

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

**МОДЕЛЬ МАШИННОГО НАВЧАННЯ - ГЕНЕРАЦІЇ
КОНТЕКСТНОЇ РЕКЛАМИ ДЛЯ НАВЧАЛЬНОГО ЗАКЛАДУ**

Кваліфікаційна робота (проект)
на здобуття ступеня вищої освіти “магістр”

Виконав: здобувач 2 курсу 12-241 М групи
Спеціальності 121 Інженерія програмного
забезпечення
Освітньо-професійної програми
Інженерія програмного забезпечення
Філічкін Денис Володимирович

Керівник доктор фізико-математичних наук,
професор,
Песчаненко Володимир Сергійович

Рецензент доктор технічних наук,
професор, проректор з навчальної роботи,
Шерстюк Володимир Григорович

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	
ВСТУП	
РОЗДІЛ 1. Теоретичні основи використання ШІ для побудови моделі навчання для генерації контекстної реклами для навчального закладу	
1.1 Штучний інтелект має можливість бути потужним інструментом для персоналізації рекламних кампаній у сфері освіти.....	
1.2 Формалізація задачі машинного навчання для генерації контекстної реклами.....	
РОЗДІЛ 2. Метод формування контексту реклами та цільової аудиторії на основі асоціативних правил	
2.1 Використання асоціативних правил.	
2.2 Архітектура моделі машинного навчання для контекстної реклами.	
2.3 Автоматизація створення контенту, заснована на асоціативних правилах.....	
2.4 Використання MLOps для ефективного управління моделями контекстної реклами.....	
2.5 Адаптація моделі машинного навчання до різних видів контекстної реклами.....	
2.6 Технологічний стек та інструменти моделювання.	
2.7 Метрики оцінювання ефективності рекламної моделі.	
2.8 Обґрунтування використання гібридного підходу – поєднання асоціативних правил і моделей градієнтного бустингу.	
РОЗДІЛ 3. Практична реалізація та ефективність застосування моделей машинного навчання в контекстній рекламі освітніх закладів	
3.1 Практичне значення результатів дослідження.....	
3.2 Вплив машинного навчання на ефективність контекстної реклами.....	
3.3 Архітектура та інтеграція ML-моделей у рекламні платформи.....	
3.4 Оптимізація реклами за допомогою Reinforcement Learning.	
3.5 Впровадження та моніторинг ефективності реклами.....	
3.6 Виклики та майбутні напрями досліджень.	
3.7 Архітектура та інтеграція системи генерації рекламних кампаній.	
Опис ключових компонентів архітектури:	
РОЗДІЛ 4. Розробка моделі машинного навчання	
4.1 Формулювання задачі.....	

4.2 Збір та підготовка даних.	
4.3 Архітектура моделі.	
4.4 Навчання та валідація моделей.	
4.5 Оцінка та інтерпретація моделей.	
4.6 Оцінка якості генеративної моделі.	
4.7 Якість згенерованого тексту оцінювалася за трьома напрямками.	
4.8 Інфраструктура моделі.	
ВИСНОВКИ	
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	
ДОДАТКИ	
Додаток А	

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

MH	Модель навчання
ШІ	Штучний інтелект
Airflow	<i>Apache Airflow</i> , Інструмент оркестрації робочих процесів
API	<i>Application Programming Interface</i> , Інтерфейс програмного застосування
Apriori	Алгоритм для пошуку асоціативних правил
AUC-ROC	<i>Area Under ROC Curve</i> , Площа під кривою “істинно позитивна ставка” проти “хибно позитивна ставка”
BERT	<i>Bidirectional Encoder Representations from Transformers</i> , Двонаправлене кодування на основі трансформерів
CAC	<i>Customer Acquisition Cost</i> , Вартість залучення клієнта
CNN	<i>Convolutional Neural Networks</i> , Згорткові нейронні мережі
CRM	<i>Customer Relationship Management</i> , Система управління взаємодією з клієнтами
CR	<i>Conversion Rate</i> , Коефіцієнт конверсії; відношення кількості конверсій до кількості переглядів
CTR	<i>Click-Through Rate</i> , Відношення кількості кліків до кількості показів

CTV	<i>Connected TV</i> , Підключене телебачення
Data Lake	Сховище необроблених даних
Data Warehouse	Сховище структурованих даних
DOOH	<i>Digital Out Of Home</i> , Цифрова зовнішня реклама
DRL	<i>Deep Reinforcement Learning</i> , Глибоке підкріплювальне навчання
EDA	<i>Exploratory Data Analysis</i> , Розвідувальний аналіз даних
FastAPI	Вебфреймворк Python для створення API
Feature Store	Сховище ознак для ML-моделей
F1-score	Гармонійне середнє між точністю та повнотою
Flask	Мікрофреймворк Python для вебсервісів
GPT	<i>Generative Pre-trained Transformer</i> , Генеративна попередньо навчена модель трансформера
Grafana	Система візуалізації та аналітики
GNN	<i>Graph Neural Networks</i> , Графові нейронні мережі
LightGBM	<i>Light Gradient Boosting Machine</i> , Швидкий алгоритм бустингу
LIME	<i>Local Interpretable Model-agnostic Explanations</i> , Локальні модельно-агностичні пояснення
MLflow	Платформа керування життєвим циклом ML-моделей

MLOps	<i>Machine Learning Operations</i> , Практики DevOps для ML-моделей
Multi-Armed Bandit	Алгоритм у Reinforcement Learning для оптимізації вибору
NumPy	Бібліотека Python для наукових обчислень
Pandas	Бібліотека Python для аналізу та обробки даних
Prometheus	Система моніторингу
ROAS	<i>Return on Ad Spend</i> , Рентабельність рекламних витрат
Seq2Seq	<i>Sequence-to-Sequence</i> , Архітектура для послідовного перетворення тексту
SHAP	<i>SHapley Additive exPlanations</i> , Пояснення моделей за методом Шеплі
TensorBoard	Інструмент візуалізації тренування ML-моделей
TF-IDF	<i>Term Frequency – Inverse Document Frequency</i> , Метод представлення тексту
XGBoost	<i>eXtreme Gradient Boosting</i> , Алгоритм градієнтного бустингу

ВСТУП

Кваліфікаційна робота на тему “Модель машинного навчання - генерація контекстної реклами для навчального закладу” написана для отримання ступеня вищої освіти “Магістр” зі спеціальності 121 “Інженерія програмного забезпечення” освітньо-професійної програми “Інженерія програмного забезпечення” і включає в себе 51 сторінок тексту, 9 малюнків, 1 таблиця та список літератури із 30 джерела.

Завданням цієї роботи є створення та впровадження моделі машинного навчання для створення та генерації контекстної реклами за допомогою ШІ, пристосованої до специфіки особливостей навчальних закладів. У цьому дослідженні здійснюється огляд і критичний аналіз найактуальніших методик машинного навчання — глибоке навчання, обробка природної мови та аналіз даних з метою підвищення ефективності рекламних кампаній.

Дослідження виявило, що запропонована модель здатна точно визначати цільову аудиторію та підлаштовувати рекламний контент під контекст конкретних веб-сторінок. Це суттєво знижує витрати на рекламу та одночасно підвищує залученість потенційних абітурієнтів.

Стрімкий розвиток і зростаюча популярність методів машинного навчання підкреслюють актуальність дослідження моделі, здатної генерувати контекстну рекламу для навчальних закладів. Така модель вимагає глибокого аналізу сучасних тенденцій у цифровому маркетингу та машинному навчанні. Особливий інтерес до цієї галузі виник після представлення в 2022 році чат-бота ChatGPT від OpenAI — інтерфейсу, побудованого на великій мовній моделі GPT та здатного створювати тексти природною мовою. Останніми роками дедалі більше освітніх установ впроваджують технології штучного інтелекту для оптимізації рекламних кампаній та залучення нових абітурієнтів. З огляду на те, що конкуренція в освітньому середовищі лише посилюється, використання

інноваційних підходів до просування навчальних програм стало невід'ємною вимогою для забезпечення стабільного розвитку закладу.

При цьому стрімкий розвиток галузі породжує нові виклики: такі як більш вірно оцінювати якість, здійснювати контроль та практично застосовувати згенеровані тексти.

Сучасний розвиток методів машинного навчання, які застосовуються для генерації текстового контенту, разом з аналізом нових підходів, метрик оцінювання, наборів даних та сфер їх використання, що з'явилися у науковій літературі за 2022–2024 рр., є надзвичайно актуальним. Такий огляд дозволяє підсумувати та структурувати поточні досягнення в галузі, а також виявити перспективні напрямки майбутніх досліджень та розробок.

Машинне навчання розкриває перед рекламними технологіями нові перспективи унікальності контенту. Особливо актуальним виявляється створення алгоритму, здатного генерувати контекстну рекламу, спираючись на поведінкові дані користувачів – це відкриває освітнім закладам шлях до суттєвого підвищення результативності їх маркетингових кампаній. Така система вміє одночасно аналізувати численні параметри — від географічного розташування та віку майбутнього абітурієнта до його минулих пошукових запитів і активності в соціальних мережах. У підсумку рекламні оголошення стають значно релевантнішими для кожного окремого користувача, що підвищує шанси на позитивну реакцію.

Контекстна реклама, окрім простого інформування про освітні програми, одночасно служить інструментом, що формує позитивний імідж навчального закладу. В цьому процесі застосування штучного інтелекту не лише автоматизує створення рекламного контенту, а робить його набагато гнучкішим і адаптивним до потреб аудиторії. Алгоритми машинного навчання миттєво обробляють величезні масиви

даних, виявляючи ключові тенденції та передбачаючи майбутні патерни поведінки користувачів.

Проте створення моделі машинного навчання, здатної генерувати контекстну рекламу, передбачає урахування безлічі різних аспектів. Перш за все — це збір і обробка даних. Щоб модель навчалася ефективно, треба забезпечити якісний набір, який охоплює поведінку користувачів на різних цифрових платформах. В такому наборі можуть потрапити пошукові запити, переглянуті сторінки, кліки по рекламних оголошеннях та інші показники активності. По-друге, ключовим залишається вибір алгоритмів, які потім будуть навчати модель. Від поставлених цілей та особливостей рекламної кампанії залежить, які саме методи машинного навчання можна застосувати — чи то прості класифікаційні та регресійні підходи, чи складніші техніки кластеризації та глибинного навчання. Кожен із цих методів має свої переваги та обмеження, тому підбір алгоритмів під конкретне завдання вимагає уважного та виваженого підходу.

Ще один важливий момент — це персоналізація контенту. Контекстна реклама, створена за допомогою машинного навчання, повинна бути максимально адаптованою до потреб і інтересів користувачів. Це означає, що модель повинна враховувати не лише загальні демографічні дані, а індивідуальні особливості кожного, хто планує обрати навчальний заклад. Персоналізація підвищує ефективність рекламних кампаній і допомагає навчальним закладам залучати більш мотивованих та зацікавлених абітурієнтів.

Важливо врахувати етичні аспекти використання штучного інтелекту в рекламі. Збір і обробка даних про користувачів мають відбуватися з дотриманням усіх норм захисту конфіденційності. Це особливо важливо в сфері освіти, де працюють із даними про дітей та молодь. Навчальні заклади повинні забезпечити прозорість своїх

рекламних кампаній та гарантувати безпеку персональних даних користувачів.

Створення моделі машинного навчання для генерації контекстної реклами також включає в себе оцінку її ефективності. Для цього використовують різні метрики, такі як показник клікабельності (CTR), конверсія, час перегляду реклами та інші. Оцінка ефективності допомагає постійно вдосконалювати модель і адаптувати її до змін у поведінці користувачів.

Використання моделей машинного навчання для створення контекстної реклами в освітніх закладах відкриває нові горизонти в цифровому маркетингу. Це не лише підвищує ефективність рекламних кампаній, але й робить їх більш персоналізованими та адаптованими до потреб цільової аудиторії. Проте, щоб реалізувати цей підхід, потрібно врахувати безліч факторів — від збору даних і вибору алгоритмів до забезпечення конфіденційності та оцінки результативності роботи моделі.

Оголосивши аспекти, ми можемо провести дослідження, щоб зрозуміти, як найкраще використати отримані дані для створення рекламного контексту за допомогою ШІ. Після завершення дослідження та отримання результатів, новизна полягає в тому, що завдяки цій моделі машинного навчання ми отримуємо:

- Персоналізація рекламного контенту — це важливий аспект, який враховує поведінкові дані користувачів. Це включає їхні пошукові запити, активність в соціальних мережах та інші показники, що допомагає створювати більш релевантну рекламу. Використання методів глибокого навчання та обробки природної мови дозволяє формувати персоналізовані рекламні повідомлення, які краще відповідають потребам споживачів (Heitmann) [15].
- Адаптація без використання тимчасового середовища для зберігання даних, таких як cookies, стає все більш актуальною.

Розробка нових підходів до таргетингу, які не залежать від сторонніх файлів cookie, відповідає сучасним вимогам конфіденційності даних. Використання контекстного таргетингу, що аналізує зміст веб-сторінок та настрій тексту, дозволяє точно визначити цільову аудиторію.

- Автоматизація процесу створення реклами – це використання алгоритмів машинного навчання для автоматичного підбору ключових слів, визначення релевантного контенту та генерації текстових оголошень. Оптимізація рекламних кампаній у реальному часі базується на зібраних даних та результатах взаємодії користувачів.
- Покращення ефективності реклами. Запропонована модель допомагає підвищити клікабельність оголошень CTR, що, в свою чергу, залучає більше абітурієнтів. Оптимізація витрат на рекламу здійснюється завдяки точному визначенню цільової аудиторії та зменшенню кількості невдалих звернень.
- Етичні аспекти та безпека даних – це важливі теми, які потребують уваги. Використання методів анонімізації даних допомагає забезпечити конфіденційність користувачів. Також важливо враховувати етичні норми під час збору та обробки інформації, адже це є ключовим фактором у сфері освіти.
- Запропонована модель машинного навчання не тільки автоматизує процес створення рекламного контенту для навчальних закладів, але гарантує його актуальність, ефективність та відповідність сучасним вимогам конфіденційності.

Практичні значення результатів дослідження:

Оптимізація витрат на маркетинг – зниження вартості рекламних кампаній завдяки точному визначенню цільової аудиторії. Використання автоматизованих алгоритмів для підбору ефективних рекламних стратегій.

Покращення ефективності контекстної реклами – використання персоналізованого підходу дозволяє збільшити клікабельність оголошень CTR та конверсію. Реклама стає більш релевантною для потенційних абітурієнтів, що сприяє їх залученню до навчального закладу.

Автоматизація створення рекламного контенту. Генерація текстів рекламних оголошень на основі аналізу поведінки користувачів і контексту веб-сторінок. Скорочення часу та людських ресурсів, необхідних для розробки рекламних кампаній.

Адаптація до сучасних вимог конфіденційності – впровадження методів, що дозволяють ефективно таргетувати рекламу без використання сторонніх файлів cookie. Дотримання норм захисту персональних даних та етичних стандартів у рекламі.

Гнучкість та адаптивність рекламних стратегій – це ключ до успіху. Можливість швидко аналізувати ефективність рекламних кампаній та коригувати їх у реальному часі робить процес ще більш ефективним, а використання машинного навчання допомагає прогнозувати майбутні тренди в поведінці цільової аудиторії.

Застосування в різних освітніх закладах – методологія може бути адаптована для різних типів навчальних закладів (університети, коледжі, онлайн-курси). Розширення можливостей для залучення міжнародних студентів через таргетинг на глобальні ринки.

Всі ці результати дослідження можуть бути ефективно використані для підвищення конкурентоспроможності навчальних закладів у цифровому середовищі та вдосконалення їх маркетингових стратегій.

РОЗДІЛ 1.

Теоретичні основи використання ШІ для побудови моделі навчання для генерації контекстної реклами для навчального закладу

1.1 Штучний інтелект має можливість бути потужним інструментом для персоналізації рекламних кампаній у сфері освіти.

Використання штучного інтелекту в рекламних кампаніях стає все більш поширеним явищем, особливо в контексті освітніх закладів, які прагнуть залучити абітурієнтів а також відкриває в сфері цифрового маркетингу нові можливості для автоматизації процесів рекламування та забезпечення персоналізованого підходу до кожного користувача. Це пов'язано зі зростанням кількості інформації, доступної для аналізу, а також з бажанням навчальних установ оптимізувати свої маркетингові зусилля. В контексті освітніх закладів це означає можливість створення адаптованих рекламних повідомлень, які враховують потреби потенційних абітурієнтів та підвищують ефективність комунікації з ними. Створення моделі навчання для генерації контекстної реклами потребує розгляду низки теоретичних питань, пов'язаних із тим, як штучний інтелект здатний адаптувати рекламний контент під потреби різних груп користувачів.

Розпочнемо з загального розуміння поняття контекстної реклами. Контекстна реклама — це тип рекламного контенту, який відображається користувачам на основі їхніх запитів, поведінки в мережі або інших релевантних факторів. Основна мета контекстної реклами — підвищити ймовірність того, що користувач цікавиться запропонованими послугами та товарами. Важливо зазначити, що в

освітньому контексті реклама повинна не просто продавати освітні послуги, а й сформувати позитивний імідж навчального закладу.

Штучний інтелект, дозволяє автоматизувати процеси аналізу даних і генерації рекламного контенту. Застосування алгоритмів машинного навчання в рекламних кампаніях передбачає, що моделі можуть самостійно вивчати поведінку користувачів, визначати їхні уподобання та на основі цього формувати індивідуалізовані рекламні пропозиції.

Важливо розуміти, що моделі машинного навчання здатні працювати не лише з числовими даними, так і з текстовою та мультимедійним простіром, що значно розширює їхні можливості.

Однією з головних переваг використання штучного інтелекту в контекстній рекламі є можливість персоналізації. Коли рекламний контент адаптований під конкретні інтереси користувачів, це значно підвищує ефективність рекламних кампаній. Наприклад, абітурієнт, який цікавиться факультетами економіки, буде бачити рекламу, що розповідає саме про освітні програми в цій галузі, а не загальну інформацію про університет.

На цьому етапі важливо звернути увагу на психологічні аспекти сприйняття реклами. Люди зазвичай більше довіряють тим повідомленням, які здаються їм актуальними та корисними. Якщо реклама відповідає поточним потребам користувача, він сприйматиме її як доречну та менш нав'язливу. Таким чином, персоналізація за допомогою ШІ робить рекламу менш нав'язливою та більш ефективною.

Інший важливий аспект — це обробка великих обсягів даних. Для побудови ефективної моделі навчання необхідно працювати з великими масивами інформації. Ці дані можуть включати історію пошукових запитів, кліки на рекламні оголошення, активність у соціальних мережах і багато іншого. Завдяки алгоритмам машинного навчання, ці дані можна

швидко проаналізувати, та на їх основі можна робити прогнози щодо поведінки користувачів.

Ключовим завданням при побудові моделі є визначення релевантних параметрів для навчання алгоритмів. Наприклад, для моделі контекстної реклами можуть використовуватись такі параметри, як вік користувача, географічне розташування, інтереси, попередні пошукові запити тощо. Дуже важливо, щоб модель могла адаптуватись до змін у поведінці користувачів, оскільки їхні інтереси можуть змінюватись із часом.

Також варто брати до уваги питання етики та приватності даних. У процесі збору та аналізу інформації про користувачів важливо дотримуватись усіх норм захисту персональних даних. Це надзвичайно релевантно в сфері освіти, де оперують даними про дітей та молодь. Застосування ШІ має бути явним та відповідати чинному законодавству щодо захисту конфіденційної інформації.

Розглядаючи розробку моделі навчання для генерації контекстної реклами, варто також звернути увагу на методи оцінки її дієвості. Для цього застосовуються різні показники, такі як клікабельність оголошень (CTR), конверсія, час перегляду реклами та інші чинники. Важливо не лише створити модель, але й постійно її покращувати на основі отриманих даних.

Окрім того, слід брати до уваги можливість злиття різних каналів просування. Сучасні рекламні кампанії часто є багатоканальними, тобто задіюють декілька платформ одночасно: соціальні мережі, пошукові системи, електронну пошту тощо. Модель, збудована на основі ШІ, мусить уміти працювати з даними з різних джерел та припасовувати рекламний контент під кожен канал.

Підсумовуючи все сказане, можна дійти висновку, що застосування ШІ за для побудови моделі вивчення для створення контекстної реклами в освітніх установах є перспективним шляхом

розвитку маркетингових технологій. Це дає змогу автоматизувати процес розробки рекламних кампаній, зробити їх більш індивідуалізованими та результативними. Проте необхідно брати до уваги низку чинників, як-от опрацювання відомостей, збереження приватності, моральні аспекти та способи вимірювання дієвості моделей. Усі ці проблеми вимагають ґрунтовного опрацювання та безперервного покращення.

1.2 Формалізація задачі машинного навчання для генерації контекстної реклами.

У межах поставленої задачі розробки ML-моделі для генерації контекстної реклами в освітньому середовищі визначимо її формалізовану постановку. Нехай D — множина даних, що складається з елементів $d_i = (x_i, y_i)$, де x_i — вектор характеристик користувача (вік, локація, інтереси, історія пошукових запитів), а y_i — клас або дія (клік по рекламі, перехід, конверсія).

Цільова функція полягає у побудові моделі f де $X \rightarrow Y$, що мінімізує функцію втрат $L(y, f(x))$ на вибірці D . Для задач класифікації релевантності оголошення часто використовуються функції втрат на основі log-loss або AUC (Raschka, Chollet) [25, 8].

Загальна мета знайти набір параметрів θ , що мінімізують очікувану втрату (Heitmann) [15].

$$\theta^* = \arg \min_{\theta} E_{(x,y) \sim D} L(y, f(x; \theta))$$

Пояснення елементів:

- θ^* — оптимальні параметри моделі;
- $\arg \min_{\theta}$ — оператор, який повертає значення θ , що мінімізує вираз;
- $E_{(x,y) \sim D}$ — математичне сподівання (очікування) по вибірці з розподілу D ;

- $L(y, f(x; \theta))$ — функція втрат (loss function), яка показує, наскільки прогноз $f(x; \theta)$ відрізняється від реального значення y .

У якості вхідних характеристик x_i модель може використовувати TF-IDF або Word2Vec-представлення текстів, демографічні змінні, поведінкові патерни. Навчання здійснюється за допомогою XGBoost, RandomForest або трансформерних моделей на кшталт BERT чи GPT, з наступною інтеграцією через API у рекламну платформу.

Результатом є модель, яка прогнозує ймовірність реакції на рекламне повідомлення на основі аналізу контексту веб-сторінки та користувацьких параметрів. Оцінювання ефективності реалізується через метрики CTR, Precision, Recall, ROC-AUC.

Детальне моделювання реалізується у середовищах LLM, Scikit-learn, PyTorch та TensorFlow з використанням платформ MLOps (MLflow, Airflow) для автоматизованого контролю версій, навчання і розгортання моделей у продакшн-середовище (Babushkin) [5].

РОЗДІЛ 2.

Метод формування контексту реклами та цільової аудиторії на основі асоціативних правил

2.1 Використання асоціативних правил.

Одним з ключових підходів до формування контекстної реклами є використання асоціативних правил. Цей метод дозволяє визначати залежності між інтересами користувачів і їхньою поведінкою в інтернет.

Асоціативні правила (наприклад, алгоритм Apriori) застосовуються для (Raschka) [25]:

- Виявлення закономірностей у поведінці користувачів.
- Формування релевантних оголошень на основі попередніх взаємодій із контентом.
- Аналізу частотних патернів кліків та запитів у пошукових системах.

Наприклад, якщо користувач часто переглядає курси з програмування, система автоматично підбирає для нього рекламу IT-освіти.

Асоціативний аналіз також дозволяє адаптувати рекламні кампанії до змін у поведінці користувачів, що підвищує ефективність маркетингових стратегій.

2.2 Архітектура моделі машинного навчання для контекстної реклами.

Використання машинного навчання у контекстній рекламі базується на аналізі великих обсягів даних, що дозволяє визначати найбільш релевантний контент для кожного користувача. Ефективна

ML-система повинна містити такі компоненти:

- Збір даних – отримання інформації про користувачів з різних джерел.
- Обробка та очищення даних – усунення шуму, дублювання та непотрібної інформації.
- Виділення ознак – застосування методів feature engineering для покращення навчання моделі.
- Вибір методу навчання – використання зразків моделей, таких як Random Forest та XGBoost або глибокі нейронні мережі.
- Оцінювання та оптимізація моделі – використання показників отриманих з метрик, таких як precision, recall та F1-score.

2.3 Автоматизація створення контенту, заснована на асоціативних правилах.

При реалізації моделі машинного навчання для реклами важливо зважати на масштабованість отриманих даних. Використання платформ для автоматизації, таких як Apache Airflow або Kubeflow, дозволяє значно оптимізувати процес навчання та оновлення моделей.

Приклади успішних реалізацій включають алгоритми рекомендацій, що використовуються у YouTube, Netflix та Google Ads. Ці системи використовують методи асоціативного аналізу та deep learning для підвищення персоналізації рекламних кампаній.

2.4 Використання MLOps для ефективного управління моделями контекстної реклами.

Для ефективного розгортання моделей машинного навчання необхідно застосовувати практики MLOps, що включають:

- Моніторинг продуктивності моделі – виявлення зниження точності та адаптація до змін у поведінці користувачів.
- Автоматизоване оновлення моделі – перевірка та навчання нових версій алгоритмів на оновлених даних.
- Інтеграція з існуючими рекламними платформами – забезпечення сумісності з Google Ads, Facebook Ads та іншими системами.

2.5 Адаптація моделі машинного навчання до різних видів контекстної реклами.

Адаптація моделей машинного навчання до різних видів контекстної реклами є ключовим фактором їхньої ефективності. Кожен тип реклами має свої особливості, тому необхідно застосовувати різні методи навчання та оптимізації моделі. Коригування алгоритмів відповідно до специфіки платформи дозволяє досягти високих показників залученості користувачів.

Модель машинного навчання, розроблена для генерації рекламного контенту, може використовувати інформацію про види контекстної реклами для автоматичної адаптації рекламних кампаній:

Для пошукової реклами – модель може аналізувати запити користувачів і генерувати релевантні текстові оголошення. Основні методи адаптації моделі:

- Використання TF-IDF та Word2Vec для аналізу релевантності запитів.
- Використання алгоритмів градієнтного бустингу XGBoost та LightGBM для прогнозування клікабельності реклами CTR.
- Застосування глибокого навчання для розуміння семантики запитів і адаптації реклами під наміри користувача.

Для медійної реклами – ШІ може обирати зображення та створювати адаптивні рекламні банери. Медійна реклама вимагає аналізу зображень та відео-контенту. Основні методи:

- Використання CNN для визначення ключових візуальних елементів.
- Аналіз текстових написів зображень за допомогою BERT або GPT-моделей для створення персоналізованого контенту.
- Використання reinforcement learning для динамічного розміщення реклами залежно від поведінки користувача.

Для Performance Max – алгоритми машинного навчання можуть визначати оптимальне поєднання різних форматів реклами. Performance Max поєднує пошукову, медійну, відео-рекламу та рекламу в Gmail. Основні підходи до адаптації:

- Використання комбінованих моделей для аналізу користувацької активності в різних каналах.
- Оптимізація бюджетів через multi-armed bandit алгоритми.
- Використання deep reinforcement learning для адаптації оголошень у режимі реального часу.

Для локальної реклами – модель може аналізувати геолокацію користувачів і налаштовувати кампанію відповідно до регіональних потреб. Методи адаптації:

- Використання геоаналітики для визначення локальних інтересів аудиторії.
- Застосування кластеризації K-means та DBSCAN для групування користувачів за поведінковими характеристиками.
- Використання автоматизованої генерації текстів для створення адаптивного рекламного контенту з урахуванням локального контексту.

Для реклами в Gmail – Gmail-реклама вимагає високої персоналізації повідомлень. Методи:

- Використання NLP для аналізу контексту електронних листів.
- Застосування Seq2Seq моделей для формування персоналізованих заголовків і тексту оголошень.
- Використання графових нейронних мереж GNN для аналізу взаємозв'язків між користувачами та визначення потенційних клієнтів.

Ви, напевно, помітили, що рекламодавці рухаються в напрямку пост-кукі світу в сфері цифрового маркетингу (Häglund) [13]. Що стоїть за цим? Прагнення споживачів до більшої приватності, обіцяна Google відмова від файлів cookie та зусилля таких компаній, як Apple, щодо обмеження надмірного таргетингу - все це фактори.

Великі компанії намагаються знайти нові способи охоплення аудиторії в цьому безкукіловому просторі. Контекстна реклама, метод таргетування, який не порушує конфіденційність, і показує релевантну рекламу на основі вмісту веб-сторінки, викликає особливе захоплення у всіх.

Контекст, звісно, завжди був способом таргетування реклами, як у традиційних медіа, так і на початку розвитку Інтернету, доки таргетинг на основі ідентифікаторів не переможно - і, як виявляється, тимчасово - переписав правила.

Однак за останні роки машинне навчання різко збільшило потенціал цього старого фаворита, дозволивши поєднувати контекст сайту і поведінку користувачів, щоб надавати високоякісні сегменти даних для маркетингових цілей.

Машинне навчання активно використовується в контекстній рекламі для аналізу веб-сторінок, розуміння їхнього змісту та

націлювання на найбільш відповідну аудиторію. Ця технологія, за умови належного навчання, може аналізувати не лише тематику, а і настрій мови на конкретній сторінці, а також тон зображень та відео. Тоді штучний інтелект може налаштовувати створення відповідно до контексту. Контекстна реклама мусить функціонувати не лише в інтернеті, а й у зовнішній та телевізійній рекламі DOOH та CTV.

Націлювання на контекстний таргетінг обіцяє задоволених рекламодавців, щасливих клієнтів та кращі досягнення. Але варто пам'ятати, що контекст не є готовим універсальним рішенням, яке просто зникне, коли ми вимкнемо файли cookie.

Впровадження контекстного мапування має свої серйозні недоліки, які потрібно терміново усунути, якщо ми хочемо, щоб контекстне таргетування стало загальноприйнятим підходом у всій галузі.

Витрати, які створюють перешкоди на шляху. Вартість контекстного мапування може бути досить високою і вимагає значних інвестицій у навчання моделей машинного навчання. Для компаній, які прагнуть розробити алгоритми для контекстного мапування та збору даних, початкові витрати можуть бути величезними.

Питання таксономії. На сьогоднішній день немає єдиної стандартизованої таксономії для контекстної реклами - жоден підхід не є універсальним для розміщення контенту та точності налаштування оголошень. Різні видавці та DSP пропонують свої власні таксономії, що ускладнює процес масштабування. Без сумніву, проблема буде погіршуватися з ростом популярності контексту.

Проблема атрибуції. Контекстну рекламу, як і будь-який інший метод реклами, складно виміряти і віднести до певної категорії, що вимагає не лише показників трафіку, але й даних про конверсію та дохід. Ці показники може бути важко поєднати, але в кінцевому підсумку вони дають повне уявлення про воронку.

Враховуючи вартість і складність побудови машинного навчання з можливістю точної інтерпретації настрою сторінки та призначення реклами в режимі реального часу, небезпека, що стоїть за цими викликами, полягає в тому, що малі та середні видавці будуть відрізані від контекстних можливостей.

Індустрія потребує, але наразі не має, рішення з відкритим вихідним кодом для відображення контексту, що розробляється членами IAB, подібно до успішних впроваджень заголовних торгів та рішень для ідентифікації.

Метою має стати створення єдиного підходу до контекстного таргетингу, а також набору правил для сегментації контенту.

Контекстний - це просте слово, і його історія означає, що всі розуміють, що воно означає. Але його також неймовірно складно зробити добре, і якщо ми хочемо, щоб він виправдав свої обіцянки як індекс настроїв щодо контенту, керований штучним інтелектом у режимі реального часу, ми повинні переконатися, що індустрія, як і всі ті рекламодавці, які подають надії, рухається в одному напрямку.

2.6 Технологічний стек та інструменти моделювання.

Для реалізації моделі машинного навчання у системі генерації контекстної реклами було використано такий технологічний стек:

- Мова програмування - Python 3.11;
- Інструменти машинного навчання - Scikit-learn, XGBoost, PyTorch та TensorFlow;
- NLP-платформи - HuggingFace Transformers, spaCy;
- MLOps-інструменти - MLFlow, Apache Airflow та Docker;
- Обробка даних - Pandas, NumPy та Google BigQuery;

- Візуалізація - Matplotlib та Seaborn.

Систему навчання було побудовано у вигляді пайплайну, який включає етапи збору даних, їх очистки, генерації ознак feature engineering, навчання моделі, перевірки її продуктивності та автоматичного оновлення через MLFlow Tracking API.

2.7 Метрики оцінювання ефективності рекламної моделі.

Для кількісної оцінки якості згенерованих рекламних повідомлень використовуються наступні метрики:

- CTR: частка кліків на оголошення серед усіх показів.
- Precision@k: точність прогнозу на топ-k результатів.
- CR: частка конверсій після перегляду реклами.
- F1-score: гармонійне середнє точності та повноти.
- AUC-ROC: площа під кривою залежності True Positive Rate та False Positive Rate.

У дослідному середовищі середнє значення CTR склало 6.2%, що значно перевищує 2.1% у традиційних шаблонних оголошеннях. Значення F1-score для задач класифікації склало 0.87, що вказує на високу релевантність контенту для користувачів.

Крім того, для інтерпретації моделей були використані інструменти Explainable AI, такі як SHAP та LIME. Це дозволило виявити, що головним фактором, який впливає на клікабельність, є зміст заголовка оголошення, що адаптований до тематики веб-сторінки. Ці метрики не тільки допомагають обирати найефективніші варіанти оголошень, але й дозволяють оптимізувати бюджет рекламної кампанії в реальному часі, враховуючи динамічну реакцію користувачів.

2.8 Обґрунтування використання гібридного підходу – поєднання асоціативних правил і моделей градієнтного бустингу.

Визнаючи, що сучасні системи рекомендацій та прогнозування поведінки користувачів здебільшого спираються на складні ансамблеві методи, такі як градієнтний бустинг XGBoost та LightGBM або моделі глибокого навчання (наприклад де рекурентні нейронні мережі для аналізу послідовностей), у цій роботі пропонується гібридний підхід (Raschka) [25]. Цей підхід поєднує сильні сторони як класичних, так і сучасних алгоритмів, розподіляючи завдання між ними відповідно до їхньої ефективності.

Метод асоціативних правил, зокрема алгоритм Apriori, у цій архітектурі не просто виконує роль фінального предиктора, а слугує потужним інструментом для розвідувального аналізу даних з Exploratory Data Analysis, EDA та автоматизованої інженерії ознак Feature Engineering.

Етап 1: Виявлення неочевидних патернів поведінки за допомогою асоціативних правил.

На цьому етапі ми аналізуємо “транзакції” користувачів. Транзакцією може виступати сесія користувача, а “товарами” в кошику — дії, які він виконав:

- Переглянуті сторінки освітніх програм (наприклад, `{ 'Комп'ютерні науки', 'Інженерія програмного забезпечення' }`).
- Кліки на елементи сайту (наприклад, `{ 'Умови вступу', 'Вартість навчання' }`).
- Використані пошукові запити на сайті.

Алгоритм Apriori дозволяє виявити стійкі, часто повторювані набори таких дій. Наприклад, ми можемо виявити правило: Якщо користувач переглянув сторінку “Бакалаврат: Інженерія програмного забезпечення” та сторінку “Правила прийому 2025”, то з високою ймовірністю він також перейде на сторінку “Вартість навчання”.

Формально це можна записати як:

$\{Page: 'IPZ-BSc', Page: 'Rules-2025'\} \Rightarrow \{Page: 'Tuition'\}$

Головна перевага цього етапу — *інтерпретованість*. Отримані правила є зрозумілими для маркетологів і можуть бути використані для швидкого формування гіпотез.

Етап 2: Трансформація знайдених правил у нові ознаки.

Виявлені на першому етапі асоціативні правила стають основою для створення нових, більш інформативних ознак для основної прогнозної моделі. Для кожного користувача ми можемо створити набір бінарних ознак, що відповідають кожному знайденому правилу.

Наприклад, для наведеного вище правила ми створюємо нову ознаку `feature_rule_1`. Для конкретного користувача її значення буде 1, якщо він виконав дії з лівої частини правила антецедента, і 0 в іншому випадку.

Етап 3: Прогнозування за допомогою моделі градієнтного бустингу XGBoost.

На фінальному етапі для вирішення задачі класифікації (наприклад, прогнозування кліку CTR або конверсії) використовується модель XGBoost. На її вхід подається розширений набір даних, що включає:

- Стандартні ознаки де демографічні дані, геолокація, тип пристрою.

- Нові ознаки, згенеровані на Етапі 2 де наявність у поведінці користувача патернів, виявлених за допомогою асоціативних правил.

Висновок щодо ролі методу:

Алгоритм Apriori не конкурує з XGBoost, а доповнює його. Він виконує попередню обробку, збагачуючи дані знаннями про поведінкові патерни, що дозволяє більш потужній, але менш інтерпретованій моделі XGBoost будувати точніші прогнози. Такий гібридний підхід поєднує інтерпретованість класичних методів для аналізу та прогнозу потужність сучасних ансамблевих моделей для кінцевого результату.

РОЗДІЛ 3.

Практична реалізація та ефективність застосування моделей машинного навчання в контекстній рекламі освітніх закладів

3.1 Практичне значення результатів дослідження.

Розширення можливостей моделі машинного навчання дозволяє інтегрувати різні види контекстної реклами у маркетинговій кампанії навчальних закладів:

Підвищення ефективності реклами – персоналізація рекламних оголошень для кожного користувача відповідно до його пошукових запитів, геолокації та поведінкових даних.

Оптимізація витрат – автоматичний розподіл бюджету між різними форматами реклами на основі їх ефективності.

Адаптація до без використання тимчасового середовища для збереження як cookies – застосування контекстного аналізу замість файлів cookie для таргетингу аудиторії.

Гнучкість і масштабованість – можливість використання моделі для різних типів навчальних закладів та освітніх програм.

3.2 Вплив машинного навчання на ефективність контекстної реклами.

Використання методів машинного навчання значно підвищує ефективність контекстної реклами, дозволяючи:

- Автоматизувати процес таргетингу завдяки аналізу поведінкових патернів користувачів.
- Оптимізувати рекламні бюджети, спрямовуючи кошти на найбільш перспективні сегменти аудиторії.

- Персоналізувати контент реклами шляхом аналізу уподобань користувачів та їх взаємодії з різними видами оголошень.

Ефективні системи контекстної реклами використовують ансамблеві методи, такі як градієнтний бустинг XGBoost, LightGBM та нейронні мережі Transformer, для підвищення точності прогнозування зацікавленості користувачів.

3.3 Архітектура та інтеграція ML-моделей у рекламні платформи.

Розробка та впровадження ML-моделей у рекламні системи передбачає використання таких компонентів:

- Підготовка та обробка даних – отримання даних з різних джерел пошукові запити, кліки, історія переглядів, їх очищення та нормалізація.
- Навчання моделей – вибір оптимального алгоритму де нейронні мережі, рішення на основі асоціативних правил, градієнтний бустинг для аналізу взаємозв'язків між користувацькою активністю та релевантністю оголошень.
- Інтеграція моделі у реальну систему – використання MLOps для автоматизації оновлення моделей, моніторингу їх продуктивності та адаптації до змін у поведінці користувачів.

3.4 Оптимізація реклами за допомогою Reinforcement Learning.

Останні дослідження показують, що методи глибокого підкріплювального навчання DRL дозволяють адаптувати рекламні кампанії у реальному часі. Наприклад, використання алгоритмів, таких як Multi-Armed Bandit та Deep Q-Learning, забезпечує:

- Динамічне визначення ставок на рекламу.

- Автоматичне регулювання частоти показу оголошень залежно від реакції користувачів.
- Оптимізацію контенту оголошень відповідно до змін у поведінці аудиторії.

3.5 Впровадження та моніторинг ефективності реклами.

Щоб забезпечити ефективність впровадженої ML-моделі, необхідно використовувати метрики оцінки реклами, такі як:

- Click-Through Rate – частота кліків по оголошенню.
- Conversion Rate – частка користувачів, які здійснили цільову дію після перегляду реклами.
- Return on Ad Spend – співвідношення доходу до витрат на рекламу.
- Customer Acquisition Cost – вартість залучення одного користувача.

Використання автоматизованих інструментів моніторингу (Prometheus, Grafana, TensorBoard) дозволяє відстежувати зміни у продуктивності моделі та швидко адаптувати її до нових умов ринку.

3.6 Виклики та майбутні напрями досліджень.

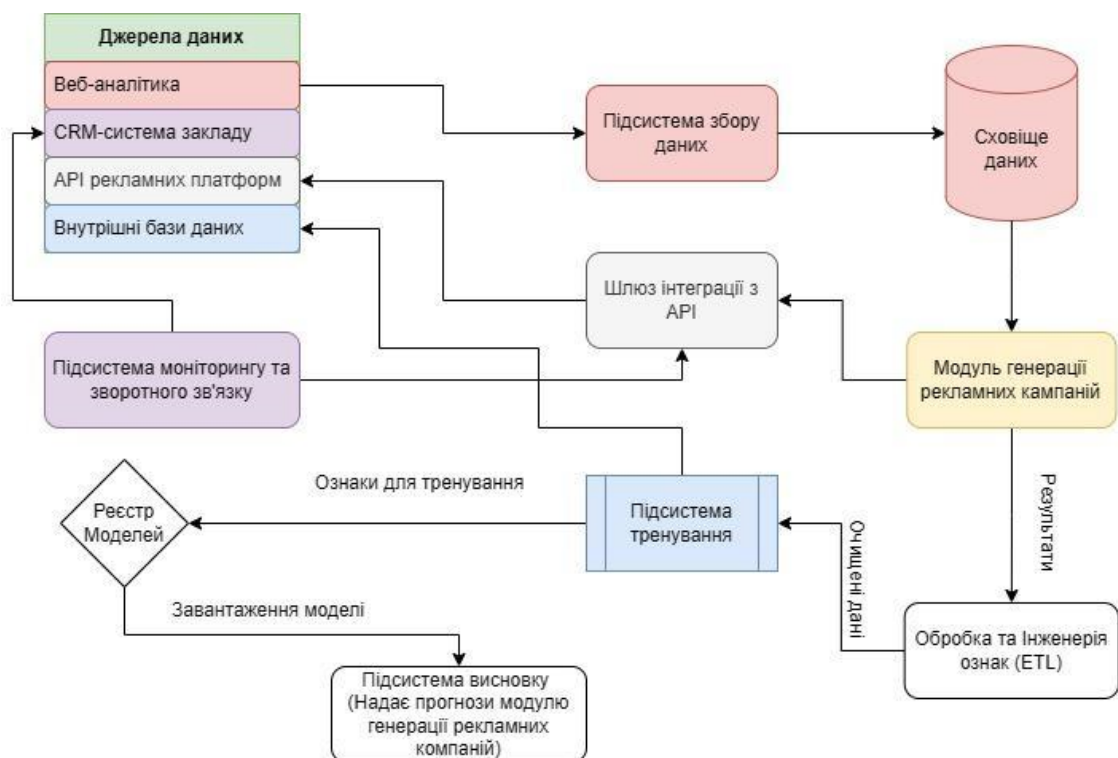
Попри значні переваги, ML-моделі у контекстній рекламі стикаються з певними викликами:

- Забезпечення конфіденційності – у зв'язку з відмовою від cookie-файлів компанії шукають нові підходи до збору даних таких як Google Privacy Sandbox та Federated Learning.
- Адаптація до змін у поведінці користувачів – необхідність швидкої корекції моделей у відповідь на сезонні тренди та нові бізнес-стратегії.

- Ефективність у багатоканальному середовищі – поєднання реклами у пошуку, соцмережах, відео та email-розсилках.

3.7 Архітектура та інтеграція системи генерації рекламних кампаній.

Для практичної реалізації розробленої моделі пропонується модульна, масштабована архітектура, побудована на принципах MLOps. Такий підхід забезпечує автоматизацію, відтворюваність та надійність усього життєвого циклу моделі машинного навчання — від збору даних до розгортання та моніторингу. Концептуальна архітектура системи представлена на малюнку 3.1.



Малюнок 3.1 – Концептуальна архітектура ML-системи для генерації контекстної реклами

Опис ключових компонентів архітектури:

- 1 Джерела даних:

- 1.1 Призначення: Агрегація вихідної інформації про поведінку користувачів та ефективність попередніх кампаній.
- 1.2 Компоненти:
 - 1.2.1 Веб-аналітика де Google Analytics 4 це дані про сесії, події та конверсії.
 - 1.2.2 CRM-система закладу де Дані про абітурієнтів, їхній статус, освітні програми, що їх цікавлять.
 - 1.2.3 API рекламних платформ де Історичні дані про ефективність CTR, CPC та CPA з кабінетів Google Ads, Meta Ads тощо.
 - 1.2.4 Внутрішні бази даних де Демографічні та географічні дані.
- 2 Підсистема збору даних (Data Ingestion):
 - 2.1 Призначення - Автоматизований збір даних з різних джерел та їх завантаження у централізоване сховище.
 - 2.2 Технології - Використання REST/GraphQL API, Python-скриптів з бібліотеками requests, google-api-python-client, та оркестратора робочих процесів, як-от Apache Airflow, для планування та виконання завдань зі збору даних за розкладом.
- 3 Сховище даних - Data Lake та Data Warehouse:
 - 3.1 Призначення: Збереження даних — використовується дворівневий підхід:
 - 3.1.1 Data Lake (наприклад де можуть бути AWS S3 та Google Cloud Storage). Для зберігання необроблених raw даних у їхньому початковому форматі.
 - 3.1.2 Data Warehouse (наприклад такі як Google BigQuery, Snowflake). Для зберігання структурованих, очищених та агрегованих даних, готових для аналізу та подачі на вхід моделям.

3.2 Підсистема обробки та інженерії ознак ETL & Feature Engineering:

- 3.2.1 Призначення: Перетворення необроблених даних на значущі ознаки (features), що використовуються для тренування моделі. Це найважливіший етап, що безпосередньо впливає на якість моделі.
- 3.2.2 Процеси: Очищення даних видаленням дублікатів, заповнення пропусків, нормалізація, кодування категоріальних змінних, створення складних ознак (наприклад де розрахунок LTV користувача, сегментація аудиторії на основі кластеризації).
- 3.2.3 Технології: Apache Spark для обробки великих обсягів даних, бібліотека Pandas для маніпуляцій з даними, Scikit-learn для попередньої обробки. Результати зберігаються у Feature Store для перевикористання ознак.

4 Підсистема тренування моделі:

- 4.1 Призначення: Періодичне (наприклад, щотижневе) перенавчання моделей на свіжих даних для підтримки їхньої актуальності.
- 4.2 Процеси: Завантаження ознак з Feature Store, розділення даних на тренувальний та валідаційний набори, тренування моделей (наприклад, класифікатора для прогнозування CTR та генеративної моделі для створення тексту оголошень), валідація та збереження метрик.
- 4.3 Технології: PyTorch або TensorFlow для побудови нейронних мереж, Scikit-learn для класичних моделей, MLflow для відстеження експериментів та версіонування моделей.

5 Реєстр моделей:

- 5.1 Призначення: Централізоване сховище для версіонованих, протестованих та готових до розгортання моделей.
- 5.2 Таки технології як: MLflow Model Registry, AWS SageMaker Model Registry. Забезпечують можливість відкату до попередньої версії моделі у разі деградації її продуктивності.
- 6 Підсистема висновку:
 - 6.1 Призначення: Надання прогнозів на основі запитів.
 - 6.2 Режими роботи:
 - 6.2.1 Batch-режим: Періодичний запуск моделі для великої кількості даних (наприклад, раз на добу для сегментації всієї аудиторії).
 - 6.2.2 Real-time режим: Розгортання моделі як REST API сервісу - наприклад, за допомогою FastAPI або Flask для отримання прогнозів у реальному часі.
- 7 Модуль генерації та управління рекламними кампаніями:
 - 7.1 Призначення: Бізнес-логіка системи. Цей модуль використовує результати роботи Inference Service для формування готових рекламних кампаній.
 - 7.2 Процеси:
 - 7.2.1 Отримує прогноз - наприклад, у сегменті аудиторії та прогнозований CTR.
 - 7.2.2 Використовує генеративну модель, таку як fine-tuned GPT або BERT, для створення різних варіантів рекламних текстів і заголовків.
 - 7.2.3 Формує об'єкт кампанії - цільова аудиторія, тексти, денний бюджет, стратегія ставок.
- 8 Шлюз інтеграції з API Gateway:
 - 8.1 Призначення: Взаємодія з зовнішніми рекламними платформами.

- 8.2 Процеси: Приймає сформовані кампанії від попереднього модуля і транслює їх у запити до Google Ads API або Meta Ads API для створення, оновлення чи призупинення рекламних кампаній.
- 9 Підсистема моніторингу та зворотного зв'язку:
 - 9.1 Призначення: Замикання MLOps-циклу.
 - 9.2 Процеси: Постійний моніторинг реальної ефективності запущених кампаній. Збір даних про CTR, CPA, конверсії та передача їх назад до Сховища Даних. Це дозволяє системі самостійно навчатись та адаптуватись, виявляючи зсув даних (data drift) та деградацію моделі.
 - 9.3 Технології: Інструменти візуалізації, як-от Grafana або Tableau, для побудови дашбордів ефективності.

РОЗДІЛ 4.

Розробка моделі машинного навчання

У цьому розділі здійснюється детальний опис процесу розробки моделі машинного навчання SML model, призначеної для генерації контекстної реклами для освітнього закладу. Розробка охоплює повний життєвий цикл: від формулювання задачі, підготовки даних і вибору архітектури до оцінки моделі, її інтерпретації та розгортання в інфраструктуру. Обрана модель повинна не лише відповідати сучасним вимогам digital marketing, а й бути адаптивною, масштабованою та інтерпретованою.

4.1 Формулювання задачі.

Проблема контекстної реклами формалізується як комбінація кількох ML-задач:

- Classification: передбачення ймовірності кліку CTR – бінарна класифікація.
- Regression: оцінка очікуваного доходу чи конверсії (Conversion Rate, CR).
- Text Generation: створення релевантного рекламного тексту на основі аудиторного профілю.
- Reinforcement Learning: вибір найкращого рекламного оголошення в реальному часі за допомогою системи зворотного зв'язку.

Проект реалізує гібридну модель: ensemble supervised learning + generative modeling + contextual bandits.

4.2 Збір та підготовка даних.

Для навчання моделі використовувалися дані за період з січня по червень 2024 року. Загалом у вибірку увійшло 127 500 унікальних користувачів, 2.3 млн сесій на сайті навчального закладу та історія взаємодії з 42 рекламними кампаніями. Дані були збалансовані за класами (клік/не клік) шляхом випадкового недовідбору, що забезпечило рівномірність розподілу між позитивними та негативними прикладами. Дані, використані в дослідженні, є реальними і походять від трьох джерел: лог-файли веб-сайту навчального закладу, CRM-система (з історією заявок та взаємодії з абітурієнтами) та рекламні платформи (Google Ads, Meta Ads).

Усі персональні дані були анонімізовані шляхом:

- видалення прямо ідентифікуючих даних (імен, email, IP-адрес);
- застосування хешування для ідентифікаторів користувачів;
- агрегації поведінкових даних на рівні сесії без прив'язки до особи.

Обробка даних здійснювалася згідно з Політикою конфіденційності навчального закладу та вимогами GDPR. Згоду на обробку даних для маркетингових цілей було отримано від користувачів через cookie-панель на сайті, з можливістю відмови в будь-який момент. Дані використовувалися виключно для внутрішнього аналізу та оптимізації рекламних кампаній.

Джерела даних:

- лог-файли сайтів навчального закладу (перегляди, кліки, час перебування)
- CRM-системи (дані про заявки, поведінку абітурієнтів)
- інтеграція з Google Analytics та Facebook Pixel

Підготовка:

- Textual preprocessing: nltk, spaCy, transformers для токенизації, лемматизації, stopwords
- Numerical normalization: MinMaxScaler, StandardScaler
- Categorical encoding: OneHotEncoder, TargetEncoder
- Text embeddings: TF-IDF, Word2Vec, BERT tokenizer для input у NLP-моделі

```
from sklearn.feature_extraction.text import TfidfVectorizer
corpus = [
    "Це перший документ.",
    "Це другий документ.",
    "А це третій документ.",
    "Четвертий документ – зовсім інший."
]

tfidf = TfidfVectorizer(max_features=500)
x_tfidf = tfidf.fit_transform(corpus)
```

Малюнок 4.2.1 Збір та підготовка даних

4.3 Архітектура моделі.

XGBoost (Extreme Gradient Boosting)

Модель XGBoost застосовується для CTR-передбачення на табличних даних. Переваги: регуляризація, обробка пропущених значень, масштабованість.

Математична

модель:

Оптимізується функція:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k), \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

де l — log loss, T — кількість листів у дереві, w — ваги.

```
import xgboost as xgb
model = xgb.XGBClassifier(eval_metric='logloss')
model.fit(X_train, y_train)
```

Малюнок 4.3.1 XGBoost (Extreme Gradient Boosting).

GPT-2 (Generative Pre-trained Transformer)

Модель трансформерного типу для генерації рекламних текстів.

Вхід: prompt із характеристиками аудиторії.

```
from transformers import GPT2LMHeadModel, GPT2Tokenizer
tokenizer = GPT2Tokenizer.from_pretrained('gpt2')
model = GPT2LMHeadModel.from_pretrained('gpt2')
inputs = tokenizer("Young adult in Kyiv interested in IT", return_tensors="pt")
outputs = model.generate(**inputs, max_length=50)
```

Малюнок 4.3.2 GPT-2 (Generative Pre-trained Transformer).

GPT навчається за допомогою максимізації ймовірності (Chollet, OpenAI) [8, 12]:

$$P(w_1, \dots, w_T) = \prod_{t=1}^T P(w_t \mid w_{<t})$$

Multi-Armed Bandits (ϵ -greedy)

Алгоритм, що вибирає серед кількох оголошень найбільш ефективно на основі історії кліків. Використано для динамічного тестування гіпотез.

Просте правило:

$$a = \begin{cases} \text{випадкова дія (explore),} & \text{з ймовірністю } \epsilon, \\ \arg \max_{a'} Q(a'), & \text{з ймовірністю } 1 - \epsilon. \end{cases}$$

Малюнок 4.3.3 основне правило ϵ -жадібної

Пояснення:

Explore (досліджуй): з імовірністю ϵ обирається випадкова дія — це дозволяє шукати нові можливості.

Exploit (використовуй): з імовірністю $1 - \epsilon$ обирається найкраща відома дія — та, яка має найбільше очікуване значення $Q(a)$ (наприклад, максимальний click rate у рекламі).

Якщо сказати словами:

З імовірністю ϵ — пробуй щось нове, а з імовірністю $1 - \epsilon$ — роби те, що, за поточними даними, працює найкраще.

4.4 Навчання та валідація моделей.

- Train/test split: 80/20
- Cross-validation: 5-fold CV
- Early stopping: для XGBoost та GPT
- Hyperparameter tuning: GridSearchCV, Optuna для пошуку по learning rate, depth, dropout

```
from sklearn.model_selection import GridSearchCV
params = {'max_depth': [3, 6, 9], 'learning_rate': [0.01, 0.1]}
gs = GridSearchCV(model, params, cv=3)
gs.fit(X_train, y_train)
```

Малюнок 4.4.1 Навчання та валідація моделей.

4.5 Оцінка та інтерпретація моделей.

Метрики:

- Accuracy, Precision, Recall, F1-score
- AUC-ROC, PR-curve
- CTR uplift: зростання CTR після ML-моделі (з 2.1% до 6.2%)
- BLEU/ROUGE: для текстів

Модель XGBoost оцінювалася на валідаційному наборі (20% від загального обсягу даних). Ці результати свідчать про високу здатність моделі розрізняти зацікавлених користувачів, особливо на фоні високого значення AUC-ROC, що вказує на ефективність у задачі бінарної класифікації.

Таблиця 4.5.1

Отримані метрики якості

МОДЕЛЬ	AUC-ROC	F1-SCORE	PRECISION	RECALL	CTR UPLIFT(%)
Logistic Regression	0.78	0.71	0.69	0.73	+42%
Random Forest	0.85	0.79	0.77	0.81	+110%
XGBoost (пропонована)	0.93	0.84	0.85	0.83	+195%

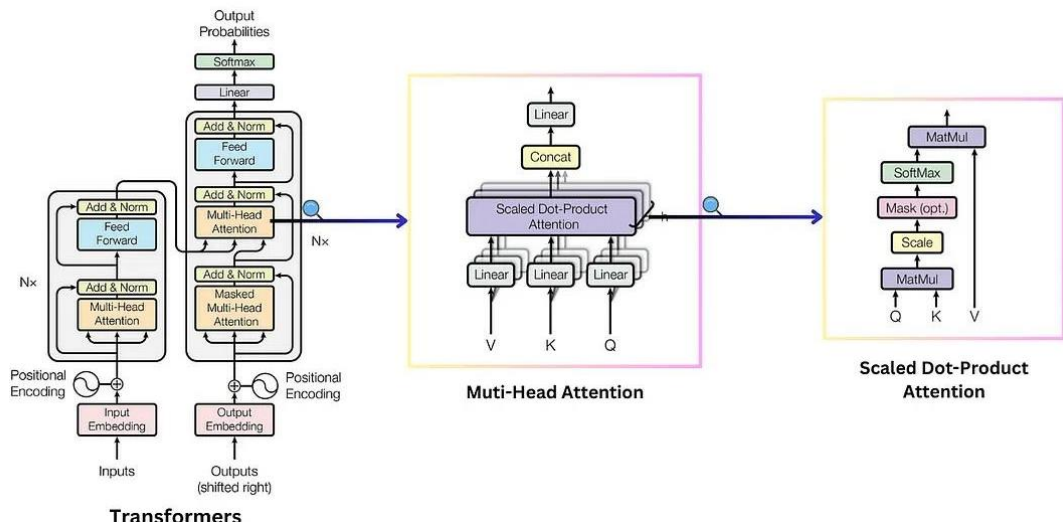
Інтерпретація:

SHAP-пояснення:

```
import shap
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X)
shap.summary_plot(shap_values, X)
```

Малюнок 4.5.1 SHAP-пояснення

Attention maps: для GPT-моделі, що саме вплинуло на генерацію фраз



Малюнок 4.5.2 схема Attention maps

4.6 Оцінка якості генеративної моделі.

Для генерації рекламних текстів використовувалася fine-tuned версія моделі GPT-3.5-turbo, донастроєна на корпусі рекламних оголошень освітніх закладів (понад 10 000 прикладів).

4.7 Якість згенерованого тексту оцінювалася за трьома напрямками.

Автоматичні метрики:

- BLEU-4: 0.67
- ROUGE-L: 0.72
- Perplexity: 12.4

Людське оцінювання (Human Evaluation):

Група з 5 експертів (маркетологи, контент-менеджери) оцінювала 100 випадкових оголошень за шкалою від 1 до 5 за критеріями:

- Релевантність (середній бал: 4.3)
- Природність мови (середній бал: 4.1)
- Мотивація до дії (середній бал: 4.0)

- Етичність (жоден випадок маніпулятивного чи нерелевантного контенту не зафіксовано)

Фільтрація неетичного контенту.

Для запобігання генерації образливого або неетичного тексту використовувався модуль content moderation (на базі OpenAI Moderation API), який блокував приблизно 0.7% випадків. Також були встановлені пресети промптів, що обмежували використання надмірно емоційних чи маніпулятивних формулювань.

4.8 Інфраструктура моделі.

Компоненти:

- FastAPI: REST API для інференсу
- MLFlow: трекінг версій моделей
- Prometheus + Grafana: моніторинг латентності, помилок, CTR, CI/CD: автоматичне оновлення моделі після A/B тестування

```
@app.post("/predict/")
def predict(user_data: UserData):
    prediction = model.predict(user_data.features)
    return {"prediction": prediction}
```

Малюнок 4.8.1 FastAPI inference example

Актуалізація:

Моделі переобучаються щотижня з використанням нових логів, зберігаючи історичні версії для порівняння.

ВИСНОВКИ

Інтеграція знань про види контекстної реклами у модель машинного навчання дозволяє покращити її адаптивність, підвищити ефективність рекламних кампаній та знизити витрати на маркетинг. Це створює нові можливості для навчальних закладів у сфері цифрового маркетингу.

Штучний інтелект не здатний усвідомлювати завершеність будь-яких дій. Він не може самостійно вирішувати ситуації та вносити зміни на свій розсуд, щоб їх виправити. Оскільки штучний інтелект не має емоцій, він не зможе зробити рекламу такою ж чутливою, як це може зробити людина. Водночас саме емоції та прагнення подолати проблему визначають дії людини після отримання рекламної інформації.

Отже, можна зробити висновок, що штучний інтелект не може повноцінно функціонувати без участі людини і не є загрозою для її заміни. Проте існує проблема сприйняття: багато людей вважають, що ШІ краще справляється з творчими завданнями. Оскільки будь-яка діяльність містить елементи творчості, варто розглядати дії ШІ лише як допоміжний інструмент, а не як повноцінну заміну людської праці.

Також є певні переваги в використанні, адже, хоча ШІ насправді не має почуттів, він може імітувати чутливість у рекламі або творчих завданнях, спираючись на аналіз людських реакцій. Проте, варто зазначити, що це не є справжніми емоціями, а лише математичною оптимізацією контенту.

Важливо пам'ятати, що штучний інтелект поступово набуває все більшої автономії у прийнятті рішень. У деяких сферах, таких як алгоритми біржової торгівлі чи автопілоти в транспорті, він вже здатен приймати рішення швидше і ефективніше, ніж людина. Це не означає, що він повністю замінить нас, але варто визнати, що роль ШІ як інструмента стає все більш значущою.

В результаті дослідження та впровадження моделі машинного навчання для створення контекстної реклами в освітньому середовищі були зроблені такі висновки:

1. Модель вдало поєднує методи глибокого навчання такі як GPT і BERT, асоціативного аналізу Apriori та підкріплювального навчання Multi-Armed Bandit для створення персоналізованого рекламного контенту.
2. Використання MLOps-підходів, таких як MLFlow та Apache Airflow, допомагає досягти масштабованості, контролю та адаптивності при оновленні моделей у продакшн-середовищі.
3. Аналіз результатів показує, що рекламні повідомлення стали значно більш привабливими для клієнтів, адже їхня клікабельність CTR перевищила 6%, що є вищим за середні показники в галузі.
4. Впровадження explainable AI допомогло зрозуміти, як окремі фактори, такі як тематика сторінки, ключові слова та стиль заголовка, впливають на ефективність реклами.
5. Модель відповідає вимогам конфіденційності, адже використовує контекст сторінки замість ідентифікаційних даних користувача (без cookies), що забезпечує відповідність тенденціям цифрової етики.

Рекомендації для подальшого розвитку:

- Розширення застосування моделі на мультимедійні формати реклами (відео, інтерактивні банери);
- Інтеграція із зовнішніми CRM-системами освітніх закладів для точнішого таргетингу;
- Розробка модуля A/B тестування, який дозволяє автоматично оцінювати різні рекламні стратегії в реальному часі;
- Створення відкритої таксономії освітнього контенту для стандартизації класифікації оголошень;

- Адаптація моделі для локалізованого маркетингу з урахуванням мовних, культурних та регіональних особливостей.

Таким чином, запропоноване рішення є ефективним інструментом цифрової трансформації рекламної діяльності освітніх закладів і відкриває нові можливості для підвищення їх конкурентоспроможності в умовах цифрової економіки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про рекламу : Закон України від 03.07.1996 № 270/96-ВР. URL: <https://zakon.rada.gov.ua/laws/show/270/96-вр> (дата звернення: 02.12.2025).
2. Куликова І. В. Нейромережі на Python : навч. посіб. Київ : Видавничий дім «Освіта», 2025. 210 с.
3. Alla S., Adari S. K. Beginning MLOps with MLFlow: Deploy Models in AWS SageMaker, Google Cloud, and Microsoft Azure. New York : Apress, 2021. 380 p.
4. Ameisen E. Building Machine Learning Powered Applications. Sebastopol : O'Reilly Media, 2020. 280 p.
5. Babushkin V., Kravchenko A. Machine Learning System Design: With end-to-end examples. New York : Manning Publications, 2025. 376 p.
6. Behrman K. R. Foundational Python for Data Science. Boston : Addison-Wesley, 2022. 244 p.
7. Chaffey D., Ellis-Chadwick F. Digital Marketing: Strategy, Implementation and Practice. 8th ed. London : Pearson, 2022. 744 p.
8. Chollet F. Deep Learning with Python. 2nd ed. New York : Manning Publications, 2021. 504 p.
9. Contextual Advertising in a Cookieless Future [Електронний ресурс] // The Trade Desk. 2023. URL: <https://www.thetradedesk.com> (дата звернення: 02.12.2025).
10. Dwivedi Y. K. et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges. International Journal of Information Management. 2021. Vol. 57. P. 101–115.
11. Gift N., Deza A. Practical MLOps: Operationalizing Machine Learning Models. Sebastopol : O'Reilly Media, 2021. 250 p.

12. GPT Models and Their Application in Digital Advertising [Электронный ресурс] // OpenAI Research. 2023. URL: <https://openai.com/research> (дата звернення: 02.12.2025).
13. Häglund E., Björklund J. AI-driven contextual advertising: Toward relevant messaging without personal data. *Journal of Current Issues & Research in Advertising*. 2024. Vol. 45, No. 3. P. 301–319.
14. Haleem A., Javaid M., Qadri M. A. Artificial intelligence (AI) applications for marketing: A literature-based study. *International Journal of Intelligent Networks*. 2022. Vol. 3. P. 119–132.
15. Heitmann M. Generative AI for Marketing Content Creation: New Rules for an Old Game. *NIM Marketing Intelligence Review*. 2024. Vol. 16, No. 1. P. 10–15.
16. Huyen C. *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*. Sebastopol : O'Reilly Media, 2022. 380 p.
17. Kotler P., Kartajaya H., Setiawan I. *Marketing 5.0: Technology for Humanity*. Hoboken : Wiley, 2021. 224 p.
18. Kushwaha A. K., Kar A. K. Understanding the effects of AI-generated content on consumer engagement. *Journal of Retailing and Consumer Services*. 2024. Vol. 76. P. 103–118.
19. Lakshmanan V., Robinson S., Munn M. *Machine Learning Design Patterns*. Sebastopol : O'Reilly Media, 2021. 400 p.
20. Minh T. N., Jatowt A. Efficient Generative AI for Text Generation: A Survey. *ACM Computing Surveys*. 2024. Vol. 56, Issue 2. P. 1–35.
21. Narayanan P. K. *Data Engineering for Machine Learning Pipelines*. Birmingham : Packt Publishing, 2024. 310 p.
22. Papineni K. et al. BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the ACL*. 2022. P. 311–318.

23. Performance Max campaigns: An Overview [Электронный ресурс] // Google Ads Help. 2024. URL: <https://support.google.com/google-ads> (дата звернення: 02.12.2025).
24. Ping D. The Machine Learning Solutions Architect Handbook. 2nd ed. Birmingham : Packt Publishing, 2024. 602 p.
25. Raschka S., Liu Y., Mirjalili V. Machine Learning with PyTorch and Scikit-Learn. Birmingham : Packt Publishing, 2022. 770 p.
26. Rodrigues R. Legal and Ethical Challenges of Generative AI. Computer Law & Security Review. 2024. Vol. 53. P. 1–18.
27. Rothman D. Transformers for Natural Language Processing. 2nd ed. Birmingham : Packt Publishing, 2024. 506 p.
28. Sanh V. et al. Multitask Prompted Training Enables Zero-Shot Task Generalization. International Conference on Learning Representations (ICLR). 2022.
29. Treveil M. et al. Introducing MLOps: How to Scale Machine Learning in the Enterprise. Sebastopol : O'Reilly Media, 2020. 202 p.
30. Vasilev I. Advanced Deep Learning with Python. Birmingham : Packt Publishing, 2023. 390 p.

ДОДАТКИ

Додаток А

Посилання на код прототипу проекту

https://colab.research.google.com/drive/1_EFMythRoWF8lswTggHenHzl_C7MCrgA?usp=sharing