ґрунтуються на аналізі словникових елементів тексту, щоб визначити його настрій. Зазвичай використовуються списки слів з позитивними та негативними зарядами, і текст аналізується на предмет наявності цих слів. Одні з найпоширеніших лексичних методів - метод AFINN та метод SentiWordNet.
- **Методи машинного** навчання використовують навчальні дані для тренування моделей класифікації настроїв. Моделі можуть використовувати різноманітні алгоритми, такі як наївний Байєсівський класифікатор, метод опорних векторів (SVM), решітки рішень тощо. Для тренування таких моделей потрібні дані з відомими мітками сентименту.

- **Глибоке навчання:** глибокі нейронні мережі, такі як рекурентні нейронні мережі (RNN), згорткові нейронні мережі (CNN) та трансформери, можуть бути використані для аналізу настроїв. Ці методи дозволяють моделям автоматично виявляти складні зв'язки між словами та контекстом в тексті. Відомі архітектури, такі як LSTM (Long Short-Term Memory) та BERT (Bidirectional Encoder Representations from Transformers), широко застосовуються в аналізі сентименту. (таблиця 3.1)[22]

Порівняння зазначених вище методів узагальнене в таблиці 3.1.

Таблиця 3.1

Переваги та недоліки методів аналізу настроїв

Методи	Переваги	Недоліки
Лексичні методи	відносно прості у використанні	менш точні через відсутність врахування контексту та семантики речень
	ґрунтуються на словниках з позитивними та негативними словами	можуть мати різний настрій у різних контекстах
Методи машинного навчання	створювати більш складні моделі	потребують великої кількості навчальних даних та часу для тренування моделей
	враховувати багато різних факторів	
	гнучкість та можливість досягнення високої точності	значних обчислювальних ресурсів
Глибокі нейронні мережі	автоматично вивчати ознаки з великих обсягів даних	потребує великої кількості даних та обчислювальних ресурсів для тренування та застосування моделей
	включаючи складні взаємозв'язки між словами та контекстом	
	переконливої точності	

Кожен із цих методів має свої переваги та обмеження, і вибір конкретного методу може залежати від потреб проєкту, доступних ресурсів та призначення аналізу настроїв.

3.2. Моделі аналізу текстів BERT і FinBERT

Пояснення, як FinBERT відрізняється від базової моделі BERT та які адаптації вона має для фінансових текстів проаналізовано в роботі [23].

Опис моделі BERT та її можливостей у роботі з природними мовами

BERT (Bidirectional Encoder Representations from Transformers) - це модель глибокого навчання, розроблена компанією Google, яка використовує трансформерну архітектуру для обробки природної мови. Однією з основних особливостей BERT є його здатність до контекстно-залежного розуміння тексту, оскільки вона розглядає слова в контексті всього речення. Це дозволяє BERT краще розуміти складний зміст тексту та враховувати семантичні зв'язки між словами.[24]

Одним із ключових елементів BERT є двонаправлене кодування, яке означає, що модель розглядає кожне слово в тексті як частину двонаправленого контексту. Це дозволяє моделі здатність "дивитися" в обидва напрямки речення для кращого розуміння контексту, на відміну від існуючих аналогів.

Можливості BERT у роботі з природними мовами:

- **Контекстно-залежне розуміння**: BERT може аналізувати кожне слово в тексті, враховуючи його контекст у всьому реченні або документі. Це дозволяє моделі краще розуміти семантичні зв'язки мовлення.
- **Універсальність**: BERT може бути застосована до різних завдань обробки природної мови, таких як розпізнавання іменованих

сутностей, аналізу настроїв, машинного перекладу, питань та відповідей тощо.

- **Переносність**: Модель BERT може бути навчена на великому обсязі даних та перенесена на різні завдання та мови без значного перетренування.
- **Досягнення переконливих результатів**: BERT показав вражаючі результати у багатьох завданнях обробки природної мови та здатний досягати високої точності на різних даних та завданнях.

Модель FinBERT є варіацією базової моделі BERT, відрізняється від базової моделі BERT і адаптована для роботи з фінансовими текстами та специфічними особливостями фінансової мови.

- **Навчання на фінансових даних**: Фінансовий текст має свою унікальну лексику та структуру, тому FinBERT була навчена на великому обсязі фінансових даних, щоб краще розуміти цю мову та враховувати її унікальні особливості.
- **Адаптація до фінансових понять**: FinBERT має покращені механізми розуміння фінансових термінів, абревіатур, числових значень та інших концепцій, які є характерними для фінансових текстів. Це дозволяє моделі краще аналізувати та інтерпретувати фінансовий контекст.
- **Спеціалізовані виходи**: FinBERT може мати спеціалізовані виходи, призначені для конкретних фінансових завдань, таких як аналіз настрою відносно фінансових новин або прогнозування ринкових тенденцій.
- **Тонке налаштування параметрів**: Модель може бути тонко налаштована для кращої адаптації до конкретних потреб та вимог фінансового аналізу. Це може включати налаштування гіперпараметрів, таких як розмір моделі, швидкість навчання та інші.
- **Виправлення недоліків базової моделі**: FinBERT може включати виправлення недоліків, які можуть бути виявлені в базовій моделі BERT під час роботи з фінансовими даними. Наприклад, вона може

враховувати особливості фінансової лексики та виключати деякі звичайні проблеми, пов'язані з неоднорідністю фінансових даних.

Загалом, FinBERT створюється з урахуванням специфічних потреб фінансової галузі та має за мету покращити результати аналізу настроїв, прогнозування та інші завдання, пов'язані з фінансовими текстами. (рис.3.1)

Рис.3.1 Етапи навчання в FinBERT

3.3. Архітектура та методи навчання FinBERT

Архітектура FinBERT містить 3 головні компоненти (рис.3.2).

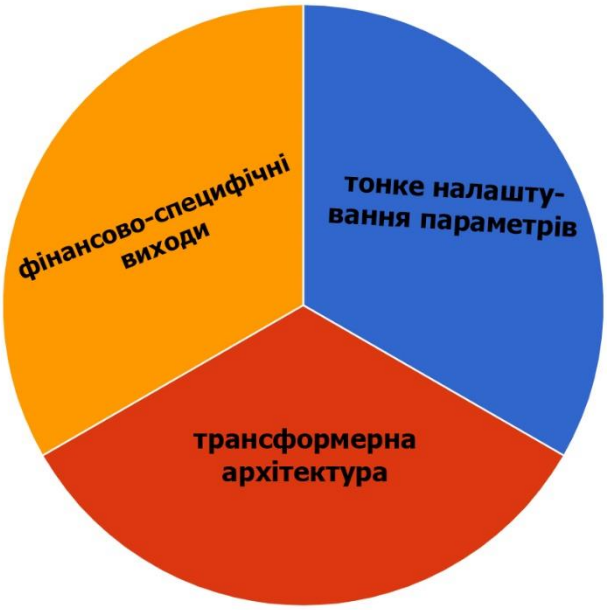


Рис.3.2 Архітектура FinBERT

FinBERT успадковує основну структуру та механізми трансформерної архітектури BERT, включаючи механізми уваги та механізми кодування-декодування, що дозволяють моделі розуміти та аналізувати текст. Для адаптації до фінансових завдань, FinBERT може мати спеціалізовані виходи, які допомагають у вирішенні конкретних фінансових задач, таких як аналіз настрою відносно фінансових новин або прогнозування ринкових тенденцій. Архітектура FinBERT може бути налаштована з урахуванням специфічних потреб фінансової галузі та завдань, що стоять перед моделлю. Це включає в себе налаштування гіперпараметрів, які дозволяють підвищити ефективність та точність моделі в контексті фінансових текстів.[25]

Методи навчання FinBERT базуються на принципах глибокого навчання та машинного навчання. Модель навчається на великому обсязі фінансових даних, включаючи новини, звіти, соціальні медіа та інші джерела текстів з фінансовою тематикою. Під час навчання модель адаптує свої ваги та параметри таким чином, щоб максимально точно відображати особливості фінансових текстів та завдань. Такий підхід дозволяє досягнути високої ефективності та точності в аналізі фінансових текстів за допомогою FinBERT.

FinBERT використовується для аналізу сентиментів у фінансових текстах шляхом використання своєї адаптованої архітектури та навченої моделі для розуміння та інтерпретації специфічної мови фінансових даних.

- Підготовка даних: Перш ніж FinBERT може проводити аналіз настрою, потрібно підготувати фінансові тексти для подальшої обробки. Це може включати усунення шуму, очищення від спеціальних символів та інші передпроцесингові операції.
- Векторизація тексту: Фінансові тексти перетворюються на числові вектори за допомогою вбудованих технік векторизації, таких як методи TF-IDF або Word Embeddings. Ці вектори використовуються для подальшого подання тексту в моделі FinBERT.

- Аналіз настрою з використанням FinBERT: Очищені та векторизовані тексти передаються до моделі FinBERT. Модель аналізує кожен текст у контексті його фінансового значення та інтерпретує семантичні зв'язки між словами та реченнями, враховуючи їхній контекст та синтаксичні особливості.
- Визначення настрою: На основі аналізу тексту FinBERT визначає настрій кожного речення або документа. Це може бути позитивний, негативний або нейтральний настрій, в залежності від того, як модель інтерпретує зміст тексту та його відтінки.
- Інтерпретація результатів: Після аналізу FinBERT повертає результати, які можна інтерпретувати для прийняття рішень у фінансових сферах. Наприклад, позитивний настроій може свідчити про оптимістичне ставлення до певної події чи компанії, тоді як негативний настрій може попереджати про потенційні ризики або проблеми.

1) Застосування FinBERT у фінансовій сфері:
    - Приклади використання FinBERT для аналізу новин, соціальних медіа, фінансових звітів та інших джерел інформації.
    - Потенційні переваги використання FinBERT для прийняття фінансових рішень та прогнозування ринкових тенденцій.

Приклади використання FinBERT

Використання FinBERT включає аналіз настрою в різних джерелах фінансової інформації з метою оцінки впливу подій та новин на ринки та інвестиційні рішення. Модель може бути використана для аналізу настрою у фінансових новинах, соціальних медіа, фінансових звітах та інших джерелах інформації. Наприклад, вона може аналізувати настрій у новинах про фінансові події, такі як оголошення прибутковості підприємства, оцінювати громадську думку щодо певних компаній на основі соціальних медіа, аналізувати фінансові звіти компаній для оцінки їхньої фінансової стійкості та прогнозування ринкових тенденцій, а також моніторити настрій на фінансових ринках з метою прийняття обґрунтованих інвестиційних рішень.

В цілому, FinBERT надає засоби для аналізу настроїв у різних джерелах фінансової інформації, допомагаючи інвесторам та аналітикам приймати обґрунтовані рішення у фінансовій галузі.

Переваги використання FinBERT для прийняття фінансових рішень та прогнозування ринкових тенденцій продемонстровані в Таблиці 3.2.

Таблиця 3.2

Переваги використання FinBERT

Переваги FinBERT	**Опис**
Аналіз настроїв ринку	FinBERT дозволяє аналізувати настрій у фінансових текстах, що дозволяє зрозуміти настрої ринку та реакції інвесторів на певні події або новини.
Швидке реагування	Завдяки використанню ШІ та аналізу настрою в реальному часі, FinBERT допомагає інвесторам швидко реагувати на зміни на ринку та приймати обґрунтовані рішення.
Покращення точності прогнозів	Аналіз настрою з використанням FinBERT може допомогти у покращенні точності прогнозів ринкових тенденцій та курсів активів.
Зниження ризиків	Фінансові рішення, підтримані аналізом настрою FinBERT, можуть допомогти зменшити ризики в інвестиціях та управлінні портфелем.
Підвищення ефективності аналізу	Використання FinBERT дозволяє автоматизувати процес аналізу фінансових текстів, що підвищує ефективність та

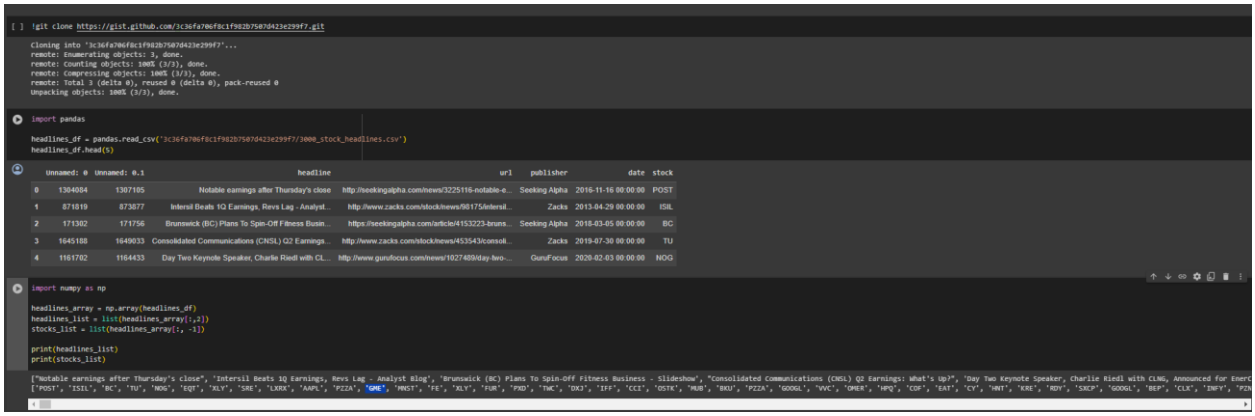
	швидкість прийняття рішень.
Додаткові джерела інформації	Аналіз настрою за допомогою FinBERT дозволяє використовувати додаткові джерела інформації для узгодження з аналітичними даними та іншими інформаційними джерелами.

Загалом, використання FinBERT може сприяти покращенню процесу прийняття фінансових рішень та прогнозуванню ринкових тенденцій шляхом аналізу настроїв у реальному часі та забезпечення більш обґрунтованих стратегій інвестування.

3.4. Аналіз настроїв на основі заголовків біржових новин

Перед тим, як працювати з будь-якою інформацією, треба мати уявлення про те, навіщо вона потрібна, кому, та у яких випадках. Відповіді на перші два питання були знайдені після вивчення матеріалів, пов’язаних з обробкою інформації та управління із застосуванням технічних інновацій. Ця попередньо навчена модель створена для бізнес аналітиків, для розбору концепції аналізу сентиментів у фінансових текстах щоб у подальшому розуміти куди направлений вектор аудиторії.

На основі підготовлених даних у вигляді 3000 заголовків біржових новин за допомогою навченої для аналізу природної мови моделі FinBERT було проаналізовано їх настрій. Для зручності сортування та кращої



візуалізації та ініціалізації отриманих даних будуть підключені таблиці W&B.(рис. 3.3)

Рис. 3.3. Підключення Git-клонування яке містить в собі 3000 заголовків біржових нових.

Git-клонування - це процес створення копії репозиторію Git, яка зазвичай знаходиться на віддаленому сервері. Важливу роль в цьому процесі виконує токенізатор. Токенізатор в обробці тексту використовується для розбиття текстового корпусу на окремі одиниці, які називаються токенами, що необхідно для аналізу семантики тексту та збору статистики і подальшої обробки (рис. 3.4).

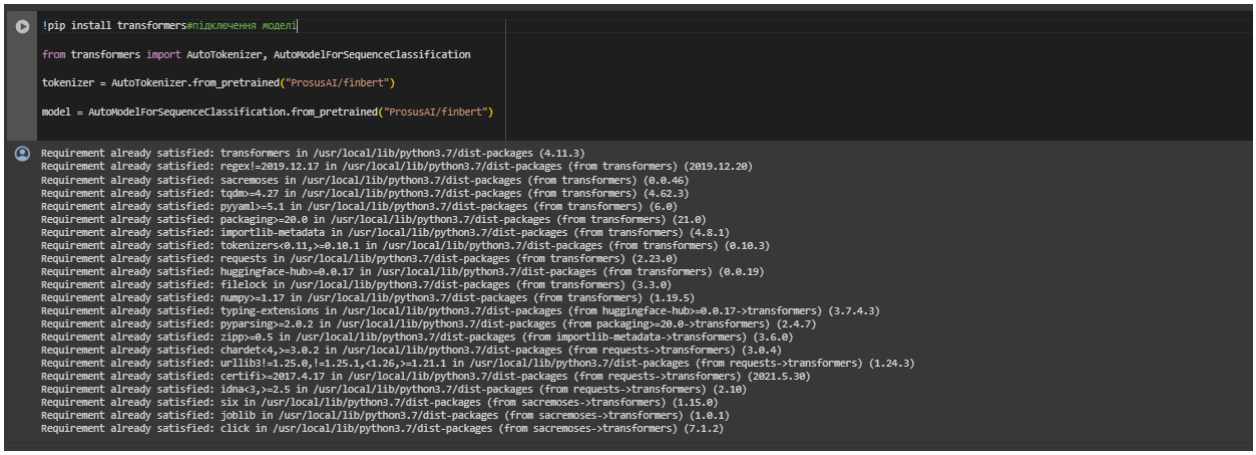


Рис. 3.4. Підключення токенізатора та моделі FinBERT

Оскільки FinBERT – це лише попередньо навчена модель NLP для аналізу настрою фінансового тексту, чіткого відображення отриманих даних в ній немає, тому за допомогою таблиці W&B є можливість візуалізувати результат тому їх підключення є необхідним (рис. 3.5).

Таблиці W&B (Weights & Biases) - це платформа для моніторингу, візуалізації та аналізу результатів експериментів у машинному навчанні та дослідженнях штучного інтелекту. Вона надає набір інструментів для відстеження метрик, візуалізації результатів, спільної роботи та зберігання моделей та даних.

```
[ ] !pip install wandb
    import wandb

    wandb.init(project="FinBERT_Stock_Sentiment_Analysis")

    headlines_table = wandb.Table(columns=["Headline", "Stock", "Positive", "Negative", "Neutral"])
```

```
Requirement already satisfied: wandb in /usr/local/lib/python3.7/dist-packages (0.12.4)
Requirement already satisfied: yaspin>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (2.1.0)
Requirement already satisfied: six>=1.13.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.15.0)
Requirement already satisfied: sentry-sdk>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.4.3)
Requirement already satisfied: GitPython>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (3.1.24)
Requirement already satisfied: shortuuid>=0.5.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.0.1)
Requirement already satisfied: docker-pycreds>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (0.4.0)
Requirement already satisfied: Click!=8.0.0,>=7.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (7.1.2)
Requirement already satisfied: subprocess32>=3.5.3 in /usr/local/lib/python3.7/dist-packages (from wandb) (3.5.4)
Requirement already satisfied: configparser>=3.8.1 in /usr/local/lib/python3.7/dist-packages (from wandb) (5.0.2)
Requirement already satisfied: protobuf>=3.12.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (3.17.3)
Requirement already satisfied: PyYAML in /usr/local/lib/python3.7/dist-packages (from wandb) (6.0)
Requirement already satisfied: requests<3,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (2.23.0)
Requirement already satisfied: pathtools in /usr/local/lib/python3.7/dist-packages (from wandb) (0.1.2)
Requirement already satisfied: python-dateutil>=2.6.1 in /usr/local/lib/python3.7/dist-packages (from wandb) (2.8.2)
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (5.4.8)
Requirement already satisfied: promise<3,>=2.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (2.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from GitPython>=1.0.0->wandb) (3.7.4.3)
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.7/dist-packages (from GitPython>=1.0.0->wandb) (4.0.7)
Requirement already satisfied: smmap<5,>=3.0.1 in /usr/local/lib/python3.7/dist-packages (from gitdb<5,>=4.0.1->GitPython>=1.0.0->wandb) (4.0.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.0.0->wandb) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.0.0->wandb) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.0.0->wandb) (2021.5.30)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.0.0->wandb) (1.24.3)
Requirement already satisfied: termcolor<2.0.0,>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from yaspin>=1.0.0->wandb) (1.1.0)
wandb: Currently logged in as: ivangoncharov (use `wandb login --relogin` to force relogin)
Syncing run fallen-oath-5 to Weights & Biases (docs).
```

Рис. 3.5. Підключені таблиці W&B

```
import torch

def chunk_list(lst, n):
    """Yield successive n-sized chunks from lst."""
    for i in range(0, len(lst), n):
        yield lst[i:i + n]


STRIDE = 100

model.eval()

n=0
for lines, stocks in zip(chunk_list(headlines_list, STRIDE), chunk_list(stocks_list, STRIDE)):

  input = tokenizer(lines, padding = True, truncation = True,  return_tensors='pt')

  outputs = model(**input)

  prediction = torch.nn.functional.softmax(outputs.logits, dim=-1)

  print(f"{n+1}/{int(len(headlines_list)/STRIDE)}")

  for headline, stock, pos, neg, neutr in zip(lines,stocks, prediction[:, 0].tolist(), prediction[:, 1].tolist(), prediction[:, 2].tolist() ):
    headlines_table.add_data(headline, stock, pos, neg, neutr)

  n+=1
```

Рис. 3.6 Виконання виводу на заголовках новин фондового ринку за допомогою моделі FinBERT, реалізованої в HuggingFace

	Headline	Positive	Negative	Neutral
1	Notable earnings after Thursday's close	0.89	0.07463	0.03535
2	Intersil Beats 1Q Earnings, Revs Lag - Analyst Blog	0.9024	0.03684	0.06077
3	Brunswick (BC) Plans To Spin-Off Fitness Business - Slideshow	0.02403	0.02708	0.9489
4	Consolidated Communications (CNSL) Q2 Earnings: What's Up?	0.03917	0.5518	0.409
5	Day Two Keynote Speaker, Charlie Riedl with CLNG, Announced for EnerCom Dallas Energy ...	0.1516	0.008106	0.8403
6	EQT Rubbishes JANA Claims on Rice Energy Buyout Synergies	0.09311	0.04176	0.8651
7	Retailers look for government help for shipping crisis	0.1109	0.08787	0.8012

1 - 7 of 3000 Export as CSV Columns... Reset table

Рис. 3.7. Отриманні результати у вигляді таблиці W&B

Взявши як приклад один із біржових заголовків який звучить як «Notable earnings after Thursday's close» модель дала результат позитивний – 0,89, негативний – 0,07, нейтральний – 0,03 що означає що це позитивна новина.

Ця попередньо навчена модель NLP для аналізу настрою фінансового тексту відкриває нові можливості у аналізі ринку. Завдяки швидкості роботи вона може стати основою для більш складних програм які будуть отримувати твіти фінансових новин оцінювати їх і створювати реальний огляд ринку з точки зору фінансових настроїв.

Це наддасть додаткову перспективу для фінансових аналітиків адже швидкість на біржі грає одну з ключових ролей. Як альтернатива FinBERT можна використовувати для вдосконалення повідомлень перед публікацією або для оцінки настроїв у корпоративних публікаціях фінансових звітах або бізнес-блогах.

ВИСНОВКИ

Розглянута інформації щодо застосування інструментів бізнес-аналітики у концепції аналізу настроїв у фінансових текстах для допомоги у прийнятті інвестиційних рішень.

Проаналізовані методологія розбору концепції аналізу настроїв та його застосування у різних галузях, включаючи фінанси.

На основі підготовлених даних у вигляді 3000 заголовків біржових новин за допомогою навченої для аналізу природної моделі FinBERT проведений аналіз їх настроїв для обґрунтування фінансових рішень.

Автоматизований аналіз настроїв здатен швидко та ефективно виявляти настрої та емоції, виражені у фінансових звітах, новинах та інших джерелах. Цей аналіз дозволяє зрозуміти реакцію ринку на певні події та тенденції, що допомагає інвесторам та фінансовим аналітикам приймати обґрунтовані рішення.

Дослідження показало, що використання методів машинного навчання в поєднанні з аналізом настроїв може значно покращити прогнозування ринкових тенденцій та ризиків, сприяючи підвищенню ефективності фінансових стратегій та прийняттю кращих інвестиційних рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. https://www.sciencedirect.com/science/article/pii/S0048733313001790?via%3Dihub Teppo Felin, Todd R. Zenger Closed or open innovation? Problem solving and the governance choice, Research Policy Volume 43, Issue 5, June 2014, Pages 914-925
2. https://www.sciencedirect.com/science/article/abs/pii/S0040162521005928?via%3Dihub Ma, T., Zhou, X., Liu, J., Lou, Z., Hua, Z., Wang, R. Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies. Technological Forecasting and Social Change, Volume 173, December 2021, 121159. https://doi.org/10.1016/j.techfore.2021.121159
3. https://www.sciencedirect.com/science/article/abs/pii/S0166497221001887?via%3Dihub Suckwon Hong, Juram Kim, Han-Gyun Woo, Young-Choon Kim , Changyong Lee, Screening ideas in the early stages of technology development: A word2vec and convolutional neural network approach, Volume 112, April 2022, 102407
4. https://aclanthology.org/2021.acl-short.96/ Federico Bianchi, Silvia Terragni, Dirk Hovy, 2021.acl-short.96 Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)
5. https://dl.acm.org/doi/10.1145/3434237Usman Naseem, Author PictureImran Razzak, Author PictureShah Khalid Khan, Author PictureMukesh Prasad A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models 30 June 2021
6. https://journals.sagepub.com/doi/10.1177/1094428120986851 Martin R. W. Hiebl, Sample Selection in Systematic Literature Reviews of Management Research, January 29, 2021

7. https://www.emerald.com/insight/content/doi/10.1108/JBS-05-2023-0094/full/html Paavo Ritala, Mika Ruokonen, Laavanya Ramaul, Transforming boundaries: how does ChatGPT change knowledge work?, 26 March 2024.
8. https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/143/6/061403/1088904/Eliciting-Attribute-Level-User-Needs-From-Online?redirectedFrom=fulltext Yi Han, Mohsen Moghaddam, Eliciting Attribute-Level User Needs From Online Reviews With Deep Language Models and Information Extraction, J. Mech. Des. Jun 2021.
9. https://ieeexplore.ieee.org/document/9851937 Randy A. Williams; Nasir Jamil Sheikh; Gazi Murat Duman; Elif KongarJ. Mech. Des. Jun 2021, Critical Success Factors of Business Intelligence Systems Implementation
10. https://ieeexplore.ieee.org/document/9468657 Julian A. García-García; C. Arevalo Maldonado; Ayman Meidan; Esteban Morillo-Baro; María José EscalonaCritical Success Factors of Business Intelligence Systems Implementation, A Tool to Assist the Automatic Extraction of Business Knowledge From Legacy Information Systems, 30 June 2021.
11. https://ieeexplore.ieee.org/document/8825774 Ahmad Aburomman; Manuel Lama; Alberto Bugarín, A Vector-Based Classification Approach for Remaining Time Prediction in Business Processes, 05 September 2019.
12. https://www.sciencedirect.com/science/article/pii/S240591881730048X?via%3Dihub Adam Atkins, Mahesan Niranjan, Enrico Gerding, Financial news predicts stock market volatility better than close price, Volume 4, Issue 2, June 2018.
13. https://www.emerald.com/insight/content/doi/10.1108/AJB-08-2015-0024/full/html Laura K. Rickett, Do financial blogs serve an infomediary role in capital markets?, 4 April 2016.

14. https://www.sciencedirect.com/book/9780123748560/data-mining-practical-machine-learning-tools-and-techniques?via=ihub= A volume in The Morgan Kaufmann Series in Data Management Systems, Practical Machine Learning Tools and Techniques, 2011.
15. https://www.nature.com/articles/srep03578 Merve Alanyali, Helen Susannah Moat & Tobias Preis, Quantifying the Relationship Between Financial News and the Stock Market, 20 December 2013.
16. https://ieeexplore.ieee.org/document/9272732 Francisco De Arriba-Pérez; Silvia García-Méndez; José Ángel Regueiro-Janeiro; Francisco J. González-Castaño, Detection of Financial Opportunities in Micro-Blogging Data With a Stacked Classification System, 27 November 2020.
17. https://onlinelibrary.wiley.com/doi/10.1002/isaf.1386 Ingrid E. Fisher, Margaret R. Garnsey, Mark E. Hughes, Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research, 01 March 2016.
18. https://www.sciencedirect.com/science/article/abs/pii/S0925527319300404?via%3Dihub You Zhu, Li Zhou, Chi Xie , Gang-Jin Wang, Truong V. Nguyen, Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach, Volume 211, May 2019, Pages 22-33.
19. https://www.sciencedirect.com/science/article/abs/pii/S0040162516301202?via%3Dihub Silvia Mayumi Takey, Marly M. Carvalho, Fuzzy front end of systemic innovations: A conceptual framework based on a systematic literature review, Volume 111, October 2016, Pages 97-109.
20. https://comsys.kpi.ua/upload/Rainforcement%20Learning%20.pdf V. N. Karazin Kharkiv National University, МАШИННЕ НАВЧАННЯ: МЕТОДИ ТА МОДЕЛІ November 2020.
21. https://www.sciencedirect.com/science/article/abs/pii/S0959652618327082 Jizhou Zhan a b, Shuting Li c, Xiangfeng Chen, The impact of

financing mechanism on supply chain sustainability and efficiency, Volume 205, 20 December 2018, Pages 407-418

22. https://habr.com/ru/companies/Voximplant/articles/446738/ Muhammad Adnan Khan; Shazia Saqib; Tahir Alyas; Anees Ur Rehman; Yousaf Saeed; Asim Zeb; Mahdi Zareei, Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning, 19 June 2020.
23. https://www.researchgate.net/publication/364070191_FinBERT_A_Large_Language_Model_for_Extracting_Information_from_Financial_Text Allen Huang, Hui Wang, Yi Yang, FinBERT: A Large Language Model for Extracting Information from Financial Text, September 2022.
24. https://www.researchgate.net/publication/342793634_FinBERT_A_Pre-trained_Financial_Language_Representation_Model_for_Financial_Text_Mining Zhuang Liu Dalian University of Technology, Degen Huang Dalian University of Technology, Huang Kaiyu Dalian University of Technology, FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining, July 2020.
25. https://analyticsindiamag.com/the-gap-between-these-state-of-the-art-generative-language-models-and-domain-specific-proficiency-remains/ Pranav Kashyap, Unveiling the AI Potential: FinBert beats ChatGPT in Financial Text Analytics, September 15, 2023.