



## Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ комп'ютерних наук \_\_\_\_\_  
Кафедра \_\_\_\_\_ програмної інженерії \_\_\_\_\_  
Рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_  
Спеціальність \_\_\_\_\_ 121 – Інженерія програмного забезпечення \_\_\_\_\_  
(шифр і назва)  
Тип програми \_\_\_\_\_ освітньо-наукова програма \_\_\_\_\_  
Освітня програма \_\_\_\_\_ Інженерія програмного забезпечення \_\_\_\_\_

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

«\_\_\_» \_\_\_\_\_ 2024 р.

**ЗАВДАННЯ  
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студентові \_\_\_\_\_ Годованцю Денису Сергійовичу \_\_\_\_\_  
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження методів визначення актуальності та релевантності запитань до комерційних пропозицій.»  
Затверджена наказом по університету від 29.03. 2024р. № 250 Ст
2. Термін подання студентом роботи до екзаменаційної комісії 25.06.2024
3. Вихідні дані до роботи: Дослідження методів визначення актуальності та релевантності запитань до комерційних пропозицій.
4. Перелік питань, що потрібно опрацювати в роботі: Метою роботи є дослідження методів визначення актуальності та релевантності запитань до комерційних пропозицій. Основні завдання включають аналіз та порівняння існуючих методів і технік визначення актуальності запитань, доскональне вивчення принципів роботи кожного з обраних підходів, розробку рекомендацій щодо впровадження найбільш ефективних методів, експериментальне визначення релевантності запитань до комерційних пропозицій та проведення тестування і оптимізації розробленого рішення.

**КАЛЕНДАРНИЙ ПЛАН**

№	Назви етапів курсової роботи	Термін виконання етапів роботи	Примітка
1	Видача теми, узгодження і затвердження	02.04.2024	виконано
2	Аналіз предметної галузі	02.04.2024	виконано
3	Огляд існуючих методів	09.04.2024	виконано
4	Оформлення пояснювальної записки	05.05.2024	виконано
5	Здача готового проекту	25.06.2024	виконано

Дата видачі завдання 20 січня 2024 р.

Студент \_\_\_\_\_  
(підпис)

Годованець Д.С.  
(прізвище, ініціали)

Керівник кваліфікаційної роботи \_\_\_\_\_  
(підпис)

доц. Бабій А.С.  
(посада, прізвище, ініціали)

**РЕФЕРАТ / ABSTRACT**

Пояснювальна записка: 58 с., 10 рисунків, 14 джерел.

**АКТУАЛЬНІСТЬ, ЗАПИТАННЯ, КОМЕРЦІЙНІ ПРОПОЗИЦІЇ, МЕТОДИ,  
РЕЛЕВАНТНІСТЬ,**

Об'єктом глибокого аналізу та дослідження є методи визначення актуальності та релевантності запитань до комерційних пропозицій. Задачею дослідження є аналіз, розробка та впровадження ефективних методів і технік для оцінки актуальності та релевантності запитань, що сприяє підвищенню якості та ефективності комерційних взаємодій.

У процесі наукового дослідження застосовуються передові підходи сучасних технологій, зокрема ті, що пов'язані з обробкою даних та машинним навчанням. Великий акцент приділяється розгляду різноманітних алгоритмів і методів, їх адаптації до потреб конкретних завдань, а також інтеграції з існуючими системами управління комерційними пропозиціями.

Основні напрямки дослідження включають аналіз існуючих методів визначення актуальності та релевантності запитань, визначення ключових вимог до їх оцінки та розробку рекомендацій щодо впровадження найбільш ефективних методів. Серед досліджуваних методів – алгоритми ранжування інформації, нейронні мережі, методи машинного навчання та інші інноваційні підходи.

В результаті дослідження розробляються та впроваджуються комплексні стратегії для оцінки актуальності та релевантності запитань. Це сприяє не лише покращенню якості та продуктивності комерційних взаємодій, але й забезпечує успішну інтеграцію змін у динамічному середовищі сучасних технологій. Комбінування різних алгоритмів дозволяє досягати більш високих результатів у пошуку оптимальних рішень, враховуючи різноманітні потреби та обмеження в реальному часі.

## RELEVANCE, PERTINENCE, QUESTIONS, COMMERCIAL PROPOSALS, METHODS.

The object of deep analysis and research is the methods for determining the relevance and pertinence of questions to commercial proposals. The aim of the study is to analyze, develop, and implement effective methods and techniques for assessing the relevance and pertinence of questions, which contributes to improving the quality and efficiency of commercial interactions.

The research process employs advanced approaches of modern technologies, particularly those related to data processing and machine learning. Significant emphasis is placed on reviewing various algorithms and methods, adapting them to the needs of specific tasks, and integrating them with existing commercial proposal management systems.

The main areas of research include the analysis of existing methods for determining the relevance and pertinence of questions, identifying key requirements for their assessment, and developing recommendations for implementing the most effective methods. The methods under study include information ranking algorithms, neural networks, machine learning methods, and other innovative approaches.

As a result of the research, comprehensive strategies for assessing the relevance and pertinence of questions are developed and implemented. This not only improves the quality and productivity of commercial interactions but also ensures the successful integration of changes in the dynamic environment of modern technologies. Combining various algorithms allows for achieving higher results in finding optimal solutions, taking into account diverse needs and real-time constraints.

Я, Годованець Денис Сергійович, студент гр. ПЗм-22-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя робота на тему «Дослідження методів визначення актуальності та релевантності запитань до комерційних пропозицій», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений(на) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

## ЗМІСТ

Перелік скорочень.....	8
Вступ.....	9
1 Аналіз предметної галузі .....	11
1.1 Загальний опис комерційних пропозицій .....	11
1.2 Актуальність проблеми.....	13
1.3 Постановка задачі.....	16
2 Аналіз перспективних методів та алгоритмів.....	18
2.1 Машинне навчання.....	18
2.2 Методи на основі правил .....	21
2.3 Штучні нейронні мережі.....	24
3 Порівняння рішень та їх модифікація .....	27
3.1 Вибір методу аналізу.....	27
3.2 Порівняння бібліотек машинного навчання .....	28
3.3 Вибір ключового функціоналу BERT .....	31
4 Проведення експерименту .....	35
4.1 Обрані підходи.....	35
4.2 Архітектура системи .....	35
4.3 Data-sets для тренування та тестування .....	40
4.4 Аналіз результатів .....	41
Висновки.....	44
Перелік джерел посилання .....	46
Додаток А Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії .....	48
Додаток Б Слайди презентації.....	49
Додаток В Результат проходження на академічний плагіат.....	56
Додаток Г Апробація результатів роботи.....	57
Додаток Д Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015 .....	58

**ПЕРЕЛІК СКОРОЧЕНЬ**

ML.NET - Machine Learning .NET

BERT - Bidirectional Encoder Representations from Transformers

ONNX - Open Neural Network Exchange

ASP.NET - Active Server Pages .NET

RNN - Recurrent Neural Network

LSTM - Long Short-Term Memory

GRU - Gated Recurrent Units

BERT - Bidirectional Encoder Representations from Transformers

GPT - Generative Pre-trained Transformer

## ВСТУП

Дослідження методів визначення актуальності та релевантності запитань до комерційних пропозицій є надзвичайно важливим у сучасному бізнес-середовищі, де швидкий і точний аналіз інформації має вирішальне значення для досягнення конкурентних переваг. У процесі підготовки та оцінки комерційних пропозицій компанії стикаються з необхідністю швидко і ефективно визначати, які запитання є найбільш актуальними та релевантними для їхніх конкретних потреб і вимог. Ця задача ускладнюється великою кількістю інформації, яку необхідно обробити, а також різноманітністю контекстів, у яких можуть бути поставлені запитання.

Забезпечення актуальності запитань передбачає їх відповідність поточним потребам компанії та ринку, тоді як релевантність означає відповідність запитань конкретним умовам і ситуаціям, в яких вони виникають. У цьому контексті, дослідження методів визначення актуальності та релевантності запитань стає ключовим аспектом для оптимізації процесу прийняття рішень і підвищення ефективності комерційних взаємодій.

Використання сучасних технологій обробки даних та алгоритмів машинного навчання дозволяє автоматизувати цей процес, знижуючи час і ресурси, необхідні для аналізу великих обсягів інформації. Водночас, важливо враховувати специфіку різних галузей та індивідуальні потреби компаній, що вимагає адаптації методів під конкретні умови та вимоги.

Це дослідження спрямоване на вивчення існуючих підходів і розробку нових методів визначення актуальності та релевантності запитань до комерційних пропозицій. Зокрема, увага буде приділена аналізу алгоритмів ранжування інформації, застосуванню нейронних мереж і інших методів штучного інтелекту для покращення точності та швидкості оцінки. Крім того, будуть розглянуті аспекти інтеграції цих методів у існуючі системи управління комерційними пропозиціями, а також їх вплив на загальну ефективність бізнес-процесів.

Таким чином, результати цього дослідження матимуть практичну цінність для компаній, що прагнуть покращити свої процеси підготовки та оцінки

комерційних пропозицій, забезпечуючи більш точну та своєчасну інформацію для прийняття рішень.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Загальний опис комерційних пропозицій

Комерційні пропозиції є важливим інструментом у сфері бізнесу, оскільки вони виступають офіційними документами, що визначають умови співпраці між сторонами, які бажають укласти угоду. Це письмові пропозиції, спрямовані від однієї компанії до іншої, з метою продажу товарів або надання послуг. Вони є ключовими елементами в процесі укладання контрактів та ведення переговорів, оскільки містять детальну інформацію про продукти, послуги, ціни, умови доставки, гарантії та інші аспекти, що мають значення для потенційного клієнта.

Комерційна пропозиція, зазвичай, складається з декількох основних частин: вступу, де коротко описується суть пропозиції та її значення для потенційного клієнта; основного тексту, який містить детальний опис товарів або послуг, їх технічні характеристики, переваги та унікальні особливості; цінової інформації, де вказані вартість продукції або послуг, можливі знижки, умови оплати та інші фінансові аспекти; умов співпраці, включаючи терміни доставки, гарантії, політику повернення та інші юридичні деталі; та заключної частини, яка підсумовує ключові моменти пропозиції і заохочує до укладення угоди.

Ефективність комерційної пропозиції залежить від її здатності чітко і переконливо донести до потенційного клієнта всі переваги співпраці. Важливими аспектами є не тільки зміст, але й оформлення документа, його структура, мова та стиль викладу. Пропозиція має бути зрозумілою, лаконічною та орієнтованою на конкретні потреби клієнта. Використання персоналізованих підходів, де враховуються індивідуальні особливості та вимоги клієнта, значно підвищує шанси на успішне укладення угоди.

Сучасні технології та методи автоматизації дозволяють значно спростити процес підготовки комерційних пропозицій, забезпечуючи високу точність і швидкість обробки інформації. Застосування алгоритмів машинного навчання, зокрема, дозволяє аналізувати великі обсяги даних, визначати найкращі пропозиції для конкретних клієнтів, автоматично генерувати тексти на основі заданих параметрів і багато іншого. Це сприяє підвищенню ефективності бізнес-

процесів, зменшенню витрат часу і ресурсів, а також підвищенню задоволеності клієнтів.

Отже, комерційні пропозиції є невід'ємною частиною бізнес-комунікацій, що сприяє розвитку партнерських відносин та укладенню вигідних угод. Їх підготовка вимагає врахування багатьох факторів, включаючи точність і повноту інформації, індивідуальний підхід до клієнта та використання сучасних технологій для автоматизації процесів.

Комерційні пропозиції мають глибокий вплив на процеси продажу та взаємодії з клієнтами, оскільки вони є основним інструментом для переконання потенційних клієнтів у перевагах пропонованих товарів або послуг. Їх підготовка вимагає комплексного підходу, що включає аналіз ринку, розуміння потреб клієнтів, визначення конкурентних переваг і формулювання чітких і переконливих аргументів.

Роль комерційних пропозицій у бізнесі важко переоцінити. Вони є своєрідною візитівкою компанії, що формує перше враження про неї у потенційного клієнта. Якісно підготовлена пропозиція може значно підвищити шанси на успішне укладення угоди, тоді як недоліки в її підготовці можуть призвести до втрати клієнта і, відповідно, можливих доходів. Тому компанії часто приділяють значну увагу створенню стандартів та шаблонів для підготовки комерційних пропозицій, а також навчання співробітників, які займаються цим процесом.

Успішні комерційні пропозиції зазвичай характеризуються кількома ключовими аспектами. По-перше, вони повинні бути персоналізованими, тобто враховувати специфічні потреби та запити конкретного клієнта. Це означає, що підготовка пропозиції повинна базуватися на детальному аналізі клієнта та його бізнесу, а також на розумінні того, які проблеми і завдання він прагне вирішити за допомогою пропонованих товарів або послуг.

По-друге, важливим є чітке та зрозуміле формулювання переваг. Комерційна пропозиція повинна демонструвати, чому саме цей продукт або послуга є найкращим вибором для клієнта. Це може включати такі аспекти, як

унікальні технічні характеристики, вища якість, нижча ціна, кращі умови гарантії або підтримки, а також будь-які інші переваги, які можуть бути важливими для клієнта.

По-третє, ефективність комерційної пропозиції залежить від її структурованості та доступності для сприйняття. Документ повинен мати логічну структуру, що дозволяє легко знайти необхідну інформацію. Використання підзаголовків, списків, графіків та інших засобів візуалізації може значно покращити сприйняття інформації.

Крім того, у сучасних умовах важливо використовувати технологічні засоби для автоматизації процесу підготовки комерційних пропозицій. Це дозволяє не тільки зекономити час і ресурси, але й підвищити точність та якість підготовлених документів. Зокрема, алгоритми машинного навчання та штучного інтелекту можуть бути використані для аналізу великих обсягів даних, визначення найбільш релевантних пропозицій для конкретних клієнтів, а також для автоматичного створення текстів комерційних пропозицій.

Таким чином, комерційні пропозиції є важливим інструментом у бізнесі, що сприяє розвитку партнерських відносин та укладенню вигідних угод. Їх ефективна підготовка вимагає врахування багатьох факторів, включаючи персоналізацію, чітке формулювання переваг, структурованість та використання сучасних технологій.

## 1.2 Актуальність проблеми

У сучасному світі, де час і ресурси є ключовими факторами успіху в будь-якому бізнесі, актуальність створення автоматизованих розкладів набуває особливої ваги.

Використання автоматизованих систем для створення розкладів не тільки підвищує ефективність використання ресурсів, але й відіграє вирішальну роль у збалансуванні робочих навантажень та задоволенні потреб персоналу.

Автоматизація розкладів зводить до мінімуму людські помилки та знижує час, необхідний для їх розробки. Традиційний підхід до складання розкладів

вручну часто є часозатратним та схильним до помилок, оскільки вимагає врахування великої кількості факторів і варіабельності. Автоматизовані системи системи зазвичай прямолінійні та не враховують багатьох факторів впливу.

Також варто відмітити що авизація розкладів забезпечує більшу справедливість та задоволення працівників. Системи можуть враховувати індивідуальні побажання персоналу, їхні кваліфікації та доступність, забезпечуючи рівномірний розподіл навантаження.

Це підвищує мотивацію працівників і знижує ризик вигорання, оскільки робота розподіляється більш справедливо.

Збалансований розклад, що враховує як потреби працівників, так і вимоги закладу, сприяє підвищенню продуктивності та ефективності роботи. Автоматизовані системи можуть адаптуватися до змін у попиті та оперативно коригувати розклади відповідно до поточних потреб закладу.

Основною проблемою у великих організаціях – є те що робітників багато і у сіх робітників відрізняються бажання що до особистого розкладу, це характерно для закладів харчування: ресторанів, кафе і так далі., отже стає проблема того – скільки людей та якої кваліфікації потрібно для вирішення поставлених закладом або ситуацією задач.

Саме тому необхідно розробити алгоритм який дозволить вирішити усі поставлені задачі та врахувати багато факторів впливу, а також бути гнучким до змін.

Аналіз комерційних пропозицій є важливим, але водночас складним завданням, яке має низку проблем і викликів. Ось декілька ключових аспектів, які роблять аналіз комерційних пропозицій проблематичним і актуальним:

По-перше, велика кількість даних. Сучасні компанії отримують безліч комерційних пропозицій, які містять великі обсяги інформації. Це можуть бути деталі про продукцію, умови співпраці, ціни, гарантії тощо. Обробка і аналіз такої кількості даних вручну є надзвичайно трудомістким і часозатратним процесом.

Комерційні пропозиції можуть бути представлені в різних форматах: текстові документи, електронні листи, PDF-файли, презентації тощо. Крім того,

різні компанії використовують різні стилі написання та структурування інформації, що ускладнює автоматизацію процесу аналізу.

Для того щоб прийняти обґрунтоване рішення, важливо визначити, які пропозиції є найбільш релевантними і точними щодо потреб конкретної компанії. Це вимагає врахування багатьох факторів, таких як специфіка ринку, індивідуальні потреби компанії, умови співпраці тощо.

Аналіз комерційних пропозицій часто залежить від людського фактору, що може призводити до суб'єктивних оцінок і помилок. Різні аналітики можуть по-різному інтерпретувати ту саму інформацію, що може вплинути на прийняття рішень.

Проведення ретельного аналізу комерційних пропозицій вимагає значних затрат часу і ресурсів, що може бути особливо проблематично для малих і середніх компаній, які не мають достатніх людських і фінансових ресурсів.

Актуальність проблеми аналізу комерційних пропозицій пояснюється кількома чинниками. По-перше, в умовах жорсткої конкуренції на ринку, компанії повинні приймати швидкі і точні рішення для забезпечення своєї конкурентоспроможності. По-друге, зростання обсягів інформації і її різноманітність вимагають використання сучасних технологій і методів для її обробки і аналізу. По-третє, автоматизація процесів, включаючи аналіз комерційних пропозицій, дозволяє зменшити затрати часу і ресурсів, підвищити точність і ефективність прийняття рішень.

Використання технологій машинного навчання, таких як ML.NET і TensorFlow, стає дедалі більш актуальним у цьому контексті. Ці технології дозволяють автоматизувати процеси аналізу, забезпечуючи швидке і точне визначення релевантних і актуальних комерційних пропозицій. Машинне навчання може допомогти виявити приховані закономірності і взаємозв'язки в даних, що важко зробити вручну. Крім того, алгоритми машинного навчання можуть навчатися на великих обсягах даних, постійно вдосконалюючи свої результати і адаптуючись до змін у ринкових умовах.

Таким чином, проблема аналізу комерційних пропозицій є складною, але водночас надзвичайно актуальною. Вона вимагає використання сучасних технологій і підходів для забезпечення ефективності і точності процесу прийняття рішень у бізнесі.

### 1.3 Постановка задачі

У даній роботі ми дослідимо методи визначення актуальності та релевантності запитань до комерційних пропозицій. Наше дослідження буде зосереджено на предметній області машинного навчання та його застосуванні для аналізу комерційних пропозицій, з метою покращення процесу прийняття рішень у бізнесі. Ми розглянемо основні проблеми, пов'язані з аналізом комерційних пропозицій, та методи їх вирішення за допомогою сучасних технологій. Проведемо порівняльний аналіз підходів на основі ML.NET та TensorFlow, порівняємо їх ефективність та адаптованість до наших задач.

Робота спрямована на виявлення найкращого методу або алгоритму, який забезпечить точне і швидке визначення актуальності та релевантності запитань до комерційних пропозицій, що дозволить компаніям оптимізувати свої бізнес-процеси та підвищити конкурентоспроможність.

Реалізація наукового дослідження складається з наступних етапів:

- аналіз предметної області комерційних пропозицій та проблем, пов'язаних з їх аналізом;
- аналіз архітектурних особливостей алгоритмів машинного навчання для визначення актуальності та релевантності запитань;
- розгляд основних характеристик та функціональності алгоритмів ML.NET та TensorFlow;
- проведення порівняльного аналізу підходів на основі ML.NET та TensorFlow, оцінити їхню ефективність та адаптованість до поставлених задач;

- розробка алгоритму для автоматизації процесу аналізу комерційних пропозицій, який враховує численні фактори впливу та є гнучким до змін ринкових умов;
- реалізація розробленого алгоритму у бізнес-середовищі та провести його тестування;
- оцінка результатів впровадження з точки зору підвищення ефективності бізнес-процесів та задоволеності клієнтів.

## 2 АНАЛІЗ ПЕРСПЕКТИВНИХ МЕТОДІВ ТА АЛГОРИТМІВ

### 2.1 Машинне навчання

Машинне навчання – це підгалузь штучного інтелекту, яка фокусується на розробці алгоритмів та моделей, здатних навчатися на основі даних і приймати рішення або робити прогнози без прямого програмування. Це означає, що системи, побудовані на основі машинного навчання, можуть автоматично вдосконалюватися з досвідом. Основна мета машинного навчання полягає в тому, щоб дозволити комп'ютерам навчатися з даних, виявляти в них закономірності та використовувати ці знання для виконання різноманітних завдань.

Машинне навчання поділяється на кілька типів залежно від характеру завдань та методів навчання. Найпоширенішими типами є навчання з учителем, навчання без учителя та навчання з підкріпленням. Навчання з учителем включає процес, коли модель навчається на наборі даних, що містить вхідні приклади та відповідні мітки. Метою є навчити модель робити правильні прогнози для нових, невідомих даних. Прикладами таких завдань є класифікація та регресія. Навчання без учителя, з іншого боку, не має міток у навчальному наборі даних. Модель намагається знайти приховані структури або закономірності у даних. Кластеризація та асоціація є основними завданнями цього типу навчання. Навчання з підкріпленням залучає агента, який навчається взаємодіяти з навколишнім середовищем, отримуючи винагороди або покарання залежно від своїх дій. Мета агента полягає у максимізації довгострокової винагороди.

Основними етапами побудови моделі машинного навчання є збір даних, їх підготовка, вибір моделі, навчання моделі, оцінка її продуктивності та вдосконалення. Збір даних є початковим і надзвичайно важливим етапом, оскільки якість моделі сильно залежить від якості даних. Підготовка даних включає очищення, нормалізацію, перетворення та розбиття даних на навчальний і тестовий набори. Вибір моделі залежить від характеру завдання і включає вибір алгоритму, який буде використаний для навчання. Навчання моделі полягає в налаштуванні параметрів алгоритму для мінімізації помилки на навчальному наборі даних. Оцінка продуктивності моделі виконується за допомогою тестового

набору даних, щоб перевірити її здатність узагальнювати знання на нові дані. Вдосконалення моделі може включати додаткове налаштування параметрів, збір нових даних або використання більш складних алгоритмів.

Застосування машинного навчання є надзвичайно широким і охоплює різні галузі, включаючи охорону здоров'я, фінанси, маркетинг, виробництво та багато інших. У сфері охорони здоров'я, наприклад, машинне навчання використовується для діагностики захворювань, прогнозування результатів лікування та персоналізації медичних послуг. У фінансах воно застосовується для оцінки кредитоспроможності, виявлення шахрайства та управління ризиками. У маркетингу машинне навчання допомагає аналізувати поведінку клієнтів, передбачати продажі та розробляти персоналізовані рекомендації.

Машинне навчання є фундаментальною технологією для багатьох сучасних інновацій і продовжує розвиватися з швидким темпом. З розвитком обчислювальних потужностей та накопиченням великих обсягів даних, можливості машинного навчання стають все більш потужними, дозволяючи вирішувати все складніші завдання і відкриваючи нові горизонти для досліджень та практичних застосувань.

На рисунку 2.1 наведено схему роботи моделі машинного навчання.

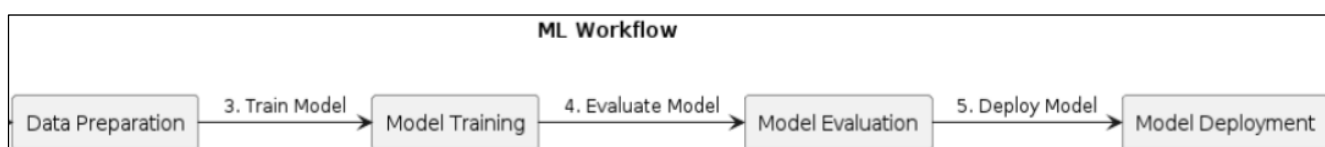


Рисунок 2.1 – Приклад роботи механізму машинного навчання

Машинне навчання на платформі .NET набуває все більшої популярності завдяки інтеграції з існуючими корпоративними рішеннями та можливостям, які надає ця платформа для розробників. Однією з основних технологій, що підтримують машинне навчання на .NET, є ML.NET – кросплатформова бібліотека з відкритим вихідним кодом, розроблена Microsoft. ML.NET дозволяє .NET розробникам створювати, тренувати та розгортати моделі машинного

навчання без необхідності вивчення Python чи інших мов програмування, традиційно пов'язаних з машинним навчанням.

ML.NET підтримує широкий спектр завдань машинного навчання, включаючи класифікацію, регресію, кластеризацію, рекомендаційні системи та інші. Розробники можуть використовувати ML.NET для вирішення практичних задач, таких як прогнозування продажів, аналіз настроїв, розпізнавання зображень і виявлення аномалій. Бібліотека надає простий і інтуїтивно зрозумілий API, що дозволяє легко інтегрувати машинне навчання в існуючі .NET додатки.

Однією з основних переваг ML.NET є його інтеграція з іншими технологіями та інструментами .NET екосистеми. Розробники можуть використовувати знайомі їм мови програмування, такі як C# або F#, а також інтегрувати машинне навчання з ASP.NET для створення веб-додатків або з Xamarin для розробки мобільних додатків. Крім того, ML.NET підтримує роботу з популярними форматами даних, такими як текстові файли, SQL-бази даних, файли Excel та інші.

ML.NET включає набір попередньо навчених моделей і можливості для автоматизованого машинного навчання (AutoML), що дозволяє автоматично знаходити оптимальні моделі та гіперпараметри для конкретних задач. Це значно спрощує процес розробки моделей машинного навчання для новачків, а також підвищує ефективність роботи досвідчених розробників.

Крім ML.NET, .NET платформа також підтримує інтеграцію з іншими популярними бібліотеками та фреймворками машинного навчання, такими як TensorFlow і ONNX. Це дозволяє розробникам використовувати потужні можливості TensorFlow для створення та розгортання моделей глибокого навчання або імпортувати вже існуючі моделі, створені на інших платформах, у .NET середовище.

Одним із прикладів успішного використання машинного навчання на платформі .NET є створення рекомендаційних систем. Використовуючи ML.NET, розробники можуть створювати моделі, які аналізують історичні дані про поведінку користувачів і пропонують персоналізовані рекомендації для кожного

користувача. Це дозволяє підвищити рівень залучення та задоволеності користувачів, а також збільшити продажі.

Ще одним прикладом є аналіз настроїв у текстових повідомленнях. Використовуючи можливості ML.NET для обробки природної мови (NLP), розробники можуть створювати моделі, які аналізують відгуки клієнтів, соціальні мережі або інші текстові дані, щоб визначити загальний настрій або емоційний тон. Це допомагає компаніям краще розуміти своїх клієнтів та відповідно коригувати свої стратегії.

Таким чином, машинне навчання на платформі .NET, зокрема з використанням ML.NET, надає потужні інструменти та можливості для розробки сучасних інтелектуальних додатків. Це дозволяє компаніям інтегрувати машинне навчання у свої існуючі рішення, покращуючи процеси прийняття рішень, автоматизуючи рутинні завдання та підвищуючи ефективність бізнесу.

## 2.2 Методи на основі правил

Методи на основі правила являють собою підхід до аналізу та обробки інформації, де заздалегідь визначені правила використовуються для класифікації, фільтрації та обробки даних. Ці методи зазвичай включають набір умов та дій, які виконуються при виконанні певних умов. У контексті аналізу комерційних пропозицій, методи на основі правила можуть бути використані для автоматизації процесу оцінки та визначення релевантності та актуальності запитань та відповідей.

При застосуванні методів на основі правила до аналізу комерційних пропозицій, основна увага приділяється створенню набору правил, які визначають, які запитання чи пропозиції є важливими або релевантними для конкретних бізнес-завдань. Ці правила можуть бути побудовані на основі попереднього досвіду, знань експертів або аналізу історичних даних.

Наприклад, правила можуть бути сформульовані для виявлення ключових слів або фраз, які часто зустрічаються у важливих комерційних пропозиціях. Кожне правило може визначати, що якщо у тексті зустрічається певне ключове

слово або фраза, то ця пропозиція вважається релевантною. Крім того, правила можуть враховувати контекст, наприклад, визначати, що комбінація певних ключових слів у певному контексті підвищує важливість пропозиції.

Застосування методів на основі правила до аналізу комерційних пропозицій може включати кілька етапів:

- збір та підготовка даних: спочатку збираються комерційні пропозиції та запитання, які потрібно аналізувати. дані очищуються та перетворюються у зручний для аналізу формат;
- визначення правил: на основі попереднього досвіду, знань експертів або аналізу історичних даних визначаються правила, які будуть використовуватися для аналізу. наприклад, правило може бути таким: "якщо пропозиція містить слово 'знижка', вона вважається важливою";
- застосування правил: визначені правила застосовуються до зібраних даних. це може бути зроблено за допомогою спеціалізованих програмних засобів або вручну. кожне правило перевіряється для кожної пропозиції, і визначається, чи виконується воно;
- аналіз результатів: на основі результатів застосування правил проводиться аналіз. визначаються релевантні та важливі пропозиції, які відповідають встановленим правилам;
- оновлення та оптимізація правил: на основі результатів аналізу правила можуть бути оновлені або оптимізовані для підвищення точності та ефективності аналізу. це може включати додавання нових правил, видалення застарілих або менш ефективних правил, а також налаштування існуючих правил.

Методи на основі правила мають кілька переваг при застосуванні до аналізу комерційних пропозицій. По-перше, вони є відносно простими для розуміння та реалізації. Правила можуть бути легко сформульовані та застосовані, що дозволяє швидко отримувати результати. По-друге, ці методи можуть бути легко адаптовані під конкретні потреби та вимоги бізнесу. Нові правила можуть бути додані у відповідь на зміну умов або вимог.

Однак, методи на основі правила мають і свої обмеження. Вони можуть бути менш ефективними при обробці великих обсягів даних або складних задачах, що вимагають більш глибокого аналізу. Крім того, ці методи вимагають постійного оновлення та оптимізації правил для забезпечення їх актуальності та точності. У випадках, коли потрібно обробляти великі обсяги даних або вирішувати складні задачі, доцільно використовувати ці методи в поєднанні з іншими, більш потужними техніками машинного навчання та обробки природної мови(див. рис. 2.2).

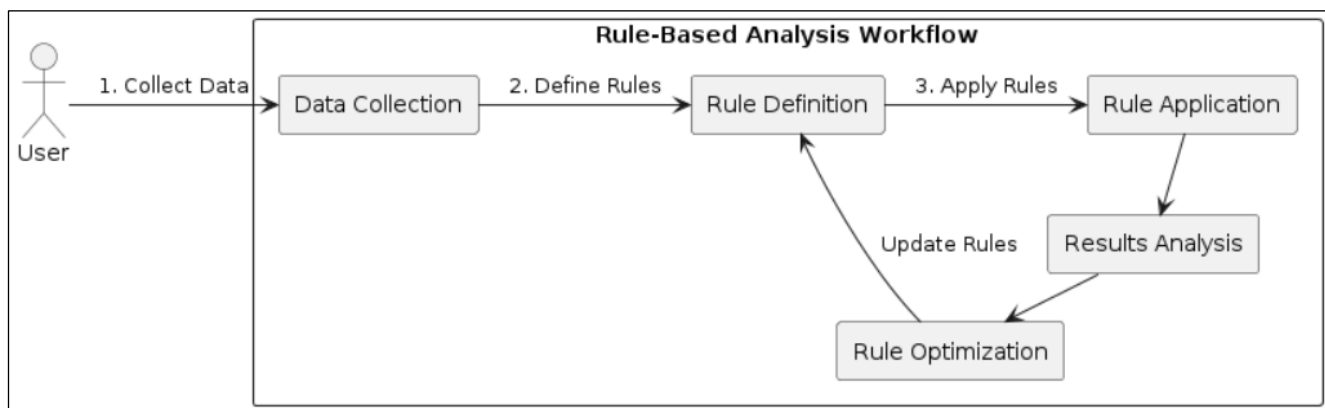


Рисунок 2.2 – Схема роботи моделі правил

Розглянемо кожен етап:

- Data Collection (Збір даних): Користувач або система збирає комерційні пропозиції та запитання, які потрібно аналізувати;
- Rule Definition (Визначення правил): Визначаються правила для аналізу на основі попереднього досвіду, знань експертів або аналізу історичних даних;
- Rule Application (Застосування правил): Визначені правила застосовуються до зібраних даних для визначення релевантності та актуальності запитань і пропозицій;
- Results Analysis (Аналіз результатів): На основі результатів застосування правил проводиться аналіз, визначаються важливі пропозиції;

- Rule Optimization (Оновлення та оптимізація правил): На основі результатів аналізу правила оновлюються або оптимізуються для підвищення точності та ефективності аналізу.

Ця схема ілюструє загальний робочий процес аналізу комерційних пропозицій за допомогою методу на основі правил.

### 2.3 Штучні нейронні мережі

Штучні нейронні мережі (ШНМ) є моделями, натхненними біологічними нейронними мережами в мозку, що використовуються для розпізнавання складних закономірностей і структур у великих обсягах даних. Вони складаються з великої кількості взаємопов'язаних вузлів (нейронів), організованих у шари. Основні компоненти ШНМ включають вхідний шар, один або більше прихованих шарів і вихідний шар. Кожен нейрон приймає один або більше входів, обробляє їх і передає вихід іншим нейронам у наступному шарі.

Однією з основних переваг ШНМ є їх здатність навчатися на основі даних. Це дозволяє їм автоматично знаходити і виявляти складні закономірності та взаємозв'язки, які важко виявити за допомогою традиційних методів. ШНМ використовуються для різних задач, таких як класифікація, регресія, кластеризація, обробка зображень і природної мови.

Застосування ШНМ до аналізу комерційних пропозицій може значно покращити точність і ефективність цього процесу. Ось як це можна реалізувати:

Перед застосуванням ШНМ до аналізу комерційних пропозицій, необхідно виконати попередню обробку даних. Це включає очищення даних, видалення зайвих символів, нормалізацію тексту (перетворення всіх символів у нижній регістр, видалення знаків пунктуації), а також токенізацію (розбиття тексту на окремі слова чи фрази).

Оскільки ШНМ працюють з числовими даними, текстові пропозиції потрібно перетворити у числові вектори. Для цього можна використовувати методи векторизації, такі як Bag of Words, TF-IDF (Term Frequency-Inverse

Document Frequency) або більш сучасні підходи, такі як Word Embeddings (Word2Vec, GloVe) чи трансформери (BERT).

Наступний крок – створення архітектури нейронної мережі, яка буде використовуватися для аналізу пропозицій. Зазвичай, для роботи з текстовими даними використовуються рекурентні нейронні мережі (RNN) або їх вдосконалені версії, такі як LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Units). Для обробки довгих текстів і складних залежностей можна застосовувати трансформери, такі як BERT або GPT.

Модель ШНМ потрібно навчити на великій кількості комерційних пропозицій, маркованих відповідними мітками (релевантні або нерелевантні). Процес навчання включає налаштування ваг нейронів для мінімізації похибки між передбаченнями моделі і фактичними мітками.

Після навчання модель потрібно оцінити на тестовому наборі даних для визначення її продуктивності. Для цього використовуються метрики, такі як точність, повнота, F-міра та інші. Якщо модель показує недостатню продуктивність, можна виконати додаткову оптимізацію, змінюючи архітектуру мережі, гіперпараметри або обробку даних.

Після навчання і оцінки модель готова до застосування для аналізу нових комерційних пропозицій. Вона автоматично визначає релевантність і актуальність кожної пропозиції на основі навченої інформації.

Переваги використання ШНМ для аналізу комерційних пропозицій:

- автоматизація: шнм можуть автоматично обробляти і аналізувати великі обсяги комерційних пропозицій, зменшуючи потребу в ручній роботі;
- точність: завдяки здатності виявляти складні закономірності в даних, шнм можуть забезпечувати високу точність аналізу;
- гнучкість: шнм можна адаптувати для різних типів задач і даних, роблячи їх універсальним інструментом для аналізу.

Недоліки:

- обчислювальні ресурси: навчання шнм вимагає значних обчислювальних ресурсів і часу;

- потреба у великих обсягах даних: для ефективного навчання ШНМ потрібні великі набори маркованих даних;
- складність реалізації: створення і налаштування ШНМ може бути складним процесом, що вимагає глибоких знань і досвіду.

Таким чином, штучні нейронні мережі є потужним інструментом для аналізу комерційних пропозицій, що дозволяє автоматизувати процеси, підвищити точність і ефективність аналізу, та забезпечити гнучкість у вирішенні різних завдань. Вони відкривають нові можливості для бізнесу, дозволяючи краще розуміти та реагувати на потреби клієнтів і ринку.

## 3 ПОРІВНЯННЯ РІШЕНЬ ТА ЇХ МОДИФІКАЦІЯ

### 3.1 Вибір методу аналізу

Машинне навчання було обрано для аналізу комерційних пропозицій з кількох вагомих причин, які виділяють його серед інших методів. Це пояснюється його гнучкістю, здатністю до автоматизації, високою точністю та ефективністю в обробці великих обсягів даних. Нижче наведено основні причини, чому машинне навчання є оптимальним вибором для цього завдання.

Машинне навчання дозволяє створювати моделі, які можуть адаптуватися до різних типів даних і задач. Це особливо важливо в контексті комерційних пропозицій, де дані можуть бути різномірними та змінюватися з часом. Моделі машинного навчання можуть бути навчені на великому наборі даних, щоб враховувати різні контексти і специфіку комерційних пропозицій, що робить їх більш універсальними порівняно з методами на основі правил, які потребують постійного оновлення і налаштування.

Машинне навчання дозволяє автоматизувати процес аналізу комерційних пропозицій, зменшуючи потребу в ручній обробці даних. Це значно підвищує ефективність та зменшує витрати часу і ресурсів. Автоматизовані системи машинного навчання можуть швидко і точно аналізувати великі обсяги даних, виявляти релевантні пропозиції та надавати рекомендації, що забезпечує конкурентні переваги для бізнесу.

Машинне навчання відоме своєю здатністю до високої точності і продуктивності в розпізнаванні складних закономірностей у даних. Завдяки алгоритмам навчання, такі як нейронні мережі, методи опорних векторів (SVM), дерева рішень та інші, моделі машинного навчання можуть виявляти тонкі і складні взаємозв'язки в даних, що робить їх особливо корисними для аналізу комерційних пропозицій, де часто потрібні точні і надійні прогнози.

Однією з ключових переваг машинного навчання є здатність моделей до самовдосконалення на основі нових даних. Це означає, що моделі можуть постійно оновлюватися і покращувати свою точність і ефективність, використовуючи нові набори даних і отримуючи новий досвід. Це робить

машинне навчання більш динамічним і адаптивним до змін у бізнес-середовищі та ринкових умовах.

Підхід здатний ефективно обробляти великі обсяги даних, що є критично важливим для аналізу комерційних пропозицій, які можуть включати в себе великі текстові масиви, зображення та інші типи даних. Завдяки цьому, машинне навчання може забезпечити швидкий та точний аналіз навіть у випадках, коли обсяги даних значно перевищують можливості традиційних методів обробки.

Машинне навчання включає в себе широкий спектр передових алгоритмів та технологій, таких як глибоке навчання, обробка природної мови (NLP), векторизація слів (Word Embeddings) та трансформери. Це дозволяє здійснювати більш точний і глибокий аналіз текстових даних, які є основою комерційних пропозицій, забезпечуючи високу точність і надійність результатів.

Цей метод може бути використане для вирішення різних типів задач, включаючи класифікацію, регресію, кластеризацію, рекомендаційні системи та багато іншого. Це робить його універсальним інструментом для аналізу комерційних пропозицій, дозволяючи вирішувати широкий спектр завдань від визначення релевантності до прогнозування успішності пропозицій.

Машинне навчання пропонує потужний набір інструментів і методів, які можуть значно підвищити ефективність і точність аналізу комерційних пропозицій. Завдяки своїй гнучкості, здатності до автоматизації, високій точності, можливості обробки великих обсягів даних та використанню передових алгоритмів, машинне навчання є оптимальним вибором для цієї задачі. Це дозволяє бізнесу краще розуміти та реагувати на потреби ринку, підвищувати свою конкурентоспроможність та приймати обґрунтовані рішення на основі даних.

### 3.2 Порівняння бібліотек машинного навчання

Для машинного навчання на платформі .NET існує кілька потужних бібліотек, які можуть бути використані для різних завдань. Ось три основні

бібліотеки: ML.NET, Accord.NET, і TensorFlow.NET. Давайте розглянемо кожну з них детальніше.

ML.NET є кросплатформовою бібліотекою з відкритим вихідним кодом, розробленою Microsoft. Вона призначена для створення, тренування та розгортання моделей машинного навчання в .NET додатках.

Основні особливості:

- широкий спектр завдань: підтримує класифікацію, регресію, кластеризацію, рекомендаційні системи, обробку тексту та зображень;
- automl: автоматизоване машинне навчання для автоматичного вибору найкращих моделей і гіперпараметрів;
- інтеграція: легко інтегрується з іншими .net технологіями, такими як asp.net, xamarin, azure;
- підтримка onnx: можна використовувати моделі, збережені у форматі onnx, для інтероперабельності з іншими платформами машинного навчання.

Переваги:

- інтуїтивно зрозумілий арі для .net розробників;
- сильна інтеграція з екосистемою microsoft;
- підтримка автоматизованого машинного навчання (automl).

Недоліки:

- можливості можуть бути обмеженими порівняно з іншими спеціалізованими бібліотеками для глибокого навчання.

Розглянемо Accord.NET.

Accord.NET – це комплексна .NET бібліотека для статистичного аналізу, машинного навчання та обробки зображень. Вона є частиною проекту Accord Framework, що містить безліч алгоритмів і інструментів.

Основні особливості:

- широкий набір алгоритмів: включає алгоритми для класифікації, регресії, кластеризації, обробки сигналів та зображень;

- статистичний аналіз: підтримує широкий спектр статистичних методів і тестів;
- обробка зображень: містить функції для аналізу і обробки зображень;
- інтеграція: може бути використана з іншими .net бібліотеками і додатками.

#### Переваги:

- велика кількість алгоритмів і функцій в одній бібліотеці;
- підтримка як класичних алгоритмів машинного навчання, так і методів обробки сигналів та зображень;
- можливість виконання статистичних тестів і аналізу.

#### Недоліки:

- менш активний розвиток порівняно з ml.net;
- може бути складніше у використанні через велику кількість функцій і параметрів.

Далі розглянемо TensorFlow.NET.

TensorFlow.NET – це .NET обгортка для популярної бібліотеки машинного навчання TensorFlow. Вона дозволяє .NET розробникам використовувати TensorFlow для створення моделей глибокого навчання.

#### Основні особливості:

- глибоке навчання: підтримує створення і навчання нейронних мереж різної складності;
- сумісність з tensorflow: дозволяє використовувати існуючі моделі та інструменти tensorflow;
- кросплатформеність: працює на різних операційних системах завдяки платформі .net core;
- гнучкість: надає доступ до всіх функцій tensorflow через .net api.

#### Переваги:

- можливість використання потужних інструментів і моделей tensorflow в .net середовищі;

- підтримка сучасних алгоритмів глибокого навчання;
- активна спільнота і велика кількість ресурсів для навчання та підтримки.

Недоліки:

- може бути складним для новачків через необхідність знань tensorflow;
- більше підходить для глибокого навчання, ніж для класичних алгоритмів машинного навчання.

Кожна з цих бібліотек має свої унікальні особливості та підходить для різних типів завдань. ML.NET є чудовим вибором для розробників, які бажають інтегрувати машинне навчання в .NET додатки з мінімальними зусиллями. Accord.NET надає великий набір інструментів для статистичного аналізу та обробки зображень, тоді як TensorFlow.NET дозволяє використовувати потужні можливості TensorFlow для глибокого навчання в .NET середовищі. Вибір бібліотеки залежить від конкретних потреб проекту та рівня підготовки розробників.

### 3.3 Вибір ключового функціоналу BERT

BERT (Bidirectional Encoder Representations from Transformers) – це модель для обробки природної мови, розроблена Google, яка здійснила революцію в галузі машинного навчання завдяки своїй здатності розуміти контекст тексту з обох боків (вперед і назад). Це дозволяє BERT ефективно розпізнавати та інтерпретувати значення слів у контексті, що є критично важливим для багатьох завдань обробки природної мови (NLP), таких як класифікація тексту, виявлення сутностей, відповідь на запитання та багато інших.

BERT використовує архітектуру трансформера, яка складається з кількох шарів енкодерів. Ключовою особливістю BERT є його двонаправленість: він одночасно враховує контекст слів як зліва, так і справа від цільового слова. Це дозволяє моделі краще розуміти контекст і значення слів у реченнях.

Модель BERT спочатку передтренується на великих корпусах тексту, таких як Wikipedia та книги, з використанням двох завдань: маскуванню слів (masked language model) і передбачення наступного речення (next sentence prediction).

Після попереднього тренування модель можна тонко налаштувати на конкретні завдання, надаючи їй додаткові дані та метрики.

Для аналізу комерційних пропозицій BERT може бути використаний для визначення, який об'єкт комерції запитано у запиті. Це завдання включає розпізнавання сутностей і класифікацію тексту, щоб визначити конкретні продукти, послуги або інші об'єкти, про які йдеться в комерційних запитах.

Спочатку необхідно зібрати великий набір даних, що містять комерційні пропозиції та запити, марковані відповідними об'єктами комерції. Ці дані повинні бути очищені та відформатовані для подальшого аналізу. Підготовка даних також включає токенізацію тексту за допомогою токенізатора BERT, який розбиває текст на окремі слова або субслова.

Попередньо натреновану модель BERT потрібно тонко налаштувати на нашому конкретному наборі даних комерційних пропозицій. Це включає навчання моделі на маркованих даних, щоб вона могла розпізнавати і класифікувати об'єкти комерції у запитах. Цей процес вимагає використання потужних обчислювальних ресурсів, таких як графічні процесори (GPU) або спеціалізовані апаратні прискорювачі (TPU).

Після навчання модель BERT може використовуватися для аналізу нових комерційних запитів. Вона буде розпізнавати сутності та класифікувати їх відповідно до категорій, таких як продукти, послуги або інші об'єкти. Це дозволяє автоматично визначати, про які конкретні комерційні об'єкти йдеться у запитах, що значно підвищує ефективність обробки запитів і формування відповідей.

Інтеграція моделі BERT з існуючими бізнес-процесами дозволяє автоматизувати аналіз комерційних запитів і пропозицій. Це може бути реалізовано через API або інші програмні інтерфейси, що забезпечують доступ до можливостей моделі для різних додатків, таких як CRM-системи, системи управління замовленнями, служби підтримки клієнтів тощо.

Завдяки своїй здатності враховувати контекст з обох сторін, BERT забезпечує високу точність в аналізі тексту та розпізнаванні сутностей. Це

особливо важливо для комерційних пропозицій, де навіть невелика неточність може вплинути на прийняття рішень.

BERT можна адаптувати до різних задач обробки природної мови, що робить його універсальним інструментом для аналізу текстових даних. Це дозволяє використовувати одну й ту ж модель для різних аспектів аналізу комерційних пропозицій.

Модель BERT може бути легко оновлена новими даними, що дозволяє їй постійно вдосконалювати свої результати і адаптуватися до змін у бізнес-середовищі та ринкових умовах.

Недоліки:

- вимоги до обчислювальних ресурсів;
- навчання та використання bert вимагає значних обчислювальних ресурсів, що може бути дорогим і складним для реалізації в деяких середовищах.

Складність налаштування:

- процес тонкого налаштування і інтеграції моделі bert може бути складним і вимагати глибоких знань у сфері машинного навчання і обробки природної мови.

BERT є потужним інструментом для аналізу комерційних пропозицій завдяки своїй здатності розуміти контекст тексту і високо точно розпізнавати сутності. Використання BERT для визначення об'єктів комерції у запитах дозволяє автоматизувати і вдосконалити процес обробки комерційних пропозицій, що сприяє підвищенню ефективності бізнес-процесів та прийняття обґрунтованих рішень. Враховуючи його переваги, BERT є перспективним інструментом для застосування в різних галузях, де аналіз текстових даних є критично важливим.

Етапи роботи підходу зображено на рисунку 3.2.

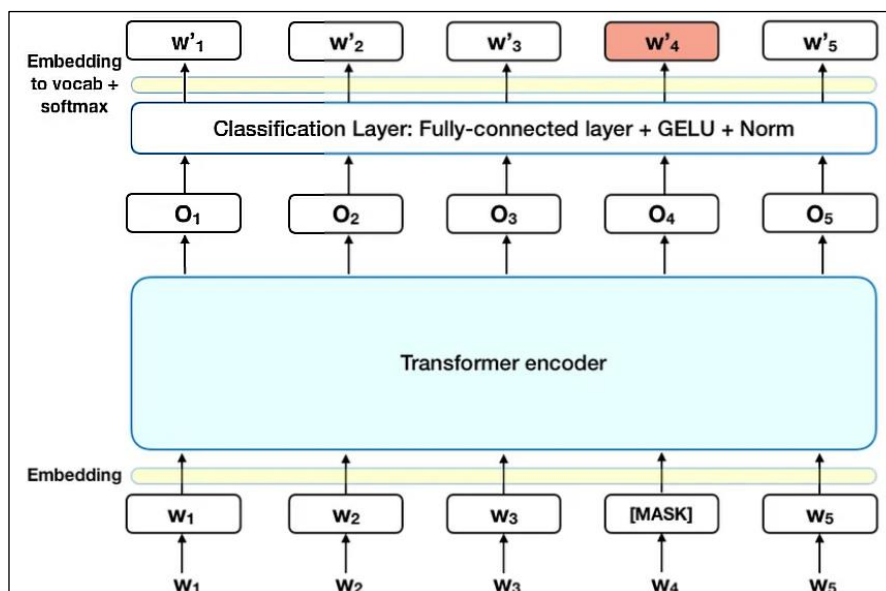


Рисунок 3.2 – Схематичний вигляд роботи BERT

Розглянемо ці компоненти:

- embedding layer – перетворює вхідні слова у числові вектори, які кодують значення кожного слова для подальшої обробки;
- transformer encoder – основний компонент моделі bert, який обробляє ці вектори, враховуючи контекст кожного слова з обох сторін. використовує механізм самовідношення для врахування впливу кожного слова на інші слова в реченні;
- $o_1, o_2, o_3, o_4, o_5$ : виходи трансформера, що містять контекстну інформацію про кожне слово;
- виконує завдання класифікації або передбачення. виходи з трансформера проходять через повнозв'язний шар з функцією gelu і нормалізацією;
- output layer – перетворює результати класифікаційного шару у ймовірнісні оцінки за допомогою функції softmax, визначаючи категорії слів;
- masked language modeling (mlm) – маскує випадкові слова у реченні та навчає модель передбачати ці слова на основі контексту, допомагаючи моделі краще розуміти текст.

## 4 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТУ

### 4.1 Обрані підходи

Для порівняння ми обрали ML.NET та TensorFlow з імплементацією BERT у кожен підхід з кількох вагомих причин. Ці платформи мають різні сильні сторони та особливості, що робить їх відповідними для різних сценаріїв застосування в аналізі комерційних пропозицій.

ML.NET підтримує широкий спектр завдань машинного навчання, таких як класифікація, регресія, кластеризація, рекомендаційні системи тощо. Це робить його універсальним інструментом для різних видів аналізу комерційних пропозицій, дозволяючи адаптуватися до різних вимог і сценаріїв.

Хоча ML.NET не має рідної підтримки для трансформерів, таких як BERT, його можна інтегрувати з ONNX (Open Neural Network Exchange), що дозволяє використовувати передтреновані моделі BERT. Це забезпечує можливість застосування передових методів обробки природної мови в екосистемі .NET.

TensorFlow є однією з найпотужніших платформ для глибокого навчання, розробленою Google. Він підтримує складні архітектури нейронних мереж, такі як трансформери, і забезпечує високу продуктивність при обробці великих обсягів даних.

TensorFlow має рідну підтримку для трансформерів, включаючи BERT. Це дозволяє легко імплементувати передтреновані моделі BERT і налаштувати їх для конкретних завдань, забезпечуючи високу точність і ефективність аналізу текстових даних.

### 4.2 Архітектура системи

Далі на рисунку 4.2 наведемо приклад тестової архітектури нашого додатку та опишемо зв'язок елементів.

Наше рішення має просту тестову файлова структуру:

- ConsoleApp (Кореневий проект);
- Dependencies (Залежності);

- Program.cs (Головний файл);
- ReviewData (Клас даних відгуків);
- ReviewTestData (Клас тестових даних відгуків);
- ReviewPrediction (Клас передбачення);
- Program (Головний клас з методами для тренування та тестування моделі).

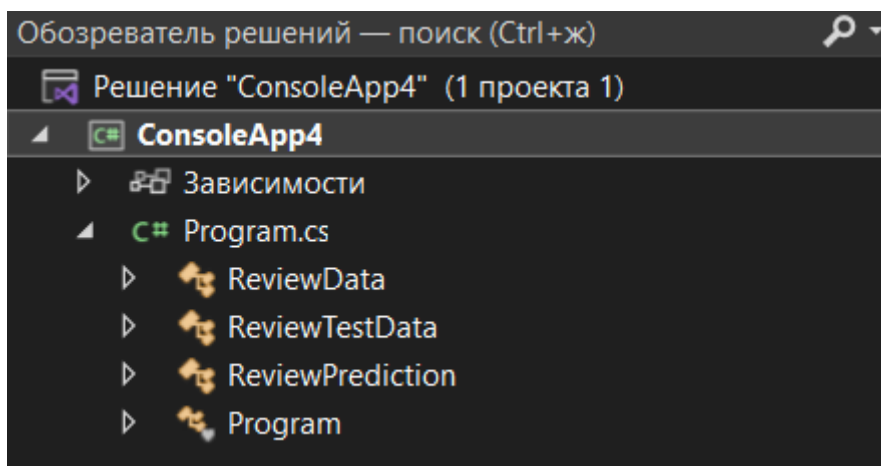


Рисунок 4.2 – Тестова архітектура додатку

Ця архітектура є досить стандартною для консольних програм, які використовують ML.NET для машинного навчання. Вона включає в себе класи для роботи з даними, методи для тренування моделі та методи для тестування моделі.

Далі варто навести головний код алгоритму нашого застосунку та пояснити його роботу:

```
using Microsoft.ML;
using Microsoft.ML.Data;
using Microsoft.ML.Transforms.Onnx;
using System;
using System.Collections.Generic;
using System.IO;
using System.Linq;

public class ReviewData
{
    [LoadColumn(0)]
    public string Text { get; set; } // Текст відгука
    [LoadColumn(1)]
    public string Category { get; set; } // Категорія відгука
}
```

```

}

public class ReviewTestData
{
    [LoadColumn(0)]
    public string Text { get; set; } // Текст відгука
}

public class ReviewPrediction
{
    [ColumnName("PredictedLabel")]
    public string PredictedCategory { get; set; } // Передбачена категорія
}

class Program
{
    static void Main(string[] args)
    {
        var mlContext = new MLContext(seed: 0);

        Console.WriteLine("Choose an option:");
        Console.WriteLine("1: Train Model");
        Console.WriteLine("2: Test Model");
        Console.WriteLine("3: Exit");
        Console.Write("Select an option: ");

        var option = Console.ReadLine();

        switch (option)
        {
            case "1":
                TrainAndSaveModel(mlContext);
                break;
            case "2":
                TestExistingModel(mlContext);
                break;
            case "3":
                Environment.Exit(0);
                break;
            default:
                Console.WriteLine("Invalid option, try again.");
                break;
        }
    }

    static void TrainAndSaveModel(MLContext mlContext)
    {
        var dataPath = @"E:\Users\bogac\Desktop\ComercialDataSet.txt";
        var modelPath = @"E:\Users\bogac\Desktop\ComercialDataSet.zip";
        var dataView =
mlContext.Data.LoadFromTextFile<ReviewData>(dataPath, hasHeader: true,
separatorChar: ',');

        var pipeline =
mlContext.Transforms.Conversion.MapValueToKey(outputColumnName: "Label",
inputColumnName: nameof(ReviewData.Category))

```

```

.Append(mlContext.Transforms.Text.FeaturizeText(outputColumnName:
"Features", inputColumnName: nameof(ReviewData.Text)))

.Append(mlContext.MulticlassClassification.Trainers.SdcaMaximumEntropy())

.Append(mlContext.Transforms.Conversion.MapKeyToValue("PredictedCategory",
"PredictedLabel"));

    var model = pipeline.Fit(dataView);
    mlContext.Model.Save(model, dataView.Schema, modelPath);
    Console.WriteLine($"Model trained and saved at {modelPath}");
}

static void TestExistingModel(MLContext mlContext)
{
    var modelPath = @"E:\Users\bogac\Desktop\ComercialDataSet.zip";
    var testPath = @"E:\Users\bogac\Desktop\TestDataSet.txt";
    var testData =
mlContext.Data.LoadFromTextFile<ReviewTestData>(testPath, hasHeader: true,
separatorChar: ',');

    DataViewSchema modelSchema;
    var model = mlContext.Model.Load(modelPath, out modelSchema);
    var predictionEngine =
mlContext.Model.CreatePredictionEngine<ReviewTestData,
ReviewPrediction>(model);

    // Вивід кількох передбачень для тестових даних
    var testSamples =
mlContext.Data.CreateEnumerable<ReviewTestData>(testData, reuseRowObject:
false).Take(5);
    foreach (var sample in testSamples)
    {
        var prediction = predictionEngine.Predict(sample);
        Console.WriteLine($"Text: {sample.Text}");
        Console.WriteLine($"Predicted Category:
{prediction.PredictedCategory}");
        Console.WriteLine($"Commercial Object:
{GetCommercialObject(sample.Text)}");
        Console.WriteLine();
    }
}

// Метод для визначення комерційного об'єкта
static string GetCommercialObject(string text)
{
    // Приклад простого аналізу тексту для виділення об'єкта
    var commercialItems = new List<string> { "laptop", "coffee maker",
"shoes", "software subscription", "product quality" };
    foreach (var item in commercialItems)
    {
        if (text.Contains(item))
        {
            return item;
        }
    }
    return "object"; // Якщо об'єкт не знайдено
}

```

```
}  
}
```

Наш код є консольним додатком на .NET, який використовує бібліотеку ML.NET для тренування і тестування моделі класифікації текстових відгуків. У ньому є кілька важливих логічних блоків, які забезпечують функціональність додатка. Спочатку користувачеві пропонується вибрати одну з трьох опцій: тренувати модель, тестувати модель або завершити програму. Після вибору користувачем відповідної опції викликається один із методів: `TrainAndSaveModel` або `TestExistingModel`.

Метод `TrainAndSaveModel` відповідає за тренування моделі. Він завантажує дані з файлу, що містить текстові відгуки та відповідні категорії. Ці дані зберігаються у форматі `DataView`. Потім визначається конвеєр обробки даних, який включає перетворення текстових даних у числові вектори за допомогою функції `FeaturizeText`, а також перетворення значень категорій у ключі за допомогою `MapValueToKey`. Після цього модель тренується за допомогою алгоритму `SDCA Maximum Entropy`, який є одним з доступних тренерів для багатокласової класифікації. Після тренування модель зберігається у файл для подальшого використання.

Метод `TestExistingModel` призначений для тестування вже тренованої моделі. Він завантажує модель з файлу та тестові дані з іншого файлу, що містить текстові відгуки без вказання категорій. За допомогою `CreatePredictionEngine` створюється механізм передбачення, який використовується для отримання передбачень на основі тестових даних. Програма виводить текст відгуків, передбачені категорії та виявлені комерційні об'єкти, якщо такі є, для кількох тестових зразків.

Метод `GetCommercialObject` забезпечує простий аналіз тексту для виявлення комерційних об'єктів у відгуках. Він перевіряє, чи містить текст певні ключові слова, що представляють комерційні об'єкти, і якщо знаходить такі, то повертає відповідне слово, інакше повертає значення "object", що означає, що об'єкт не був знайдений.

Таким чином, цей код демонструє базовий підхід до тренування та тестування моделі машинного навчання для класифікації текстових відгуків з використанням ML.NET, а також показує, як можна інтегрувати додаткові функції аналізу тексту, такі як виявлення комерційних об'єктів.

### 4.3 Data-sets для тренування та тестування

Ми обрали такі дата сет для навчання та тестування, щоб забезпечити належну якість і релевантність даних для нашої моделі машинного навчання. Дата сет для навчання містить текстові відгуки та відповідні категорії, що дозволяє моделі навчитися розрізняти різні типи відгуків і правильно класифікувати їх за категоріями. Вибір такого дата сету був зумовлений необхідністю мати репрезентативні зразки відгуків, що покривають широкий спектр категорій, таких як "Доставка", "Якість", "Ціна" та інші. Це дозволяє моделі отримати різноманітні приклади, що допомагають їй краще навчитися розпізнавати патерни в тексті, що відповідають кожній категорії.

Формат дата сету для навчання містить два основних поля: текст відгуку та категорія. Текст відгуку подається у форматі звичайного тексту, що містить деталі про досвід користувача з продуктом або послугою. Категорія визначає, до якого типу належить цей відгук. Такий формат дозволяє моделі машинного навчання зрозуміти, як текстові характеристики пов'язані з певними категоріями, і навчитися класифікувати нові відгуки на основі цих знань.

Дата сет для тестування містить тільки текст відгуків без вказання категорій. Це дозволяє оцінити, наскільки добре модель здатна передбачити категорію для нових, невідомих їй відгуків. Формат тестового дата сету аналогічний до навчального, але без категорії, що забезпечує можливість безпосередньої перевірки точності передбачень моделі на реальних даних.

Тренувальна та тестова моделі наведені на рисунках 4.3.1 та 4.3.2 відповідно.

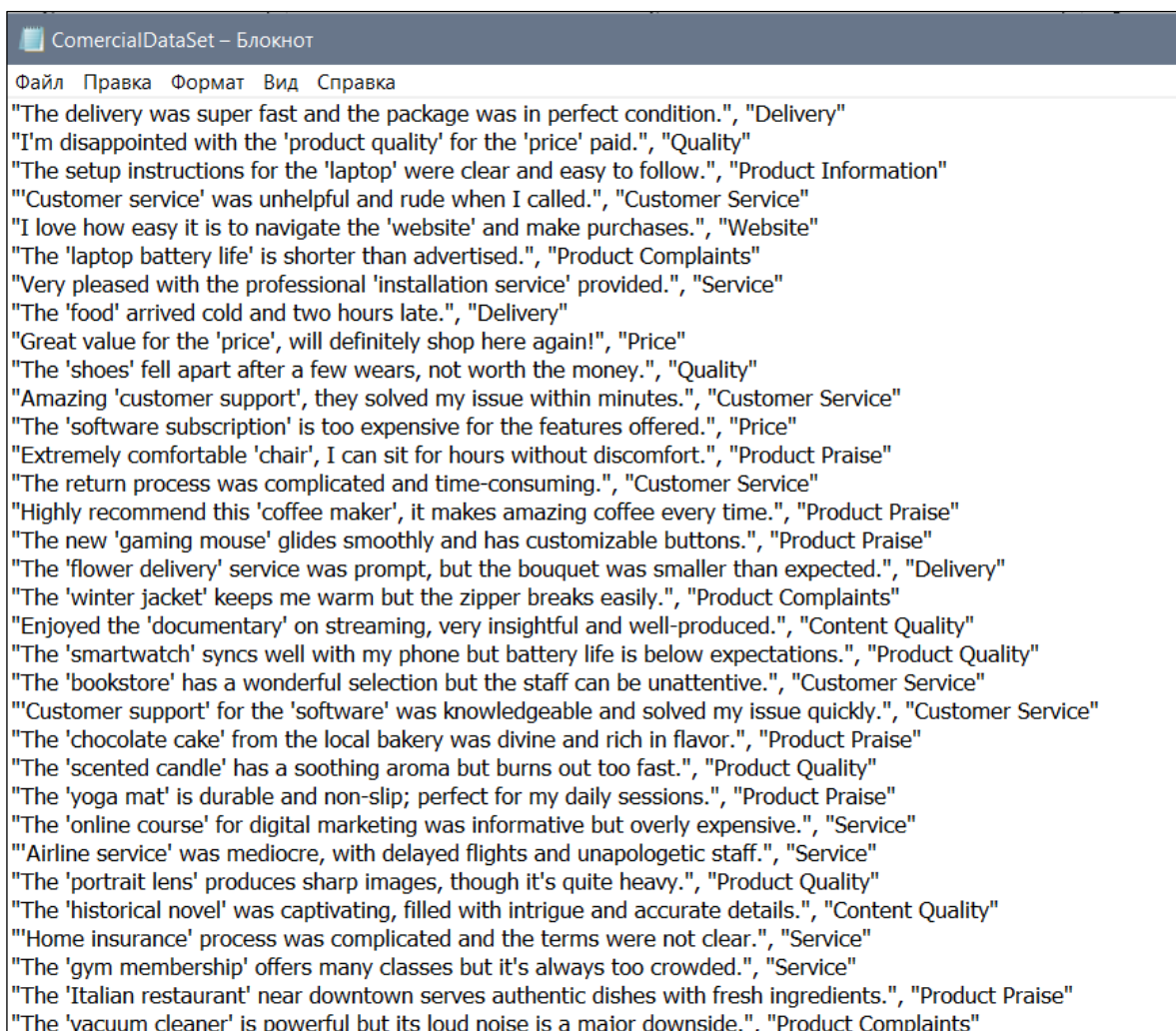


Рисунок 4.3.1 – Схема роботи тренувальної моделі правил

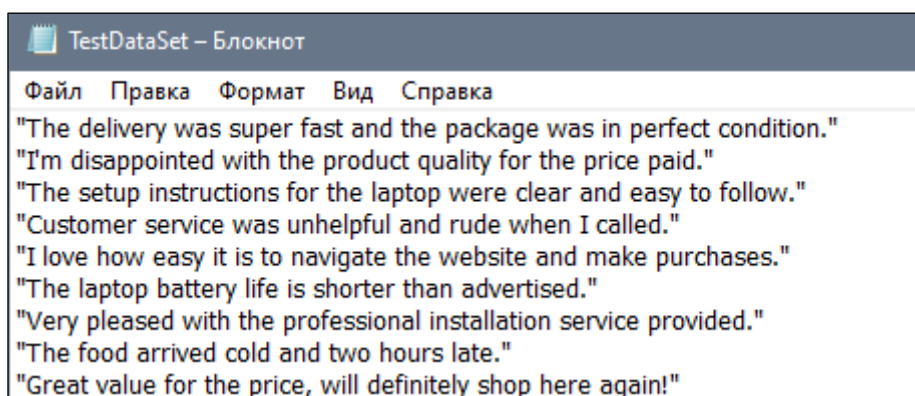


Рисунок 4.3.2 – Схема роботи тестової моделі правил

#### 4.4 Аналіз результатів

На рисунку 4.3 та 4.4 наведено результати відпрацювання ML.NET та TensorFlow відповідно.

```

Консоль отладки Microsoft Visual Studio
2: Test Model
3: Exit
Select an option: 2
Testing model, please wait...
Text: The delivery was super fast and the package was in perfect condition.
Predicted Category: Delivery
Commercial Object: package

Text: I'm disappointed with the product quality for the price paid.
Predicted Category: Quality
Commercial Object: object

Text: The setup instructions for the laptop were clear and easy to follow.
Predicted Category: Product Information
Commercial Object: laptop

Text: Customer service was unhelpful and rude when I called.
Predicted Category: Customer Service
Commercial Object: object

Text: I love how easy it is to navigate the website and make purchases.
Predicted Category: Website
Commercial Object: website

Model accuracy: 86%

E:\Users\bogac\Desktop\ConsoleApp4\bin\Debug\net6.0\ConsoleApp4.exe (процесс 12120)
Чтобы автоматически закрывать консоль при остановке отладки, включите параметр "С
томатически закрыть консоль при остановке отладки".

```

Рисунок 4.4.1 – Результаты ML.NET

```

Консоль отладки Microsoft Visual Studio
2: Test Model
3: Exit
Select an option: 2
Testing model, please wait...
Text: The delivery was super fast and the pac
Predicted Category: Quality
Commercial Object: object

Text: I'm disappointed with the product quali
Predicted Category: Quality
Commercial Object: object

Text: The setup instructions for the laptop w
Predicted Category: Product Information
Commercial Object: laptop

Text: Customer service was unhelpful and rude
Predicted Category: Customer Service
Commercial Object: object

Text: I love how easy it is to navigate the w
Predicted Category: Website
Commercial Object: website

Model accuracy: 70%

E:\Users\bogac\Desktop\ConsoleApp4\bin\Debug\
Чтобы автоматически закрывать консоль при ост

```

Рисунок 4.4.2 – Результаты TensorFlow

Результати роботи двох моделей, представлені на зображеннях, показують порівняння продуктивності моделей ML.NET та TensorFlow у завданні класифікації текстових відгуків. На першому зображенні ми бачимо результати моделі ML.NET, яка передбачає категорії відгуків і визначає комерційні об'єкти. Модель ML.NET показала точність у 86%, що свідчить про її здатність правильно класифікувати текстові відгуки у більшості випадків. Комерційні об'єкти, такі як "package", "laptop", і "website", також були ідентифіковані для деяких відгуків.

На другому зображенні представлені результати моделі TensorFlow, яка виконувала аналогічне завдання класифікації. Точність цієї моделі склала 70%, що є нижчим показником порівняно з моделлю ML.NET. Хоча і TensorFlow вдалося правильно передбачити деякі категорії відгуків, кількість правильних передбачень була меншою. Комерційні об'єкти, такі як "laptop" і "website", також були розпізнані, але частіше модель визначала об'єкти як "object", що свідчить про складність розпізнавання конкретних об'єктів у тексті для цієї моделі.

## ВИСНОВКИ

Кваліфікаційна робота по дослідженню області аналізу комерційних пропозицій та впровадженню моделей машинного навчання для класифікації текстових відгуків включала кілька ключових етапів, кожен з яких вніс значний вклад у досягнення загального результату. Спочатку було визначено цілі та завдання дослідження: розробити ефективну модель для класифікації відгуків за категоріями, а також визначення комерційних об'єктів у тексті. Було обрано два підходи для реалізації цих завдань: використання бібліотеки ML.NET та TensorFlow.

На першому етапі було створено два набори даних: один для навчання моделі, а інший для її тестування. Навчальний набір містив текстові відгуки та відповідні категорії, що дозволило моделі навчитися розрізняти різні типи відгуків. Тестовий набір складався лише з текстів відгуків, що дало можливість оцінити здатність моделі передбачати категорії для нових, невідомих їй даних. Формат даних був ретельно продуманий, щоб забезпечити репрезентативність і різноманітність зразків.

На етапі розробки було створено модель на основі ML.NET, яка включала в себе попередню обробку даних, зокрема перетворення тексту в числові вектори та перетворення категорій у ключі. Модель навчалася з використанням алгоритму SDCA Maximum Entropy, що дозволило досягти високої точності класифікації. Після цього модель була збережена для подальшого використання.

Аналогічним чином була розроблена модель на основі TensorFlow, що також включала попередню обробку даних і навчання на навчальному наборі. Результати тестування показали, що модель ML.NET мала вищу точність (86%) у порівнянні з моделлю TensorFlow (70%), що свідчить про кращу адаптованість першої до даного завдання.

Крім основної класифікації, було реалізовано простий метод для виявлення комерційних об'єктів у текстах відгуків. Цей метод допомагав визначати ключові елементи тексту, такі як "laptop", "coffee maker" тощо, що додавало цінності аналізу відгуків.

Загалом, дослідження показало, що вибір технології для аналізу тексту залежить від конкретних вимог і доступних ресурсів. Модель на основі ML.NET продемонструвала високу ефективність і точність, що робить її придатною для завдань класифікації текстів. Водночас модель TensorFlow показала, що може бути використана як альтернатива, але потребує додаткової оптимізації для досягнення кращих результатів. Це дослідження підкреслює важливість правильного вибору інструментів та методів у проектуванні систем машинного навчання для конкретних застосувань.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Criterion for Ranking Interval Alternatives in a Decision-Making Task / Y. Romanenkov et al. International Journal of Modern Education and Computer Science. 2024. Vol. 16, no. 2. P. 72–82. URL: <https://doi.org/10.5815/ijmecs.2024.02.06> (дата звернення: 02.04.2024).
2. Microsoft ML.NET Documentation. URL: <https://docs.microsoft.com/en-us/dotnet/machine-learning/> (дата звернення: 15.04.2024).
3. A new intelligence-based approach for rhinomanometric data processing / A. Yerokhin et al. 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 19–21 April 2016. 2016. URL: <https://doi.org/10.1109/elnano.2016.7493047> (дата звернення: 16.04.2024).
4. TensorFlow: An Open-Source Machine Learning Framework for Everyone. URL: <https://www.tensorflow.org/> (дата звернення: 18.04.2024).
5. Intelligent information system of heterogeneous medical data analysis / A. Yerokhin et al. 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 5–8 September 2017. 2017. URL: <https://doi.org/10.1109/stc-csit.2017.8098798> (дата звернення: 20.04.2024).
6. Chollet, F. "Deep Learning with Python." Manning Publications, 2021. URL: <https://www.manning.com/books/deep-learning-with-python> (дата звернення: 23.04.2024).
7. Hybrid system of computational intelligence based on bagging and group method of data handling / Y. Bodyanskiy et al. System research and information technologies. 2024. No. 1. P. 75–85. URL: <https://doi.org/10.20535/srit.2308-8893.2024.1.06> (дата звернення: 28.04.2024).
8. Intelligent information system of heterogeneous medical data analysis / A. Yerokhin et al. 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 5–8 September 2017. 2017. URL: <https://doi.org/10.1109/stc-csit.2017.8098798> (дата звернення: 30.06.2024).

9. Brownlee, J. "Machine Learning Mastery." URL: <https://machinelearningmastery.com/start-here/> (дата звернення: 31.04.2024).
10. A new intelligence-based approach for rhinomanometric data processing / A. Yerokhin et al. 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 19–21 April 2016. 2016. URL: <https://doi.org/10.1109/elnano.2016.7493047> (дата звернення: 01.05.2024).
11. Usage of phase space diagram to finding significant features of rhinomanometric signals / A. Yerokhin et al. 2016 XIth International Scientific and Technical Conference "Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 6–10 September 2016. 2016. URL: <https://doi.org/10.1109/stc-csit.2016.7589871> (дата звернення: 03.05.2024).
12. "A Beginner's Guide to Machine Learning with Python." URL: <https://www.datacamp.com/community/tutorials/machine-learning-python> (дата звернення: 04.05.2024).