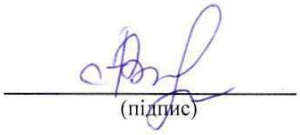
МІНІСТЕРСТВО ФІНАНСІВ УКРАЇНИ
ДЕРЖАВНИЙ ПОДАТКОВИЙ УНІВЕРСИТЕТ

ФАКУЛЬТЕТ ФІНАНСІВ ТА ЦИФРОВИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНИХ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ І СИСТЕМ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: ДОСЛІДЖЕННЯ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ ORACLE MACHINE LEARNING У ЗАДАЧАХ МАРКЕТИНГУ

Здобувача вищої освіти другого (магістерського) рівня,
ІІ курсу, групи КМІз-24-1, спеціальності 122 «Комп'ютерні науки»,
освітня програма «Інформаційні управляючі системи і технології в економіці»


(підпис)

Василенкова Анна Юріївна
(прізвище, ім'я, по батькові)

Науковий керівник


(підпис)

Редич Олександр Володимирович, канд. екон. наук, доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

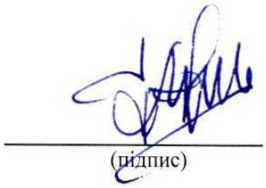
Керівник освітньо-професійної програми


(підпис)

Омельчук Антон Анатолійович, канд. техн. наук
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту за рішенням кафедри протокол від «09» січня 2026 № 11

Завідувач кафедри


(підпис)

Філоненко Михайло Миколайович, канд. фіз.-мат. наук, доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Ірпінь – 2026

Реєстрація № 36 02.02.2026 (дата) (підпис відповідальної особи)

Результати перевірки до захисту (до захисту, на доопрацювання) 09.01.2026 (дата) (підпис)

Результат захисту 4 (добре) (оцінка) 11.02.2026 (дата) (підпис)

Голова комісії: (підпис) О.С. Дюбанов

Члени комісії:

(підпис) А.Ю. Горбовий

(підпис) М.М. Філоненко

(підпис) І.Д. Погорєловська

БЛАНК ОЦІНЮВАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Об'єкти контролю	Діапазон оцінки в балах	Фактична кількість балів	Примітки	Дата, П.І.Б.: керівника, голови комісії членів комісії,секретаря, (підписи)
ОЦІНЮВАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ КОМІСІЄЮ ПО ЗАХИСТУ				
1. Зміст кваліфікаційної роботи (якість теоретичного матеріалу, його зв'язок з практичним, вміння представляти матеріал в логічній послідовності, наявність висновків та пропозицій)	від 0 до 30	23		11.02.2026 Дюбанов О.С.
2. Дотримання вимог щодо оформлення	від 0 до 10	9		Горбовий А.Ю.
3. Вчасне виконання роботи на всіх етапах у встановлені кафедрою терміни	від 0 до 5	3		Філоненко М.М. Погорєловська І.Д.
4. Рецензент	від 0 до 5	3		Суховецька Л.П.
РАЗОМ БАЛІВ	від 0 до 50	38		

Оцінювання кваліфікаційної роботи науковим керівником - 37 балів.

ПЕРЕВЕДЕННЯ ОЦІНКИ ЗА 100-БАЛЬНОЮ ШКАЛОЮ ДО НАЦІОНАЛЬНОЇ ШКАЛИ ТА ШКАЛИ ЄКТС

Оцінка за 100-бальною шкалою	Оцінка за шкалоюЄКТС	Оцінка за національною шкалою	Дата	П.І.Б.: голови, членів комісії, секретаря, (підпис)
75	C	4	11.02.2026	Дюбанов О.С. Горбовий А.Ю. Філоненко М.М. Погорєловська І.Д. Суховецька Л.П.

ПРАВИЛА ПЕРЕВЕДЕННЯ ОЦІНОК ДО НАЦІОНАЛЬНОЇ ШКАЛИ ТА ШКАЛИ ЄКТС

Оцінка за 100-бальною шкалою	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою
90-100	A	5 (відмінно)
80-89	B	4 (добре)
70-79	C	
60-69	D	3 (задовільно)
50-59	E	
35-49	FX	2 (незадовільно)

Факультет фінансів та цифрових технологій
Кафедра комп'ютерних та інформаційних технологій і систем
Галузь знань 12 «Інформаційні технології»
Спеціальність 122 «Комп'ютерні науки»
Освітня програма «Інформаційні управляючі системи і технології в економіці»

Затверджую
Завідувач кафедри _____ Філоненко М.М.
«11» грудня 2024

Завдання на кваліфікаційну роботу
здобувачу вищої освіти
Василенковій Анні Юріївні

1. Тема кваліфікаційної роботи:
Дослідження класифікаційних моделей Oracle Machine Learning у задачах маркетингу
Затверджена наказом ректора від «10» грудня 2024 № 1852
2. Строк здачі закінченої роботи «09» січня 2026
3. Мета роботи: *вивчення можливостей, переваг і обмежень класифікаційних алгоритмів Oracle Machine Learning у застосуванні для аналізу маркетингових даних і оптимізації маркетингових стратегій.*
Об'єкт дослідження: *класифікаційні моделі у маркетинговій діяльності.*
Предмет дослідження: *інструментарій класифікаційних моделей Oracle Machine Learning.*
Орієнтовна база дослідження: *інформаційні процеси у маркетинговій діяльності.*
4. Вихідні дані до роботи: *нормативно-інформаційною основою дослідження є офіційні матеріали корпорації Oracle, документація з Oracle Database та Oracle Machine Learning.*
5. Перелік графічного матеріалу: *схеми алгоритмів обробки даних, математичні алгоритми машинного навчання, формули, візуалізації та скріншоти інтерфейсу.*
6. Зміст кваліфікаційної роботи (перелік питань за кожним розділом) :
РОЗДІЛ 1. Теоретико-методологічні основи застосування класифікаційних моделей у маркетингу
1.1. Класифікаційні задачі та їх роль у сучасному маркетингу
1.2. Огляд класифікаційних моделей машинного навчання
1.3. Технологічна платформа Oracle Machine Learning (OML) та її інструментарій
РОЗДІЛ 2. Проєктування та підготовка даних для класифікаційного моделювання
2.1. Визначення цільової маркетингової задачі та формування набору даних
2.2. Етапи очищення, трансформації та збагачення маркетингових даних
2.3. Розділення даних та вибір контрольних параметрів моделювання

РОЗДІЛ 3. Дослідження та оцінка ефективності класифікаційних моделей Oracle Machine Learning
3.1. Реалізація класифікаційних моделей у середовищі OML
3.2. Порівняльний аналіз результатів моделювання та оцінка ефективності

7. Календарний план виконання роботи

№ п/п	Назва етапів кваліфікаційноїроботи	Строк виконання етапів роботи	
		За планом	Фактично
1	2	3	4
1	Вибір теми, підбір літературних джерел, формування структури кваліфікаційної роботи	До 15.12.2024	15.12.2024
2	**І розділ** (розкрити сутність і стан дослідження наукової проблематики роботи, після чого навести її загальну характеристику; огляд літератури за темою і вибір конкретних напрямів досліджень)	26.05.2025	26.05.2025
3	**ІІ розділ** (аналіз експериментальної частини про методики досліджень)	02.09.2025	01.09.2025
4	**ІІІ розділ** (проведені теоретичні і (або) експериментальні дослідження та проектні рішення)	24.11.2025	25.11.2025
5	**Висновки**	18.12.2025	18.12.2025
6	Доопрацювання роботи з урахуванням зауважень наукового керівника	02.01.2025	02.01.2025
7	Отримання відгуку керівника	02.02.2026	02.02.2026
8	Отримання зовнішньої рецензії	05.02.2026	05.02.2026

8. Дата видачі завдання «11» грудня 2024

9. Керівник кваліфікаційної роботи _____ (підпис) Редич О.В (прізвище, ініціали)

10. Керівник освітньо-професійної програми _____ (підпис) Омельчук А.А. (прізвище, ініціали)

11. Завдання до виконання отримав здобувач вищої освіти _____ (підпис) Василенкова А.Ю (прізвище, ініціали)

АНОТАЦІЯ

«Дослідження класифікаційних моделей Oracle Machine Learning у задачах маркетингу»

тема кваліфікаційної роботи

Мета роботи полягає у вивченні можливостей, переваг і обмежень класифікаційних алгоритмів Oracle Machine Learning та визначенні доцільності їх застосування для аналізу маркетингових даних і оптимізації маркетингових стратегій.

Завданнями роботи є охарактеризувати класифікаційні задачі та їх роль у сучасному маркетингу; провести огляд класифікаційних моделей машинного навчання; вивчити специфіку технологічної платформи Oracle Machine Learning (OML) та її інструментарій; розкрити особливості цільової маркетингової задачі та формування набору даних; проаналізувати дані та вибір контрольних параметрів моделювання; здійснити порівняльний аналіз результатів моделювання та оцінка ефективності.

Предметом дослідження є класифікаційні моделі Oracle Machine Learning у задачах маркетингу.

Об'єктом дослідження виступає оцінка ефективності класифікаційних моделей Oracle Machine Learning.

Базою дослідження виступає програмні та технічні засоби Oracle Database з підтримкою Oracle Machine Learning; SQL та/або Python (OML4SQL, OML4Py); середовище Oracle SQL Developer або Oracle Autonomous Database.

У ході виконання даного дослідження було проаналізовано можливості використання класифікаційних моделей Oracle Machine Learning для розв'язання маркетингових задач, зокрема прогнозування реакції клієнтів на маркетингові кампанії. Розглянуто основні підходи до побудови моделей машинного навчання в середовищі Oracle Database та досліджено їхню ефективність на основі маркетингових даних. За результатами порівняльного аналізу встановлено, що найбільш ефективні результати у задачі прогнозування відгуку клієнтів продемонстрували ансамблеві методи, зокрема модель Random Forest, яка забезпечила найвищі значення інтегральних показників якості. Водночас логістична регресія показала стабільні та інтерпретовані результати, що є важливим фактором для практичного застосування в маркетинговій аналітиці.

Одержані результати можуть бути використані як при написанні тез, доповідей на науково-практичних конференціях, так і застосовані при проведення лекційних і семінарських занять із студентами профільних факультетів.

Кваліфікаційна робота містить 58 сторінки, 3 рисунки, список використаних джерел з 34 найменування.

Ключові слова: Oracle Machine Learning, машинне навчання, класифікаційні моделі, маркетингова аналітика, аналіз даних, прогнозування поведінки клієнтів, бінарна класифікація.

Рік виконання кваліфікаційної роботи: 2025.

Рік захисту роботи: 2026.

ABSTRACT

«Research on Oracle Machine Learning classification models in marketing tasks»

The purpose of this work is to study the capabilities, advantages, and limitations of Oracle Machine Learning classification algorithms and to determine the feasibility of their application for marketing data analysis and marketing strategy optimization.

The objectives of the thesis are to characterize classification tasks and their role in modern marketing; to review machine learning classification models; to study the specifics of the Oracle Machine Learning (OML) technology platform and its tools; to reveal the features of the target marketing task and the formation of a data set; to analyze data and select control parameters for modeling; to perform a comparative analysis of modeling results and evaluate effectiveness.

The subject of the study is Oracle Machine Learning classification models in marketing tasks.

The object of the study is to evaluate the effectiveness of Oracle Machine Learning classification models.

The study is based on Oracle Database software and hardware with Oracle Machine Learning support; SQL and/or Python (OML4SQL, OML4Py); Oracle SQL Developer or Oracle Autonomous Database environment.

During this study, the possibilities of using Oracle Machine Learning classification models to solve marketing problems, in particular predicting customer responses to marketing campaigns, were analyzed. The main approaches to building machine learning models in the Oracle Database environment were considered and their effectiveness was investigated based on marketing data. The comparative analysis showed that ensemble methods, in particular the Random Forest model, demonstrated the most effective results in predicting customer response, providing the highest values of integral quality indicators. At the same time, logistic regression showed stable and interpretable results, which is an important factor for practical application in marketing analytics.

The results obtained can be used both in writing abstracts and reports at scientific and practical conferences, and in conducting lectures and seminars with students of specialized faculties.

The thesis consists of 58 pages, 3 figures, and a list of 34 references.

Keywords: Oracle Machine Learning, machine learning, classification models, marketing analytics, data analysis, customer behavior forecasting, binary classification.

Year of qualification work completion: 2025.

Year of thesis defense: 2026.

ЗМІСТ:

ВСТУП……..……..3

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ ЗАСТОСУВАННЯ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ У МАРКЕТИНГУ…… 8

1.1. Класифікаційні задачі та їх роль у сучасному маркетингу……..8

1.2. Огляд класифікаційних моделей машинного навчання……..…11

1.3. Технологічна платформа Oracle Machine Learning (OML) та її інструментарій……...16

РОЗДІЛ 2. ПРОЄКТУВАННЯ ТА ПІДГОТОВКА ДАНИХ ДЛЯ КЛАСИФІКАЦІЙНОГО МОДЕЛЮВАННЯ……..…… 21

2.1. Визначення цільової маркетингової задачі та формування набору даних……..21

2.2. Етапи очищення, трансформації та збагачення маркетингових даних……...……27

2.3. Розділення даних та вибір контрольних параметрів моделювання…...…29

РОЗДІЛ 3. ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ ORACLE MACHINE LEARNING 33

3.1. Реалізація класифікаційних моделей у середовищі OML …….……33

3.2. Порівняльний аналіз результатів моделювання та оцінка ефективності……...43

ВИСНОВОК ……...……..51

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ…….53

ДОДАТОК 1……..……58

ВСТУП

Актуальність теми. Сучасний маркетинг функціонує в умовах стрімкого зростання обсягів даних, різноманіття каналів взаємодії зі споживачами та підвищених вимог до точності управлінських рішень. Традиційні підходи до аналізу ринку дедалі частіше виявляються недостатніми для глибокого розуміння поведінки клієнтів, прогнозування їхніх потреб і формування персоналізованих пропозицій. У цьому контексті особливої актуальності набувають методи машинного навчання, які дозволяють автоматизувати процеси аналізу даних і виявляти приховані закономірності, недоступні для класичних статистичних методів. Одним із потужних інструментів для впровадження аналітики та машинного навчання в корпоративному середовищі є Oracle Machine Learning – платформа, що поєднує можливості роботи з великими масивами даних безпосередньо в базі даних Oracle та широкий спектр алгоритмів інтелектуального аналізу. Особливе місце серед цих алгоритмів посідають класифікаційні моделі, які широко застосовуються у маркетингу для вирішення таких завдань, як сегментація клієнтів, прогнозування відтоку, оцінка ймовірності відгуку на маркетингові кампанії, визначення кредитоспроможності та формування рекомендацій.

Дослідження класифікаційних моделей Oracle Machine Learning у маркетингових задачах є важливим як з наукової, так і з практичної точки зору. З одного боку, воно дає змогу проаналізувати ефективність різних алгоритмів класифікації – логістичної регресії, дерев рішень, наївного байєсівського класифікатора, методів ансамблю – з урахуванням специфіки маркетингових даних. З іншого боку, таке дослідження сприяє обґрунтованому вибору моделей для підтримки управлінських рішень у бізнесі, підвищенню точності прогнозів і оптимізації маркетингових стратегій.

Стан і ступінь розробки проблеми в спеціальній літературі. Ключові теоретичні засади стосовно машинного навчання, а також основні моделі

структуровано описано у дослідженнях Білоус А. [2], Бондар В. О., Кравець О. І. [3]. Відносно поточного стану та перспектив розвитку машинного навчання ведуть мову Бортник К. Я., Ольшевський О. В., Кирилюк А. Л. [4]. До питання можливостей застосування штучного інтелекту і машинного навчання у маркетингу висловлюється Волнянський А. [5], Мокін В. Б., Дратований М. В. [14]. Практичні кейси а також реальні механізми імплементації розкриті у наукових працях Федько В. В. [22].

Мета і завдання дослідження. Мета дослідження полягає у вивченні можливостей, переваг і обмежень класифікаційних алгоритмів Oracle Machine Learning та визначенні доцільності їх застосування для аналізу маркетингових даних і оптимізації маркетингових стратегій. Для реалізації окресленої мети доцільно передбачити наступні **завдання**, які включають в себе:

1. охарактеризувати класифікаційні задачі та їх роль у сучасному маркетингу;
2. провести огляд класифікаційних моделей машинного навчання;
3. вивчити специфіку технологічної платформи Oracle Machine Learning (OML) та її інструментарій;
4. розкрити особливості цільової маркетингової задачі та формування набору даних;
5. проаналізувати дані та вибір контрольних параметрів моделювання;
6. здійснити порівняльний аналіз результатів моделювання та оцінка ефективності;

Об’єкт дослідження – класифікаційні моделі у маркетинговій діяльності

Предмет дослідження – інструментарій класифікаційних моделей Oracle Machine Learning.

Методологічною основою дослідження є системний підхід, який передбачає розгляд маркетингової діяльності підприємства як цілісної системи, де дані, аналітичні інструменти та управлінські рішення перебувають у тісному взаємозв’язку. У межах цього підходу класифікаційні моделі Oracle

Machine Learning розглядаються не як ізольований технічний інструмент, а як складова інформаційно-аналітичного забезпечення маркетингу, що підтримує стратегічне та оперативне управління взаємовідносинами з клієнтами. Системний підхід дозволяє враховувати вплив зовнішніх і внутрішніх факторів ринку, особливості поведінки споживачів та специфіку бізнес-процесів підприємства. Аналітичну складову методології формують методи економічного та статистичного аналізу, які застосовуються на всіх етапах дослідження. До них належать методи описової статистики, аналізу розподілів, кореляційного аналізу та виявлення закономірностей у маркетингових даних. Зазначені методи використовуються для попереднього аналізу інформації, оцінки якості даних і вибору релевантних ознак для класифікаційного моделювання. Окреме місце в методології посідають методи машинного навчання, зокрема алгоритми класифікації, реалізовані в Oracle Machine Learning, які дозволяють формалізувати маркетингові задачі у вигляді прогнозних моделей.

Практико-орієнтовану частину методології становлять методи інтелектуального аналізу даних та експериментального моделювання. Для кожного алгоритму визначаються контрольні параметри, здійснюється навчання моделей на підготовлених вибірках і проводиться їх тестування. Порівняльний аналіз результатів дозволяє оцінити переваги та обмеження кожного підходу в контексті конкретних маркетингових задач, таких як прогнозування відтоку клієнтів або оцінка ймовірності відгуку на маркетингову кампанію. Важливою складовою методології є використання методів валідації та оцінювання якості моделей. У дослідженні застосовуються підходи до розділення даних на навчальні та тестові вибірки, а також методи оцінювання ефективності класифікації за допомогою показників точності, повноти, F-міри та площі під ROC-кривою. Оцінювання результатів моделювання здійснюється не лише з точки зору статистичних метрик, а й з позицій маркетингової доцільності, зокрема потенційного впливу

моделей на підвищення ефективності маркетингових кампаній і оптимізацію витрат підприємства.

Базу дослідження становлять як теоретичні, так і практичні джерела інформації. Теоретичною базою є праці вітчизняних і зарубіжних науковців у галузі маркетингу, управління даними, економічного аналізу та машинного навчання, а також концепції клієнтоорієнтованого маркетингу й аналітики даних. Нормативно-інформаційну основу формують офіційні матеріали корпорації Oracle, зокрема документація з Oracle Database та Oracle Machine Learning, методичні рекомендації щодо використання інструментів інтелектуального аналізу даних і стандарти роботи з великими масивами інформації.

Наукова та практична значимості результатів. Наукову цінність також становить розроблена та обґрунтована методологія підготовки маркетингових даних і побудови класифікаційних моделей у середовищі Oracle Machine Learning. У межах дослідження узагальнено підходи до очищення, трансформації, збагачення та розділення даних, а також до вибору контрольних параметрів моделювання з урахуванням специфіки маркетингових задач. Отримані результати можуть бути використані як теоретичне підґрунтя для подальших наукових досліджень у сфері маркетингової аналітики, цифрового маркетингу та управління даними.

Практична значимість результатів дослідження полягає у можливості безпосереднього застосування отриманих висновків і розроблених підходів у діяльності підприємств різних галузей. Запропонована практична модель використання Oracle Machine Learning дозволяє підприємствам підвищити точність прогнозування поведінки клієнтів, зокрема ризику відтоку або ймовірності відгуку на маркетингові кампанії. Це створює передумови для більш ефективного використання маркетингових бюджетів, зниження витрат на залучення клієнтів і підвищення рівня їх лояльності.

Структура роботи. Кваліфікаційна робота складається із вступу, 3 розділів, висновку та списку використаних джерел (включає в себе 34 найменування). Загальна кількість сторінок – 58.

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ ЗАСТОСУВАННЯ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ У МАРКЕТИНГУ

1.1. Класифікаційні задачі та їх роль у сучасному маркетингу

Маркетингова діяльність сучасних компаній є одним з головних напрямків того, щоб забезпечити економічну ефективність, безпеку, конкурентоспроможність і фінансову стійкість стратегічного розвитку бізнесу. Однак під впливом цифровізації та створенням нових інновацій формується все частіше використання різних технологій у вдосконаленні маркетингу, серед яких технології штучного інтелекту – машинне навчання [1, с. 71]. Популярність впровадження та застосування цифрових технологій та інновацій у маркетингу з метою забезпечення вдосконалення маркетингової діяльності компанії обумовлена впливом таких факторів:

- по-перше, необхідністю проведення стратегічних змін при маркетинговому просуванні бренду і продукції компанії;
- по-друге, необхідністю адаптації під цифрову трансформацію бізнесу, включаючи цифровізацію всіх бізнес-процесів і в цілому бізнес-моделі;
- по-третє, необхідністю адаптації до змін стосовно реалізації продукції в умовах війни, адже традиційні способи продажу тривалий період часу були закриті.

Через карантинні обмеження та загрози зі сторони країни-агресора відбулося закриття багатьох офісів і магазинів. Велика кількість людей перейшла в мережу Інтернет, де здійснюється купівля необхідних товарів і послуг [19, с. 233].

Основним способом маркетингового просування продукції є традиційна реклама в засобах масової інформації та телебаченні. Вона дозволяє охоплювати найширшу аудиторію потенційних споживачів, проте її якість є незадовільною. Охоплення цільової аудиторії стає все нижчим, адже в

традиційних джерелах реклами відсутні можливості персоналізації рекламної пропозиції для тих клієнтів, які можуть бути потенційними покупцями [9].

Використання інструментів інноваційного маркетингу виводить таргетування на новий рівень ефективності та дозволяє уникнути взаємодії з незацікавленою аудиторією. Так, цифрові технології маркетингового просування дозволяють орієнтуватися не тільки на демографічну ознаку споживача, але і на його інтереси та сферу діяльності, що дозволяє з більшою точністю персоналізувати рекламні пропозиції для споживачів [17, с. 59-60].

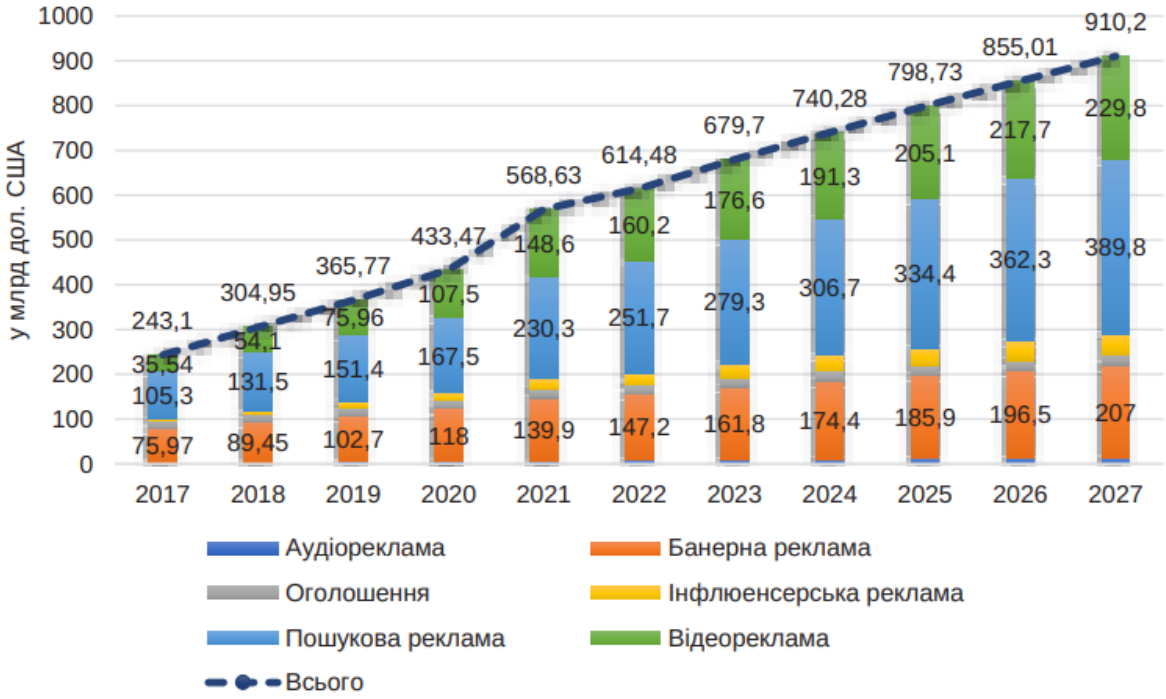


Рис. 1. Загальносвітові витрати на Інтернет-маркетинг за видами реклами [13, с. 4]

У маркетингу цифрові та інноваційні технології застосовуються з метою координації зусиль підприємства щодо досягнення стратегічних і оперативних орієнтирів з урахуванням ринкових вимог та забезпечення взаємодії з клієнтами і бізнес-партнерами за допомогою передачі інформації про товари і послуги, а також отримання, обробки та аналізу маркетингової інформації [12, с. 39]. На нашу думку, сучасним компаніям економіки України необхідно проводити плавний перехід від мультиканального до омніканального маркетингу, сутністю якого є можливість поєднати безліч каналів комунікацій,

що використовуються споживачами. Актуальність застосування омніканального маркетингу обумовлена необхідністю адаптації маркетингової стратегії та рекламної діяльності до нових умов зовнішнього середовища.

Побудова систем машинного навчання є на сьогоднішній день однією з найпопулярніших, актуальних і сучасних областей цифровізації підприємницької діяльності в Україні на стику інформаційних технологій, математичного аналізу і статистики. Машинне навчання все глибше проникає в бізнес-процеси українських організацій за допомогою користувацьких продуктів, створених при участі штучного інтелекту. Очевидно, що ці технології будуть розвиватися і далі, поступово стаючи частиною операційної діяльності будь-якого бізнесу. Крім того, застосування технології машинного навчання сприяє вдосконаленню маркетингу, що включає в себе наступний ряд переваг [5]:

1. Збільшується швидкість обробки інформації, її аналітика та прийняття рішень при обробці результатів.
2. Максимізуються можливості клієнтського обслуговування в режимі безперервної роботи.
3. З'являється фундамент для застосування та розвитку інших цифрових технологій та інформаційних систем, де машинне навчання збільшує ефективність.
4. Персоналізується реклама та маркетингова пропозиція окремим групам цільової аудиторії.
5. Відбувається оптимізація фінансових і трудових ресурсів, оскільки технології машинного навчання вимагають мінімальної кількості обслуговуючого персоналу.

Класифікаційні задачі в машинному навчанні полягають у побудові моделей, здатних визначати, до якого класу належить той чи інший об’єкт. Наприклад, система може класифікувати електронні листи як «спам» або «не спам», клієнтів – як «лояльних» або «схильних до відтоку», а відгуки – як

«позитивні», «негативні» чи «нейтральні». Для цього використовуються різні алгоритми, зокрема логістична регресія, дерева рішень, метод опорних векторів, наївний баєсівський класифікатор та нейронні мережі. У маркетингу класифікаційні моделі відіграють надзвичайно важливу роль.

По-перше, вони застосовуються для сегментації клієнтів. На основі поведінкових, демографічних та транзакційних даних компанії можуть поділяти аудиторію на групи з подібними характеристиками, що рівноцінно дає змогу створювати персоналізовані маркетингові кампанії, підвищуючи їхню релевантність та ефективність.

По-друге, класифікаційні задачі широко використовуються для прогнозування відтоку клієнтів (churn prediction). Модель аналізує історію взаємодії клієнта з брендом і визначає ймовірність того, що він припинить користуватися продуктом або послугою. Завдяки цьому маркетологи можуть завчасно вжити заходів для утримання клієнтів, наприклад запропонувати спеціальні знижки чи персональні пропозиції. Не менш важливим напрямом є аналіз настроїв (sentiment analysis). За допомогою класифікації текстів відгуки в соціальних мережах, коментарі та оцінки клієнтів автоматично визначаються як позитивні або негативні, що дозволяє компаніям швидко реагувати на проблеми, покращувати якість продуктів і формувати позитивний імідж бренду.

Крім того, класифікаційні моделі сприяють у визначенні ефективності рекламних кампаній. Наприклад, користувачів можна класифікувати за ймовірністю кліку на рекламу або здійснення покупки. Такий підхід оптимізує витрати на рекламу та підвищує рентабельність маркетингових інвестицій.

1.2. Огляд класифікаційних моделей машинного навчання

Машинне навчання, завдяки адаптивному характеру, використовується в сферах, в яких дані постійно змінюються, поставлені завдання нестабільні

або написати програмний код практично неможливо. Процес використання машинного навчання складається з чотирьох етапів:

1. Збір і підготовка даних. Тип використовуваних даних допоможе визначити, які алгоритми машинного навчання можна застосовувати. При підготовці даних розробляється структура і усуваються проблеми з цілісністю даних.

2. Навчання моделі. Підготовлені дані поділяються на дві групи: набір для навчання та набір для перевірки. Набір для навчання – це більша частина даних, за допомогою яких моделі машинного навчання налаштовуються з максимальною точністю.

3. Перевірка моделі. При виборі кінцевої моделі даних продуктивність і точність оцінюється за допомогою набору для перевірки.

4. Інтерпретація результатів. Результати вивчаються, щоб отримати відомості для аналізу і спрогнозувати результати.

Всі методи машинного навчання можна класифікувати власне за способом навчання [15]. Основними групами методів є навчання з вчителем (методи класифікації) і без вчителя (методи кластерного аналізу). У разі навчання з вчителем для кожного об'єкта з навчальної вибірки задається пара «прецедент-рішення», і завдання зводиться до визначення класу об'єкта за його ознаками. При навчанні без вчителя навчальна вибірка не містить заздалегідь відомих відповідей для набору прецедентів, і від моделі вимагається самостійно об'єднати дані в кластер, що узагальнює ці дані.

До моделей класифікації відносяться дерева рішень, нейронні мережі, вирішальні правила, алгоритми покриття та інші [21, с. 98-99]. Найбільш наочною і легко інтерпретованою є модель на основі вирішального дерева. Методи кластеризації поділяються на ієрархічні та неієрархічні. Методи ієрархічної кластеризації можуть бути агломеративними, що передбачають послідовне об'єднання вихідних об'єктів окремих кластерів і зменшення числа кластерів, і дивізійними, в яких на початку роботи алгоритму всі об'єкти належать до одного кластера. Методи неієрархічної кластеризації є

ітеративними методами поділу вихідної вибірки, при яких нові кластери формуються до моменту виконання правила зупинки.

Алгоритм роботи будь-якої моделі машинного навчання можна розбити на такі етапи [24]:

1. Збір даних. На цьому етапі збираються матеріали для навчання і розміщуються в єдиному джерелі – електронній таблиці, текстовому файлі або базі даних.
2. Дослідження та підготовка даних. Етап передбачає виправлення або очищення забруднених даних, видалення непотрібної інформації та перекодування відомостей відповідно до завдань, які ставляться перед моделлю.
3. Навчання моделі. Алгоритм навчання підбирається з урахуванням специфіки даних аналізу та завдання машинного навчання.
4. Оцінка моделі. Оскільки будь-яка модель машинного навчання вирішує завдання навчання упереджено, після отримання результату важливо оцінити, наскільки добре алгоритм враховує власний досвід. Методи оцінки точності залежать від типу використовуваної моделі: для одних моделей підійдуть тестові набори даних, для інших потрібний підбір специфічних для передбачуваної області використання показників продуктивності.
6. Поліпшення моделі. Для підвищення продуктивності потрібно використання поліпшених стратегій, доповнення або додаткова підготовка даних. У ряді випадків потрібна зміна моделі машинного навчання.

Результат, який видають моделі машинного навчання, залежить не тільки від вхідного набору даних, але і від конфігурації: при різній конфігурації, тобто гіперпараметрах алгоритму, на одному наборі даних буде виходити різний результат [2]. На практиці вибір алгоритму і налаштування гіперпараметрів часто виконується вручну, що вимагає значних трудових і часових витрат. Однак в умовах збільшення обчислювальних потужностей

комп'ютерів розвиток отримують системи автоматизованого машинного навчання – AutoML, які допомагають автоматизувати основні процеси навчання зі збереженням прогностичної точності цільової моделі [14]:

- підготовка даних – інтеграція, перетворення, очищення і скорочення даних;
- вилучення і вибір ознак – вибір підмножини ознак з набору даних, що покращує узагальнення навчання;
- вибір алгоритму, що забезпечує найбільш точні результати;
- оптимізація гіперпараметрів – налаштування параметрів алгоритмів для подальшого поліпшення результатів.

Системи використовують різні методи і прийоми оптимізації гіперпараметрів для досягнення бажаної точності і продуктивності моделі. Класифікаційні моделі схожі на регресійні, проте прогноз вони дають у вигляді поділу за будь-якими заздалегідь заданими ознаками. Як у розвивальних іграх для дітей, в яких потрібно помістити різні геометричні фігури у відповідні отвори. Це найпопулярніші моделі. Наприклад, за допомогою них працюють спам-фільтри, автоматична модерація контенту, розпізнавання мов і рукописного тексту. Класифікаційні моделі також можуть бути корисні для виявлення аномалій, тобто об'єктів, які не підходять під жоден заданий критерій.

Найпопулярніші алгоритми класифікаційних моделей включають:

Naive Bayes – заснований на теоремі Байєса і використовується, коли задані ознаки незалежні одна від одної. Наприклад, у завданнях класифікації текстів і фільтрації спаму, коли достатньо сортування за будь-якими словами.

Decision Trees – алгоритм будується у вигляді дерева, де кожна «гілка» представляє собою умову за однією ознакою. Наприклад, для видачі банком кредиту клієнт повинен відповідати відразу за кількома ознаками: кредитна історія, розмір зарплати, наявність фінансових зобов'язань та інше [31]. Для наочної демонстрації ми пропонуємо звернути увагу на Додаток 1, де видно, що кожен внутрішній вузол дерева рішень представляє деяку властивість,

наприклад, борг або дохід. Кожному можливому значенню цієї властивості відповідає гілка дерева. Вузли-листя відображають результати класифікації, зокрема, низький або середній ризик. За допомогою цього дерева можна класифікувати клієнта, тип якого невідомий: для кожного внутрішнього вузла перевіряється значення відповідної властивості для даного клієнта і здійснюється перехід по відповідній гілці [20, с. 450]. Процес завершується досягши кінцевого вузла, що визначає клас об’єкта.

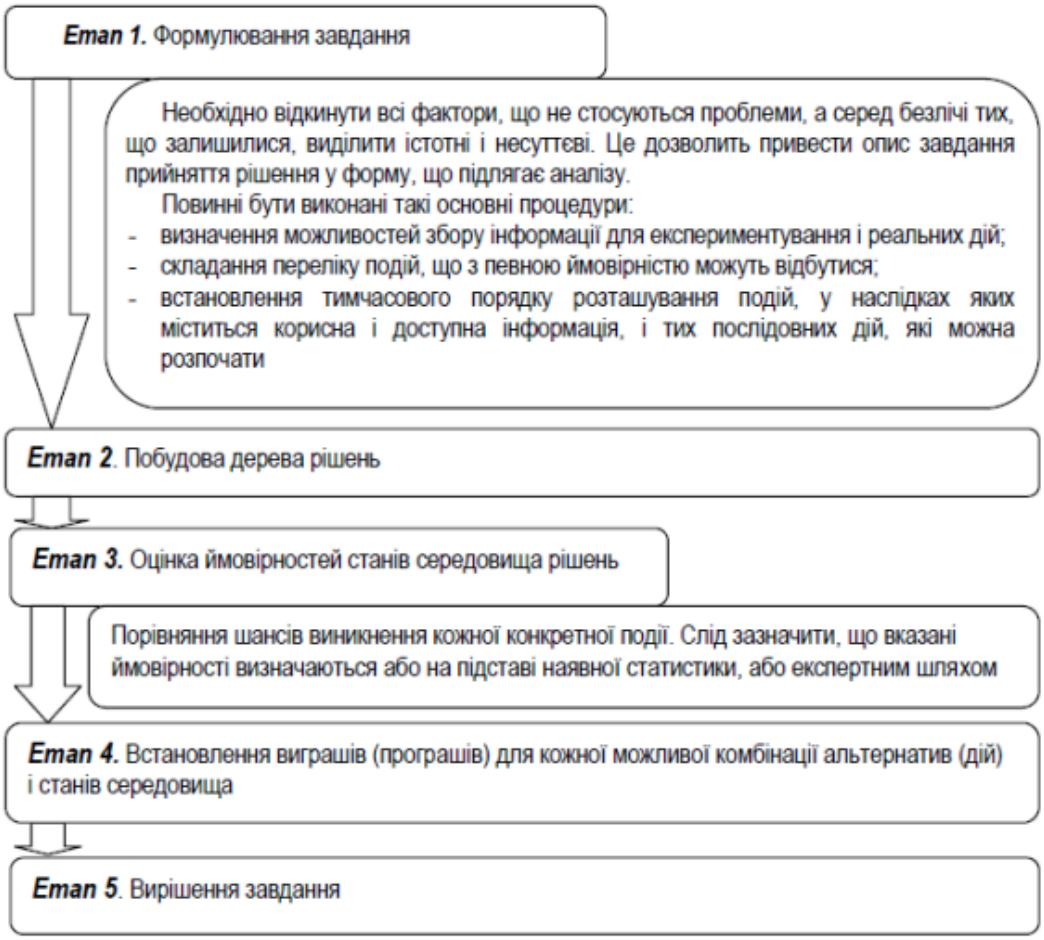


Рис. 2. Основні етапи прийняття рішення за допомогою методу побудови дерева рішень (сформовано автором)

Support Vector Machines – використовується для бінарної класифікації, коли між двома типами ознак проводиться вектор, що визначає найбільший зазор. Далі за допомогою допоміжних векторів визначається, яких ознак у об'єкта найбільше. Такі моделі часто використовують для класифікації тварин і рослин.

K-Nearest Neighbors – містить в основі теорему компактності, тобто об'єкту присвоюється ознака, яка переважає у схожих об'єктах. Таким чином, наприклад, можна класифікувати товари за популярністю.

AdaBoost – алгоритм спрямований на поступове вдосконалення класифікації. Кожен наступний класифікатор будується на основі об'єктів, які слабо задовольняють попередньому, тобто прагне підвищити точність його передбачень. На основі AdaBoost працює, наприклад, розпізнавання облич.

Таким чином, із вищевказаного виходить, що модель машинного навчання – це алгоритм, параметри якого були визначені або задані при допомозі алгоритмічного статистичного процесу навчання. На практиці модель – це комп'ютерна програма, яка може виконувати певні типи завдань без явних вказівок людини [32]. У традиційному програмуванні розробник (людина) пише програму, визначаючи набір правил, яких комп'ютер повинен дотримуватися під час виконання програми, тоді як в машинному навчанні практик навчає модель, даючи комп'ютеру команду виконати алгоритм навчання над деякими даними, доки алгоритм навчання за допомогою

комп'ютера (машини) не створить набір правил, що відповідають певним критеріям.

1.3. Технологічна платформа Oracle Machine Learning (OML) та її інструментарій

Штучний інтелект існує в тій чи іншій формі вже понад 50 років. Однак в останні роки штучний інтелект здійснив значний прорив завдяки розвитку обчислювальної потужності, доступності даних і новим алгоритмам. Він сьогодні присутній практично скрізь – може використовуватися в робототехніці та аналізі великих даних, а також у наших широко розповсюджених цифрових помічниках, таких як голосовий пошук у додатку нашого смартфона [4, с. 86-87]. Крім того, він успішно зарекомендував себе практично у всіх сферах, де його використовували, включаючи охорону здоров'я, банківську справу, освіту та виробництво.

Машинне навчання (Machine Learning) – це пікатегорія штучного інтелекту, яка фокусується на тому, щоб дати комп'ютерам можливість вчитися без явного програмування. Алгоритми машинного навчання аналізують дані, виявляють у них закономірності та використовують ці закономірності для прогнозування та прийняття рішень [26].

Однією з провідних корпоративних платформ у цій галузі є Oracle Machine Learning (OML) – комплексна технологічна платформа, інтегрована безпосередньо в екосистему баз даних Oracle. Її головною особливістю є можливість виконання аналітики та машинного навчання без переміщення даних за межі бази, що суттєво підвищує безпеку, продуктивність і масштабованість. Oracle Machine Learning орієнтована на принцип «in-database machine learning», який передбачає виконання обчислень безпосередньо в СУБД Oracle [29]. Завдяки цьому мінімізуються затрати на передачу даних, знижується ризик витоків інформації та забезпечується робота з актуальними,

узгодженими наборами даних. Такий підхід особливо важливий для великих організацій, що оперують критично важливою інформацією в реальному часі.

Інструментарій OML є багатокомпонентним і охоплює кілька рівнів взаємодії з даними. Основу платформи становить Oracle Machine Learning Algorithms: набір вбудованих алгоритмів машинного навчання, що включає методи класифікації, регресії, кластеризації, асоціативного аналізу та виявлення аномалій. Серед них – Decision Trees, Naive Bayes, Support Vector Machines, k-Means, а також сучасні ансамблеві методи [28]. Алгоритми оптимізовані для роботи з великими обсягами структурованих даних і підтримують автоматичний підбір параметрів.

Як провідний постачальник бізнес-додатків і хмарних сервісів Oracle пропонує AI-рішення, які дозволяють замовникам отримати результати для широкого спектру бізнес-процесів. Серед них:

OCI Language: дає змогу реалізувати практики текстової аналітики для глибокого розуміння неструктурованого тексту в документах, зворотного зв'язку від клієнтів, звернень до служби підтримки, постів у соціальних ЗМІ. Завдяки наявності вже підготовлених моделей OCI Language виключає потребу в наявності експертів з машинного навчання, пропонуючи розробникам вбудовувати в свої рішення інструменти для аналізу смислів, вилучення ключових фраз, класифікації текстів тощо.

OCI Speech: забезпечує можливість автоматичного розпізнавання мови за рахунок вже вбудованих моделей, навчених на тисячах зразків розмов носіїв мов та іноземців. Модуль дозволяє розпізнавати мову в реальному часі і перетворювати записи аудіо в точний текст, відкриваючи можливості для пошуку розділів, індексації контенту і розширеної аналітики аудіо та відеозаписів.

OCI Vision: попередньо навчена система комп'ютерного зору використовується для розпізнавання зображень і аналізу документів. Інструмент дозволяє розширити можливості рішення для будь-якої галузі або провести навчання для специфічних кейсів, таких як моніторинг певних зон,

виявлення дефектів і обробка складних документів. OCI Vision можна використовувати для виявлення візуальних аномалій у виробництві, вилучення текстів з форм, підрахунку товарів і автоматизації бізнес-процесів.

OCI Anomaly Detection: пропонує готову платформу для створення специфічних моделей виявлення аномалій на ранніх етапах. Подібний підхід дозволяє швидше вирішувати проблеми, що виникають, знижуючи їх негативний вплив на операційні процеси. OCI Anomaly Detection має інтерфейси REST API і набори SDK для декількох мов програмування, які відкривають двері розробникам для подальшої інтеграції модулів задля виявлення аномалій у свої бізнес-додатки. Модуль працює на базі запатентованого алгоритму MSET2, який використовується в усьому світі для аналітики складних процесів, таких як стан ядерних реакторів.

OCI Data Labeling: допомагає користувачам проводити розмітку даних для подальшого навчання AI-моделей. Користувачі можуть збирати дані, створювати та переглядати дата-сети, а також наносити мітки на записи через користувацький інтерфейс або за допомогою спеціальних API. Розмічені набори даних можна експортувати і використовувати для розробки моделей в будь-яких сервісах Oracle AI, включаючи OCI Vision або, наприклад, OCI Data Science.

Нові сервіси OCI AI є основою сімейства рішень Oracle AI, набору готових продуктів на базі штучного інтелекту, машинного навчання і Data Science. До їх числа входять Oracle Digital Assistant, а також OCI Data Science і Oracle Database Machine Learning — рішення, які використовуються для створення, управління і впровадження моделей машинного навчання. Безліч організацій по всьому світу покладаються на Oracle AI для автоматизації операційної діяльності, впровадження інновацій і прийняття більш зважених стратегічних рішень.

Важливим елементом платформи є OML4SQL, OML4Python та OML4R, які дозволяють використовувати можливості машинного навчання через знайомі мови програмування. Аналітики можуть створювати, навчати та

застосовувати моделі, використовуючи SQL-запити, Python або R, при цьому обчислення виконуються на сервері бази даних [33]. Така інтеграція поєднує зручність сучасних мов аналізу даних із потужністю корпоративної СУБД.

Для візуальної роботи та спільної аналітики Oracle пропонує Oracle Machine Learning Notebooks – вебсередовище, подібне до Jupyter Notebook. Воно забезпечує інтерактивну роботу з даними, побудову моделей, візуалізацію результатів та документування досліджень [34]. Такий підхід сприяє командній роботі між аналітиками, дата-сайєнтистами та бізнес-користувачами. Окрему роль відіграють інструменти управління життєвим циклом моделей, які дозволяють зберігати, версіонувати, повторно використовувати та впроваджувати моделі безпосередньо в бізнес-застосунки. Завдяки цьому результати машинного навчання можуть бути оперативно інтегровані в транзакційні системи, аналітичні звіти та сервіси підтримки прийняття рішень.

Отже, Oracle Machine Learning є потужною корпоративною платформою, що поєднує зберігання даних, аналітику та машинне навчання в єдиному середовищі. Її інструментарій забезпечує гнучкість для фахівців різного профілю, високу продуктивність та надійність, що робить OML важливим компонентом сучасних інформаційних систем і ефективним засобом перетворення даних на цінні знанн

РОЗДІЛ 2. ПРОЄКТУВАННЯ ТА ПІДГОТОВКА ДАНИХ ДЛЯ КЛАСИФІКАЦІЙНОГО МОДЕЛЮВАННЯ

2.1. Визначення цільової маркетингової задачі та формування набору даних

У сучасних умовах стрімкого розвитку цифрових технологій та зростання обсягів інформації дедалі більшого значення набувають методи аналізу даних і машинного навчання.



Рис. 3. Модель маркетингової діяльності [10, с. 551]

Одним із ключових напрямів у цій сфері є класифікаційне моделювання, яке широко застосовується в маркетингу для прогнозування поведінки споживачів, сегментації ринку, оцінювання ризиків, персоналізації пропозицій та підтримки управлінських рішень. Проте успішність класифікаційної моделі значною мірою залежить не лише від вибору

алгоритму, а й від коректного визначення цільової маркетингової задачі та якісного проєктування й підготовки набору даних. Саме на цих етапах закладається фундамент для отримання достовірних, інтерпретованих і практично корисних результатів дослідження.

Визначення цільової маркетингової задачі є першим і одним із найважливіших кроків у межах дослідження, орієнтованого на класифікаційне моделювання. Маркетингова задача повинна бути чітко сформульована, релевантна бізнес-цілям організації та такою, що допускає формалізацію у вигляді задачі класифікації [11, с. 3]. На цьому етапі важливо відповісти на запитання, яку саме проблему необхідно вирішити та яку управлінську цінність матиме результат. Наприклад, це може бути прогнозування ймовірності відтоку клієнтів, визначення потенційних покупців нового продукту, класифікація споживачів за рівнем лояльності або виявлення клієнтів із високою ймовірністю реагування на маркетингову кампанію. Кожна з цих задач передбачає наявність цільової змінної, яка поділяє об’єкти на певні класи, що й визначає характер майбутнього моделювання. Чіткість формулювання цільової задачі обумовлює не лише опис бажаного результату, а й уточнення контексту використання моделі.

Необхідно визначити, хто саме буде користувачем результатів моделювання, як часто модель застосовуватиметься, у якому часовому горизонті мають бути актуальними прогнози та які обмеження існують з точки зору ресурсів і даних. Наприклад, у маркетингових дослідженнях важливим є баланс між точністю моделі та її інтерпретованістю, оскільки менеджери часто потребують пояснень, чому система віднесла клієнта до того чи іншого класу. Таким чином, цільова маркетингова задача повинна бути узгоджена як із стратегічними цілями бізнесу, так і з аналітичними можливостями дослідницької команди. Після визначення цільової задачі наступним логічним кроком є формування набору даних, придатного для класифікаційного моделювання. Цей процес починається з ідентифікації джерел даних, які можуть містити інформацію, релевантну для розв’язання поставленої задачі.

У маркетингових дослідженнях такими джерелами можуть бути CRM-системи, транзакційні бази даних, результати опитувань, дані вебаналітики, соціальних мереж, програм лояльності та зовнішні відкриті джерела. Важливо не лише зібрати якомога більше даних, а й оцінити їхню якість, повноту, актуальність та відповідність поставленій цілі. Формування набору даних передбачає вибір об'єктів спостереження та ознак, які описують ці об'єкти.

У контексті класифікаційного моделювання об'єктами зазвичай є клієнти, транзакції, сесії користувачів або інші маркетингові одиниці аналізу. Ознаки, або фактори, повинні відображати суттєві характеристики об'єктів і потенційно впливати на належність до певного класу. На цьому етапі важливу роль відіграє предметна експертиза, оскільки саме знання специфіки ринку, продукту та поведінки споживачів дозволяє відібрати найбільш інформативні змінні та уникнути включення надлишкових або нерелевантних даних. Підготовка даних є комплексним і багатокроковим процесом, який включає очищення, трансформацію та попередній аналіз інформації. Очищення даних спрямоване на усунення помилок, пропусків, дублювань і аномальних значень, що здатні негативно вплинути на якість моделі. У маркетингових наборах даних часто трапляються пропуски, пов'язані з неповними профілями клієнтів або технічними збоями збору інформації, тому необхідно обґрунтовано обирати методи їх обробки, зокрема видалення, заміщення середніми значеннями або застосування більш складних методів імпутації.

Трансформація даних передбачає приведення ознак до формату, зручного для подальшого моделювання. Як правило, подібного роду ситуація може включати кодування категоріальних змінних, нормалізацію або стандартизацію числових показників, агрегацію даних за певними періодами часу чи створення похідних ознак [25]. У рамках маркетингових досліджень створення нових ознак, так званий feature engineering, є особливо важливим, оскільки дозволяє краще відобразити поведінкові патерни клієнтів, наприклад, частоту покупок, середній чек або давність останньої взаємодії з брендом. Окрему увагу в процесі підготовки даних слід приділяти формуванню цільової

змінної для класифікаційного моделювання. Вона повинна бути чітко визначена, коректно вимірювана та логічно пов’язана з маркетинговою задачею. Наприклад, у задачі прогнозування відтоку клієнтів необхідно однозначно визначити критерії, за якими клієнт вважається таким, що пішов, і зафіксувати відповідний часовий інтервал. Неправильне або нечітке визначення цільової змінної може призвести до спотворених результатів і зниження практичної цінності моделі.

Завершальним етапом формування набору даних є його аналіз і перевірка на готовність до класифікаційного моделювання, що переважно включає дослідження розподілів ознак, оцінювання балансу класів, виявлення кореляцій та потенційних проблем, таких як сильна незбалансованість даних. У маркетингових задачах часто трапляється ситуація, коли один із класів є значно менш представленим, наприклад, у випадку прогнозування рідкісних подій. Усвідомлення цих особливостей на етапі підготовки даних дозволяє заздалегідь обрати відповідні методи моделювання та оцінювання результатів. Таким чином, визначення цільової маркетингової задачі та формування набору даних є взаємопов’язаними та критично важливими етапами в рамках дослідження з проєктування й підготовки даних для класифікаційного моделювання. Чітке розуміння бізнес-цілей, грамотне формулювання задачі та ретельна робота з даними забезпечують основу для побудови ефективних моделей, здатних приносити реальну практичну користь у маркетинговій діяльності. Саме системний підхід до цих етапів дозволяє перетворити великі масиви даних на інструмент обґрунтованого прийняття рішень і досягнення конкурентних переваг на ринку.

Стратегії маркетингу можуть істотно відрізнятися залежно від напрями діяльності підприємства міста і її цілей над ринком. Давайте розглянемо основні стратегії, які ви перерахували:

Стратегія розширення ринку. Стратегія розширення ринку фокусується на пошуку нових ринків для існуючих продуктів чи послуг. Це може містити вихід на нові географічні ринки, цільові сегменти або нові канали розподілу.

Основна мета тут – збільшити базу споживачів та продажу, використовуючи вже розроблені та перевірені продукти. Окреслена стратегія часто потребує дослідження нових ринків, адаптації маркетингових повідомлень та тактик під нову аудиторію.

Стратегія інновації чи інтегрованого зростання. Стратегія інновації спрямована на розробку та впровадження нових продуктів чи послуг, а також інновації в бізнес-процесах. Інтегроване зростання може також включати розширення через злиття та поглинання, партнерство з іншими компаніями для спільної розробки продуктів або входження до нових галузей. Метою тут є зміцнення конкурентних переваг та збільшення частки ринку за рахунок інновацій та розвитку.

Стратегія диверсифікації. Диверсифікація – це стратегія, спрямовану розширення лінійки продуктів чи послуг з допомогою запровадження новинок, не пов'язаних безпосередньо з поточним асортиментом [6, с. 67]. Мова йде про диверсифікацію суміжну (розширення в межах схожих галузей) або несуміжну (вихід у нові для компанії області). Диверсифікація спрямована на зменшення ризиків за рахунок розподілу їх по різних продуктах та ринках, а також на пошук нових джерел доходу. Компанії, що використовують диференціацію, можуть цілитися на ширший ринок, але з упором на те, що робить їхню пропозицію унікальною – це може бути якість, дизайн, бренд, технології або особливий рівень обслуговування. Метою тут є залучення та утримання клієнтів, готових платити премію за високу якість чи ексклюзивність.

Стратегія скорочення. Стратегія скорочення включає зменшення масштабів діяльності, продаж або закриття певних підрозділів, продуктових ліній або ринків. Ситуація часто викликана низькою рентабельністю, посиленою конкуренцією чи змінами споживчих переваг [18]. Метою скорочення є зосередження ресурсів на більш прибуткових чи перспективних напрямках, покращення фінансової стійкості та оптимізація операцій.

Кожна з цих стратегій має свої ризики та потенційні вигоди і має бути ретельно спланована та реалізована відповідно до загальної стратегічної мети компанії. Вибір конкретної стратегії залежить від багатьох факторів, включаючи стан ринку, ресурси компанії, конкурентне середовище та довгострокові цілі.

Стратегії маркетингу з охоплення ринку відрізняються підходами до сегментації ринку та цільової аудиторії. Ось короткий огляд кожної з них:

Стратегія масового маркетингу. Масовий маркетинг, також відомий як недиференційований маркетинг, не ділить ринок на сегменти та цільові групи. Натомість він спрямований на звернення до широкої аудиторії з єдиним продуктом або повідомленням [16]. Підхід ґрунтується на припущенні, що продукт чи послуга матимуть універсальний попит серед різних груп споживачів. Прикладами можуть бути базові продукти харчування, побутова хімія та деякі види споживчих товарів. Головна перевага масового маркетингу – економія на масштабі за рахунок широкого охоплення та спрощення маркетингової стратегії.

Стратегія індивідуалізації. Стратегія індивідуалізації (чи мікромаркетинг) спрямовано задоволення специфічних потреб і бажань окремих клієнтів чи дуже вузьких сегментів. Він вимагає глибокого розуміння потреб цільової аудиторії та здатності адаптувати продукти, послуги та маркетингові повідомлення для відповідності цим індивідуальним уподобанням. Індивідуалізація часто використовується у висококласних та нішевих ринках, де клієнти шукають продукти та послуги, які ідеально відповідають їхнім унікальним вимогам.

Маркетинговий план – це документ, у якому докладно описуються цілі маркетингової стратегії компанії, і навіть способи досягнення у певний період [8, с. 10]. У ньому встановлюються маркетингові цілі, вибираються цільові сегменти ринку, визначаються маркетингові заходи та канали комунікації, планується бюджет та оцінюються ризики. Маркетинговий план допомагає компанії сфокусуватися на ключових напрямках розвитку та забезпечує

координоване виконання маркетингових завдань. Кожен аспект маркетингового плану відіграє важливу роль у досягненні кінцевих бізнес-цілей. Збалансована увага до деталей на кожному етапі планування забезпечує комплексний підхід до маркетингу та сприяє сталому розвитку компанії на ринку. У цьому важливо пам'ятати, що маркетинговий план перестав бути статичним документом; він повинен регулярно переглядатися та коригуватися відповідно до змін у зовнішньому середовищі та внутрішній ситуації в компанії.

2.2. Етапи очищення, трансформації та збагачення маркетингових даних

Етап очищення маркетингових даних у дослідженні класифікаційних моделей OML спрямований на усунення помилок, неточностей і неузгодженостей, які неминуче виникають під час збору інформації з різнорідних джерел, таких як CRM-системи, ERP, онлайн-платформи, системи вебаналітики або зовнішні маркетингові бази. Насамперед виконується аналіз повноти даних, що дозволяє виявити пропущені значення у числових, категоріальних і часових атрибутах. В Oracle Machine Learning це може здійснюватися за допомогою SQL-запитів, аналітичних функцій або вбудованих засобів попереднього аналізу даних у середовищі OML Notebooks. Пропущені значення обробляються відповідно до їхньої природи та значущості для моделі: у деяких випадках доцільним є видалення записів, у інших – застосування імпутації на основі середніх значень, медіани або модальних категорій, що особливо важливо для стабільності класифікаційних алгоритмів.

Окрему увагу на етапі очищення приділяють виявленню та усуненню дублюючих записів, які можуть виникати внаслідок повторної реєстрації клієнтів, інтеграції кількох маркетингових джерел або помилок синхронізації. В Oracle Database такі проблеми вирішуються за допомогою групування, унікальних ключів та спеціальних правил дедуплікації, що дозволяє

сформувати єдине уявлення про клієнта. Також важливим аспектом очищення є робота з аномальними значеннями, наприклад, нетипово великими сумами покупок або нереалістично високою частотою транзакцій. За допомогою статистичних методів і аналітичних функцій Oracle ці викиди ідентифікуються та або коригуються, або виключаються з подальшого аналізу, щоб не спотворювати результати класифікації. Після очищення даних наступає етап трансформації, який полягає у приведенні маркетингових даних до формату, оптимального для побудови класифікаційних моделей у Oracle Machine Learning.

Однією з ключових трансформацій є кодування категоріальних змінних, таких як стать, регіон, канал комунікації або сегмент клієнта. В OML це може реалізовуватися як автоматично під час навчання моделі, так і заздалегідь за допомогою SQL-перетворень або спеціальних пакетів Oracle Data Mining. Числові ознаки, зокрема фінансові показники чи поведінкові метрики, часто потребують нормалізації або стандартизації, що підвищує ефективність роботи алгоритмів, чутливих до масштабів даних, наприклад логістичної регресії або методів на основі відстаней. Важливим елементом трансформації є агрегація даних, яка дозволяє перейти від транзакційного рівня до рівня клієнта чи сегмента, що є типовим для маркетингових задач класифікації. У середовищі Oracle це досягається шляхом використання SQL-агрегацій, віконних функцій та матеріалізованих подань, які дозволяють обчислювати такі показники, як середня сума покупки, кількість взаємодій за певний період або час від останнього контакту з брендом. Такі трансформації сприяють узагальненню інформації та формуванню стабільних ознак, що краще відображають поведінкові патерни клієнтів.

Етап збагачення маркетингових даних є логічним продовженням трансформації та спрямований на підвищення інформативності набору ознак шляхом додавання нових даних або створення похідних характеристик. У контексті Oracle Machine Learning збагачення часто здійснюється за рахунок інтеграції внутрішніх корпоративних даних з зовнішніми джерелами,

наприклад соціально-демографічною статистикою, географічними довідниками або макроекономічними показниками. Завдяки потужним засобам роботи з даними в Oracle Database така інтеграція може виконуватися безпосередньо на рівні SQL, що забезпечує цілісність і узгодженість інформації. Особливе значення у збагаченні даних має створення нових ознак, орієнтованих на специфіку маркетингової задачі та вимоги класифікаційної моделі. У дослідженнях з використанням OML широко застосовується підхід feature engineering, який включає розрахунок індексів лояльності, показників життєвої цінності клієнта, скорингових балів або бінарних індикаторів поведінкових подій. Oracle Machine Learning дозволяє реалізовувати такі обчислення як за допомогою SQL, так і з використанням вбудованих мов R або Python, інтегрованих у середовище бази даних, що забезпечує гнучкість і масштабованість процесу [7].

У підсумку, етапи очищення, трансформації та збагачення маркетингових даних у рамках дослідження класифікаційних моделей Oracle Machine Learning формують цілісний і послідовний процес підготовки даних, орієнтований на підвищення якості моделей і достовірності результатів. Поєднання глибокої предметної експертизи з технічними можливостями OML та Oracle Database дозволяє ефективно працювати з великими обсягами маркетингової інформації, створюючи надійну основу для побудови точних, інтерпретованих і практично значущих класифікаційних рішень.

2.3. Розділення даних та вибір контрольних параметрів моделювання

Розділення даних є необхідною умовою об’єктивного оцінювання ефективності класифікаційної моделі. У маркетингових дослідженнях, де моделі часто використовуються для прогнозування майбутньої поведінки клієнтів, критично важливо перевіряти їх здатність узагальнювати знання на нові, раніше невідомі дані. У середовищі Oracle Machine Learning зазвичай застосовується класичний підхід поділу вибірки на навчальну та тестову

частини, а за потреби – і на валідаційну. Навчальна вибірка використовується безпосередньо для побудови класифікаційної моделі, тестова – для оцінювання її якості, а валідаційна – для підбору параметрів і порівняння альтернативних моделей.

Процес розділення даних у Oracle Database може бути реалізований різними способами, залежно від обсягу даних і характеру маркетингової задачі. Найпоширенішим є випадковий поділ із фіксованими пропорціями, наприклад 70 % даних для навчання та 30 % для тестування. Такий підхід легко реалізується за допомогою SQL-функцій випадкового розподілу або спеціальних атрибутів, що позначають приналежність запису до певної підвибірки. Водночас у маркетингових сценаріях, пов’язаних із часовою динамікою, наприклад аналізом поведінки клієнтів у різні періоди, доцільним є хронологічне розділення даних, коли модель навчається на історичних даних і тестується на більш пізніх спостереженнях.

Окремої уваги потребує розділення даних у випадку незбалансованих класів, що є типовим для багатьох маркетингових задач, зокрема прогнозування відтоку клієнтів або виявлення рідкісних реакцій на кампанії. У таких ситуаціях важливо забезпечити збереження пропорцій класів у навчальній і тестовій вибірках. У Oracle Machine Learning це досягається шляхом стратифікованого розподілу, який дозволяє уникнути ситуації, коли один із класів майже повністю відсутній у тестових даних, що унеможливлює коректну оцінку моделі [27]. Таким чином, правильно організоване розділення даних безпосередньо впливає на достовірність результатів дослідження. Після формування навчальної та контрольних вибірок ключовим етапом стає вибір контрольних параметрів моделювання, або гіперпараметрів, які визначають поведінку алгоритмів класифікації в Oracle Machine Learning. До таких параметрів належать налаштування, що не навчаються безпосередньо з даних, але суттєво впливають на точність, стабільність і інтерпретованість моделі. У середовищі OML користувач має доступ до широкого набору алгоритмів класифікації, зокрема логістичної регресії, наївного баєсівського

класифікатора, дерев рішень, випадкових лісів і методів опорних векторів, кожен з яких має власний набір контрольних параметрів.

Вибір контрольних параметрів зазвичай здійснюється на основі валідаційної вибірки або з використанням перехресної перевірки. В Oracle Machine Learning ці підходи можуть бути реалізовані через вбудовані механізми оцінювання моделей або шляхом багаторазового навчання моделей із різними конфігураціями параметрів. Наприклад, для дерев рішень важливими параметрами є максимальна глибина дерева, мінімальна кількість об’єктів у вузлі та критерій розбиття, тоді як для логістичної регресії ключову роль відіграють метод регуляризації та коефіцієнти штрафу. Правильний підбір цих параметрів дозволяє досягти оптимального балансу між точністю прогнозу та узагальнюючою здатністю моделі. У маркетингових дослідженнях особливе значення має вибір контрольних параметрів з урахуванням бізнес-інтерпретації результатів. Наприклад, модель з максимальною точністю може виявитися надто складною для пояснення маркетинговим менеджерам, тоді як дещо простіша модель із зрозумілими правилами класифікації може мати більшу практичну цінність. Oracle Machine Learning надає інструменти для аналізу важливості ознак і впливу параметрів на результат, що дозволяє досліднику усвідомлено керувати процесом налаштування моделей [31, с. 8]. Завершальним аспектом вибору контрольних параметрів є встановлення критеріїв оцінювання якості класифікації, які використовуються для порівняння альтернативних моделей і конфігурацій.

У маркетингових задачах це можуть бути не лише стандартні метрики, такі як точність або повнота, а й показники, що безпосередньо відображають бізнес-ефект, наприклад рівень правильно ідентифікованих цільових клієнтів або зменшення витрат на маркетингові кампанії. В Oracle Machine Learning ці метрики, яу правило, інтегровані у процес моделювання, що забезпечує прозорий і керований вибір оптимальних параметрів.

Таким чином, розділення даних і вибір контрольних параметрів моделювання в рамках дослідження класифікаційних моделей Oracle Machine

Learning є взаємопов’язаними етапами, які забезпечують надійність, об’єктивність і практичну значущість результатів. Системний підхід до формування вибірок і налаштування алгоритмів дозволяє створювати ефективні класифікаційні моделі, здатні підтримувати прийняття обґрунтованих маркетингових рішень у реальних бізнес-умовах.

РОЗДІЛ 3. ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ ORACLE MACHINE LEARNING

3.1. Реалізація класифікаційних моделей у середовищі OML

Для багатьох завдань виявлення основних характеристик та ознак даних, які дозволять досягти кращих якісних показників, є складним та трудомістким процесом. Одночасно з цим на практиці при обробці інформаційних потоків виникають явища, пов'язані з дрейфом концепції, коли відбувається зміщення діапазонів цільових змінних та зміни розподілів даних. Все це наводить до того, що з часом будь-яка модель може погіршити свої якісні показники обробки. На нижньому рівні відбувається обробка інформаційних потоків, що надходять на вхід і послідовностей, але в верхньому – вирішуються завдання поділу потоків на частини, визначення поточних властивостей даних у сегментах та, виходячи з їх властивостей, призначення найбільш підходящої за якісними показниками моделей нижнього рівня.

Більшість методів машинного навчання використовують «централізовані дані», де у вибірках зберігається вся інформація щодо об'єктів спостереження. Процеси збору здійснюються протягом певного періоду часу і, як правило, містять кортежі значень, коли система, що спостерігається, знаходиться в різних станах і її впливає безліч різнорідних чинників [22]. В результаті виникають явища, пов'язані з трансформацією властивостей, зміщення діапазонів значень, одержаних від реєструючих елементів. Це призводить до виникнення неоднорідності даних у вибірках. В окремих послідовностях усередині вибірки може з'являтися дисбаланс класів, відбуватися зміна розподілів, зміна ймовірностей виникнення подій та частоти появи об'єктів спостереження.

Розв'язання задач класифікації та регресії методами машинного навчання може бути утруднено у разі виникнення різних статистичних ефектів. Наприклад, якщо у даних спостерігається парадокс Сімпсона,

стандартний підхід до централізованого інтелектуального аналізу вибірки може не дозволяти досягати заданих якісних показників обробки даних, а результат обробки не відповідати справжньому станом. Сучасні підходи до побудови моделей обробки пов'язані з формуванням, аналізом та об'єднанням локальних результатів, у якому використовуються методи агрегування. Методи та алгоритми, що вирішують завдання класифікації та регресії, можуть мати різні результати обраного показника якості на тому самому наборі даних, наприклад, у разі різних навчальних прикладів. Отримані під час обробки об'єктів спостереження результати різних класифікаторів можуть відрізнятиметься. Їхні відповіді поєднуються агрегуючою функцією для формування загальної відповіді.

За рахунок інтеграції кількох моделей у ряді випадків стає можливим підвищити якість класифікації. На сьогоднішній день домінуючими є ансамблеві методи. Серед них найбільш відомі підходи на основі простого, комбінованого голосування [30], а також застосування ряду агрегуючих функцій, що обчислюють максимальні, середні, медіанні та інші ймовірності класів, що усереднюють результат передбачення щодо сукупності відповідей. В якості альтернативи використовуються різні агрегатори, засновані на рейтингах класифікуючих алгоритмів, арбітри, комбінатори, які можуть бути застосовані як до бінарних, так і до багатокласових завдань.

Останніми роками у низці завдань обробки інформації застосовуються гібридні класифікатори. Використовуються комбінації методів, де різні моделі на основі відносно простих класифікуючих алгоритмів та складних нейронних мереж досягають високих показників повноти та точності [23]. Однак можливості негібридних моделей залежать від властивостей навчальної вибірки та, у разі зміни характеристик даних, якісні показники можуть суттєво знизитися. Точність, повнота, результати обробки залежать від багатьох факторів. Застосування таких підходів часто наводить до виникнення різних ситуацій, коли агрегація різних моделей не тільки не сприяє підвищенню якісних показників, а навпаки, погіршує результати. Причому подібні ефекти

часто нівелюються на великій вибірці даних, але явно простежуються її окремих сегментах. Проте подібна ситуація призводить до того, що при обробці потоків даних можливі помилки внаслідок різних налаштувань класифікуючих моделей.

Таким чином, необхідно розробляти нові та адаптувати існуючі стратегії, що дають можливість проводити точне та надійне навчання в рамках поділу функцій та вибірок.

Практично всі пропоновані підходи, методи, алгоритми машинного навчання на сьогоднішній день є вузько спеціалізованими. Кожна модель досягає певних якісних показників для тих предметних областей, де вона оптимізувалася, і на яких вона навчалася. Одна з основних проблем досягнення якісних показників у методах машинного навчання пов'язана з тим, що при зміні властивостей даних, що надходять, виникає необхідність у додатковому навчанні. Більшість моделей, вирішальні завдання класифікації та регресії, що навчаються на заздалегідь визначеній сукупності об'єктів спостереження. У разі появи трансформації властивостей інформаційних послідовностей якість обробки знижується.

В основі запропонованого рішення лежить можливість побудови ієрархій, коли модель верхнього рівня застосовується для призначення найбільш ефективної моделі нижнього рівня окремий сегмент вибірки. Для цього на різних рівнях ієрархії поділяються ролі та виконувані функції моделей.

На нижніх рівнях розташовуються класифікатори та алгоритми, реалізації яких призначені для вирішення вузькоспеціалізованих завдань. На верхньому рівні застосовуються моделі, що вирішують задачі керування алгоритмами нижнього рівня. Вони «оцінюють» інформаційні послідовності, що надходять, визначають їх властивості, сегментують дані та призначають найбільш підходящу для вирішення поточного завдання модель нижнього рівня. Надалі на верхньому рівні здійснюється моніторинг помилок та відхилень результатів від реальних значень, організується безперервне

навчання моделей нижнього рівня. Застосування ієрархії спрямоване забезпечення паралельного функціонування алгоритмів.

Ієрархічна система координує дії для досягнення заданих якісних показників обробки даних, здійснює аналіз поточних властивостей даних, що використовуються призначення ефективних алгоритмів. В основі реалізації багаторівневих моделей для класифікації та регресії лежить поділ їх функцій обробки. З метою управління класифікаторами, алгоритмами, працюючими безпосередньо з даними, формуються моделі верхнього рівня. Вони, залежно від властивостей даних та необхідних якісних показників, призначають найбільш підходящу модель нижнього рівня. До їхнього функціоналу входять рішення ряду перерахованих нижче завдань:

− ідентифікація та сегментування вступника на вхід послідовності;

− формування пулу передвиборних моделей для її опрацювання;

− вибір показників та призначення найбільш підходящої моделі для обробки поточної послідовності;

− аналіз результатів обробки;

− ухвалення рішення про формування нової вибірки та додаткового навчання моделей.

Для їх вирішення моделями верхнього рівня здійснюється управління та запуск стандартних алгоритмів кластеризації, пошуку точок розладки для поділу послідовності, що надходить, виконання процесів навчання наперед визначеного пулу моделей нижнього рівня, визначення їх властивостей та призначення на сегменти тих, що мають кращі якісні показники обробки. Вони також відстежують властивості даних, що надходять і, у разі виявлення змін та погіршення результатів обробки, ухвалюють рішення про запуск процедур навчання. Функції класифікаторів та обробних алгоритмів нижнього рівня спрямовані на вирішення заздалегідь визначених завдань обробки та обчислення інформації. Для їхнього навчання на вхід подаються сформовані моделями верхнього рівня підвиборки, де властивості об'єктів спостереження різняться. У процесі навчання визначаються моделі, досягають найкращих

показників якості на сегментах. В результаті, залежно від властивостей вибірки даних визначаються моделі обробки, які використовуються при поява близьких за властивостями сегментів.

Використання декількох моделей для підвищення якості результатів передбачення відомо як ансамблеві методи. Незважаючи на різні комбінації, що поєднують окремі алгоритми у модель, виникають різні ситуації, де таке об'єднання не лише може не підвищити, а й навіть погіршити результат. У зв'язку з цим необхідно запобігати таким ситуаціям. На відміну від класичних підходів, коли методи ансамблю намагаються об'єднати кілька моделей з певними стратегіями для покращення якісних показників класифікації, у запропонованому рішенні робиться вибір найбільш підходящої моделі для поточних властивостей даних. Таким чином простіше уникати різних ситуацій, коли слабкі класифікатори можуть погіршити якісні показники сильнішої моделі обробки даних. Для цього розглядається застосування дворівневих моделей, де на нижньому рівні йде обробка послідовності, а на верхньому – відбуваються процедури оцінювання даних та призначення алгоритмів.

Якість результатів вирішення будь-якого завдання методами машинного навчання залежить від властивостей даних. Моделі верхнього рівня призначені для відстеження змін в інформаційних послідовностях внаслідок впливу зовнішніх та внутрішніх факторів. Для цього вони запускають процеси, розбиття на окремі сегменти, виконують алгоритми визначення точки розладки для тимчасових рядів, визначають зміни властивостей, а також проводять аналіз і здійснюють призначення алгоритмів, моделей, які мають кращі характеристики для обраних сегментів.

У загальному підсумку вирішення завдань класифікації та регресії здійснюється на багатовимірних інформаційних даних послідовностей, у тому числі формуються вибірки. Усі інформаційні послідовності піддаються процедурам обробки та сегментуються.

Постановка практичної задачі

У межах практичної частини розглядається задача прогнозування ймовірності відтоку клієнтів (customer churn) для компанії, що працює у сфері роздрібних послуг. Відтік клієнтів є критичним показником ефективності маркетингової діяльності, оскільки безпосередньо впливає на доходи та витрати на залучення нових споживачів. Цільовою змінною у дослідженні є бінарний показник, який набуває значення «1» у разі припинення співпраці клієнта з компанією протягом визначеного періоду та «0» у протилежному випадку. Основним завданням є побудова класифікаційних моделей, здатних з високою точністю ідентифікувати клієнтів із підвищеним ризиком відтоку для подальшого застосування таргетованих маркетингових заходів утримання.

Опис і підготовка набору даних

Для дослідження використовується маркетинговий набір даних, сформований на основі інформації з CRM-системи. Набір даних містить записи про клієнтів і включає демографічні, транзакційні та поведінкові ознаки, зокрема вік клієнта, тривалість співпраці, кількість звернень до служби підтримки, середню суму платежів, частоту використання послуг і тип тарифного плану. Дані збережені у таблицях Oracle Database, що дозволяє здійснювати всі етапи підготовки та моделювання безпосередньо в середовищі бази даних.

На першому етапі виконано очищення даних: виявлено та усунено дублікати клієнтів, оброблено пропущені значення шляхом імпутації та видалено аномальні спостереження, які могли б спотворити результати моделювання. Далі проведено трансформацію ознак, що включала кодування категоріальних змінних, нормалізацію числових показників і агрегацію транзакційних даних до рівня окремого клієнта.

На етапі збагачення сформовано додаткові похідні показники, зокрема індикатори активності клієнта та узагальнені показники лояльності.

Розділення даних і налаштування експерименту

Підготовлений набір даних було розділено на навчальну та тестову вибірки у співвідношенні 70/30 із збереженням пропорцій класів цільової

змінної. Такий підхід дозволив забезпечити репрезентативність даних і коректність оцінювання результатів. Навчальна вибірка використовувалася для побудови моделей, тоді як тестова – для незалежної перевірки їхньої якості.

У межах практичного дослідження було реалізовано кілька класифікаційних моделей за допомогою Oracle Machine Learning, зокрема логістичну регресію, дерево рішень і випадковий ліс. Для кожного алгоритму здійснювався підбір контрольних параметрів з урахуванням маркетингової інтерпретованості результатів і ризику перенавчання.

Налаштування моделей проводилося з використанням вбудованих механізмів OML та SQL-інтерфейсів. Оцінювання ефективності побудованих моделей здійснювалося на тестовій вибірці з використанням стандартних метрик класифікації, зокрема точності, повноти, F-міри та площі під ROC-кривою. Отримані результати показали, що ансамблеві методи, зокрема випадковий ліс, забезпечують вищу загальну точність і кращу здатність виявляти клієнтів із високим ризиком відтоку. Водночас логістична регресія продемонструвала дещо нижчі значення метрик, але відзначалася високою інтерпретованістю, що є важливим чинником для прийняття маркетингових рішень.

Порівняльний аналіз моделей також включав дослідження матриць помилок і аналіз важливості ознак. Було встановлено, що ключовими факторами відтоку є тривалість співпраці, частота звернень до служби підтримки та зниження активності користування послугами. Ці результати мають безпосередню практичну цінність, оскільки дозволяють маркетинговим підрозділам розробляти превентивні заходи для утримання клієнтів.

Як приклад, розглянемо особливості застосування дерева рішень в Oracle Machine Learning.

Формально, дерево рішень можна подати як орієнтований ациклічний граф

$$T = (V, E),$$ де:

V – множина вузлів (внутрішні вузли + листки),

E – множина ребер, кожен внутрішній вузол відповідає перевірці умови над ознакою x_j,

кожен листок відповідає прогнозу y.

Алгоритм використовується для:

- класифікації (ціль – дискретна змінна),
- регресії (ціль – неперервна змінна).

В Oracle Machine Learning алгоритм дерева рішень реалізований як Oracle Decision Tree (ODT) – власна реалізація Oracle, оптимізована для виконання безпосередньо в СУБД Oracle (in-database learning).

Ключові особливості ODT:

- відсутність переміщення даних за межі БД;
- масштабованість за рахунок SQL-оптимізації;
- підтримка як classification, так і regression;
- автоматична обробка пропущених значень;
- вбудований механізм pruning (обрізання).

До прикладу, ми задаємо навчальну вибірку:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

$x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ – вектор ознак,

y_i – цільова змінна.

Мета: побудувати дерево T, яке мінімізує функцію втрат:

для класифікації – ентропію або Gini,

для регресії – середньоквадратичну помилку (MSE).

Критерії розбиття в Oracle Decision

Ентропія (Classification)

$$H(S) = -\sum_{k=1}^{K} p_k \log_2 p_k$$

де pk – ймовірність класуk.

Information Gain:

$$IG(S, x_j) = H(S) - \sum_{v \in Values(x_j)} \frac{|S_v|}{|S|} H(S_v)$$

Oracle автоматично вибирає ознаку з максимальним IG.

Для регресії мінімізується:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Алгоритмічно процес виглядає так:

- Ініціалізація кореневого вузла.
- Для кожної ознаки – оцінка всіх можливих розбиттів.
- Вибір оптимального розбиття.
- Рекурсивне повторення для дочірніх вузлів.
- Зупинка за умов: досягнення max depth, мінімальна кількість записів, відсутність покращення критерію.
- Обрізання (pruning) для уникнення overfitting.

Постановка задачі із ціллю y = DEFAULT

AGE	INCOME	LOAN	DEFAULT
35	50000	0	NO
42	72000	1	YES
28	32000	0	NO

Побудова моделі в OML (SQL API)

```
BEGIN
 DBMS_DATA_MINING.CREATE_MODEL(
  model_name          => 'DT_DEFAULT_MODEL',
  mining_function     => DBMS_DATA_MINING.CLASSIFICATION,
  data_table_name     => 'CLIENTS',
  case_id_column_name => 'CLIENT_ID',
  target_column_name  => 'DEFAULT',
  settings_table_name => 'DT_SETTINGS'
```

```
  );
END;
/
```

Логіка дерева (інтерпретація)

```
IF LOAN = 1
  IF INCOME < 60000 → DEFAULT = YES
  ELSE → DEFAULT = NO
ELSE
  DEFAULT = NO
```

Наукова інтерпретація:

- Ознака LOAN має максимальний Information Gain.
- INCOME уточнює підмножину ризикових клієнтів.
- Листки дерева апроксимують апостеріорну ймовірність класу.

Ціль – прогноз ціни нерухомості.

PRICE = f (AREA, ROOMS, LOCATION)

У листку дерева зберігається:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Формальні переваги:

- інтерпретованість (white-box модель);
- несприйнятливість до масштабування ознак;
- мінімальні припущення щодо розподілу даних;
- інтеграція з SQL;
- In-Database Learning → мінімальна латентність.

Застосування дерева рішень в Oracle Machine Learning є ефективним підходом для побудови інтерпретованих моделей класифікації та регресії безпосередньо в середовищі СУБД. Oracle Decision Tree поєднує класичні принципи теорії інформації з оптимізацією виконання на рівні бази даних, що робить його доцільним для промислових аналітичних систем. Єдині застереження в даному випадку: схильність до overfitting; нестабільність при

малих змінах даних; нижча точність порівняно з ensemble-методами (Random Forest, Gradient Boosting).

Оцінка маркетингової ефективності

Для оцінки практичної користі моделей результати класифікації було інтерпретовано з точки зору бізнес-показників. Зокрема, моделювалися сценарії таргетованих кампаній утримання клієнтів, у яких контактування здійснювалося лише з клієнтами, ідентифікованими моделлю як високоризикові. Такий підхід дозволив зменшити витрати на маркетингові комунікації та підвищити їхню ефективність порівняно з масовими кампаніями.

Практична частина дослідження підтвердила, що використання Oracle Machine Learning для побудови класифікаційних моделей у маркетингу є ефективним інструментом підтримки управлінських рішень. Комплексний підхід до підготовки даних, коректне розділення вибірок і порівняльний аналіз моделей дозволяють не лише досягти високої якості прогнозування, а й забезпечити реальну бізнес-цінність результатів. Отримані висновки можуть бути використані як основа для впровадження інтелектуальних маркетингових систем у практику діяльності сучасних компаній.

3.2. Порівняльний аналіз результатів моделювання та оцінка ефективності

У межах дослідження та оцінки ефективності класифікаційних моделей Oracle Machine Learning (OML) порівняльний аналіз результатів моделювання є завершальним і водночас одним із найважливіших етапів, оскільки саме він дозволяє визначити, наскільки побудовані моделі відповідають поставленій маркетинговій задачі та забезпечують очікуваний бізнес-ефект. На цьому етапі відбувається зіставлення результатів, отриманих за допомогою різних алгоритмів класифікації, конфігурацій контрольних параметрів і наборів

ознак, а також комплексна оцінка їхньої ефективності з точки зору якості прогнозування та практичної цінності для маркетингових рішень.

Порівняльний аналіз у середовищі Oracle Machine Learning ґрунтується на використанні єдиної тестової або контрольної вибірки, сформованої на попередніх етапах дослідження, що забезпечує коректність порівняння, оскільки всі моделі оцінюються в однакових умовах і на одних і тих самих даних. У рамках маркетингових задач класифікації, таких як прогнозування відтоку клієнтів, визначення ймовірності відгуку на кампанію або сегментація за рівнем лояльності, зазвичай порівнюються кілька базових і розширених моделей, зокрема логістична регресія, наївний баєсівський класифікатор, дерева рішень, випадкові ліси та інші алгоритми, доступні в OML. Одним із ключових аспектів порівняльного аналізу є використання кількісних метрик якості класифікації.

```
SELECT
   COUNT(*) AS total_customers,
   SUM(CASE WHEN churn = 1 THEN 1 ELSE 0 END) AS churned_customers
FROM marketing_customers;
```

Запит дозволяє оцінити загальний обсяг вибірки та ступінь незбалансованості класів цільової змінної churn.

У Oracle Machine Learning доступні стандартні показники, такі як точність (accuracy), повнота (recall), точність позитивного прогнозу (precision), F-міра та площа під ROC-кривою (AUC). Кожна з цих метрик має різну інтерпретацію та по-різному відображає ефективність моделі в маркетинговому контексті. Наприклад, у задачах таргетованого маркетингу особливу цінність має висока точність позитивного прогнозу, оскільки вона зменшує витрати на контакт із клієнтами, які з високою ймовірністю не відреагують на пропозицію. Водночас у задачах утримання клієнтів критично важливою є повнота, яка відображає здатність моделі виявляти максимальну кількість потенційних «ризикових» клієнтів. Порівнюючи моделі за цими

метриками, дослідник отримує можливість не лише визначити формального лідера за точністю, а й оцінити компроміси між різними характеристиками якості. Oracle Machine Learning дозволяє будувати та аналізувати матриці помилок, ROC-криві та інші аналітичні візуалізації безпосередньо в середовищі бази даних або через OML Notebooks, що значно спрощує інтерпретацію результатів. Завдяки цьому можна виявити, наприклад, що дві моделі мають схожі значення AUC, але суттєво відрізняються за кількістю хибнопозитивних або хибнонегативних прогнозів, що безпосередньо впливає на маркетингову стратегію.

Вичвлення пропущених значень відбувається наступним чином:

```
SELECT
    COUNT(*) AS missing_income
    FROM marketing_customers
    WHERE monthly_income IS NULL;
```

Тоді як імпутація – таким способом:

```
    UPDATE marketing_customers mc
    SET monthly_income = (
       SELECT ROUND(AVG(monthly_income), 2)
       FROM marketing_customers
       WHERE monthly_income IS NOT NULL
    )
    WHERE mc.monthly_income IS NULL;
```

Для видалення дублікатів клієнтів було застосовано:

```
    DELETE FROM marketing_customers a
    WHERE ROWID > (
       SELECT MIN(ROWID)
       FROM marketing_customers b
       WHERE a.customer_id = b.customer_id
    );
```

Важливим елементом порівняльного аналізу є оцінка стабільності та узагальнюючої здатності моделей. У маркетингових дослідженнях моделі часто застосовуються до нових сегментів клієнтів або в інших часових періодах, тому важливо перевіряти, наскільки результати моделювання є стійкими до змін даних. У середовищі Oracle Machine Learning це може бути реалізовано шляхом багаторазового тестування моделей на різних підвибірках або шляхом аналізу їхньої поведінки на відкладених часових інтервалах. Моделі, які демонструють високу точність лише на навчальних даних, але втрачають ефективність на тестових, розглядаються як менш придатні для практичного використання.

Створення похідних маркетингових ознак:

```
ALTER TABLE marketing_customers ADD (
   avg_transaction_value NUMBER,
   activity_index NUMBER
);
UPDATE marketing_customers
SET avg_transaction_value = total_spent / NULLIF(total_transactions, 0),
   activity_index = total_transactions / months_active;
```

Окремий вимір порівняльного аналізу пов’язаний з інтерпретованістю моделей і можливістю пояснення результатів маркетинговим фахівцям. У цьому контексті логістична регресія та дерева рішень часто мають перевагу над більш складними ансамблевими методами, оскільки дозволяють безпосередньо аналізувати вплив окремих ознак на результат класифікації. Oracle Machine Learning надає інструменти для оцінювання важливості змінних і візуалізації правил класифікації, що дозволяє включити якісний аналіз у процес порівняння моделей. Для маркетингу це має принципове значення, адже розуміння причин прогнозу часто є не менш важливим, ніж сам прогноз. Завершальним етапом порівняльного аналізу є оцінка ефективності моделей з точки зору бізнес-результатів.

Кодування категоріальних змінних відббувається наступним чином:

```
SELECT
   customer_id,
   CASE tariff_plan
      WHEN 'BASIC' THEN 1
      WHEN 'STANDARD' THEN 2
      WHEN 'PREMIUM' THEN 3
   END AS tariff_code
FROM marketing_customers;
```

У маркетинговому контексті це означає зіставлення аналітичних показників із реальними або очікуваними економічними ефектами, такими як зростання конверсії, зниження рівня відтоку, оптимізація витрат на кампанії або підвищення довічної цінності клієнта. У рамках Oracle Machine Learning результати класифікації можуть бути інтегровані з операційними даними, що дозволяє моделювати сценарії використання різних моделей і оцінювати їхній вплив на ключові показники ефективності маркетингу. Таким чином, порівняльний аналіз результатів моделювання та оцінка ефективності в маркетингу в рамках дослідження класифікаційних моделей Oracle Machine Learning є багатовимірним процесом, який поєднує статистичні, аналітичні та бізнес-орієнтовані підходи. Саме комплексне зіставлення моделей за якісними метриками, стабільністю, інтерпретованістю та маркетинговою цінністю дозволяє обґрунтовано обрати оптимальне рішення та забезпечити практичну результативність застосування машинного навчання в маркетинговій діяльності.

Побудова класифікаційних моделей в Oracle Machine Learning:

```
BEGIN
 DBMS_DATA_MINING.CREATE_MODEL(
   model_name          => 'CHURN_GLM_MODEL',
   mining_function     => DBMS_DATA_MINING.CLASSIFICATION,
   data_table_name     => 'TRAIN_DATA',
   case_id_column_name => 'CUSTOMER_ID',
```

```
    target_column_name  => 'CHURN',
    settings_table_name => 'GLM_SETTINGS'
  );
END;
/
```

Приклад таблиці параметрів:

```
INSERT INTO glm_settings VALUES (
  DBMS_DATA_MINING.ALGO_NAME,
  DBMS_DATA_MINING.ALGO_GENERALIZED_LINEAR_MODEL
);
```

Застосування моделей та методів штучного інтелекту можливо у проблемних областях, де виникає невизначеність та важко визначити оптимальні рішення. З метою підвищення показників якості класифікаційних та регресійних моделей існує можливість реалізації процесів попередньої обробки вибірки даних, щоб витягти певні атрибути, які представляють найбільш важливі властивості інформації. У різних сегментах доводиться реалізовувати розділяючі поверхні різною складності, що призводить до того, що на різних підвиборках краще працюють різні моделі. Збір об'єктів спостереження є трудомістким завданням, де всередині кортежів можуть виникати різні усунення через налаштування реєструючих пристроїв або втрат точності значень окремих параметрів. Вилучення ознак може втрачати свою актуальність у разі виникнення дрейфу концепції.

Отимання прогнозів:

```
SELECT
    customer_id,
    PREDICTION(CHURN_GLM_MODEL USING *) AS predicted_churn,
    PREDICTION_PROBABILITY(
       CHURN_GLM_MODEL, 1 USING *
    ) AS churn_probability
FROM test_data;
```

Матриця помилок:

```
SELECT
   churn AS actual,
   PREDICTION(CHURN_GLM_MODEL USING *) AS predicted,
   COUNT(*) AS total
FROM test_data
GROUP BY churn, PREDICTION(CHURN_GLM_MODEL USING *);
```

Аналіз важливості ознак:

```
SELECT *
FROM TABLE(
  DBMS_DATA_MINING.GET_MODEL_DETAILS_GLM(
   'CHURN_GLM_MODEL'
  )
);
```

Подібний аналіз дозволяє визначити, які маркетингові показники найбільше впливають на ризик відтоку клієнтів.

Інтерпретація для маркетингу виглядає наступним чином, результат чого може бути використаний для формування таргетованої кампанії утримання клієнтів:

```
SELECT
   COUNT(*) AS high_risk_customers
FROM test_data
WHERE PREDICTION_PROBABILITY(
   CHURN_RF_MODEL, 1 USING *
) > 0.8;
```

У зв'язку з цим необхідно постійно обробляти вибірки даних, що надходять на вхід та аналізувати кожен сегмент. Інформація про властивості даних у сегментах сильно залежить від способу сегментації та поділу вибірки. Обробка цих даних необхідно отримати інформацію про роздільність класів, сформувати розділяючу поверхню та підвищити якісні показники

класифікуючого алгоритму. Використання декількох моделей для підвищення якості результатів передбачення у вигляді ансамблевих методів, наводить до того, що, незважаючи на різні комбінації, що об'єднують окремі алгоритми у модель, виникають ситуації, де таке об'єднання не тільки може не підвищити, а й навіть погіршити результат. У зв'язку з цим необхідно запобігати таким ситуації. Для цього пропонуються використовувати багаторівневі моделі, де на нижньому рівні йде обробка послідовності, а на верхньому – відбуваються процедури оцінювання даних та призначення алгоритмів. Виконання процесу сегментації та обчислення характеристик має здійснюватися постійно для налаштування моделей обробки.

Запропоновані SQL та Oracle Machine Learning приклади демонструють повний цикл практичного дослідження: від підготовки маркетингових даних до побудови, порівняння та оцінки класифікаційних моделей. Використання OML безпосередньо в середовищі Oracle Database дозволяє забезпечити масштабованість, відтворюваність і тісний зв’язок між аналітичними результатами та бізнес-процесами маркетингу.

ВИСНОВОК

У результаті виконання даної роботи було комплексно досліджено теоретичні, методологічні та прикладні аспекти застосування класифікаційних моделей машинного навчання в сучасному маркетингу з використанням технологічної платформи Oracle Machine Learning. Отримані результати дозволяють сформулювати такі узагальнені наукові висновки.

У першому розділі обґрунтовано ключову роль класифікаційних задач у маркетинговій діяльності підприємств в умовах цифрової економіки. Показано, що задачі сегментації клієнтів, прогнозування відтоку, оцінювання ймовірності відгуку на маркетингові кампанії та виявлення потенційно цінних споживачів є критично важливими для підвищення ефективності управлінських рішень. Доведено, що класифікаційні моделі машинного навчання забезпечують вищу точність, адаптивність та масштабованість порівняно з традиційними статистичними підходами, що робить їх доцільним інструментом аналітичної підтримки маркетингу.

У ході огляду класифікаційних моделей машинного навчання систематизовано їх основні типи, зокрема логістичну регресію, дерева рішень, наївний баєсівський класифікатор, метод опорних векторів, ансамблеві методи та нейронні мережі. Визначено їхні переваги, обмеження та сфери доцільного застосування в маркетингових задачах. Окрему увагу приділено ролі інтерпретованості моделей, що є важливим чинником у практичному використанні результатів моделювання для прийняття бізнес-рішень. Також проаналізовано архітектуру та функціональні можливості платформи Oracle Machine Learning.

Встановлено, що Oracle Machine Learning забезпечує тісну інтеграцію алгоритмів машинного навчання з корпоративними базами даних, підтримує in-database аналітику та дозволяє виконувати повний цикл моделювання без перенесення даних у зовнішні середовища, що підвищує продуктивність, безпеку та відтворюваність аналітичних процесів.

Другий розділ присвячено проєктуванню та підготовці даних для класифікаційного моделювання. Визначено цільову маркетингову задачу та обґрунтовано вибір ознак, що формують інформаційну основу моделі. Доведено, що якість і репрезентативність набору даних є вирішальним чинником ефективності класифікаційних моделей. Детально розглянуто етапи очищення даних, обробки пропущених значень, трансформації змінних та збагачення даних, які дозволяють зменшити вплив шуму та підвищити інформативність ознак. У роботі також обґрунтовано підхід до розділення даних на навчальну та тестову вибірки, а також вибір контрольних параметрів моделювання. Показано, що коректне налаштування гіперпараметрів та використання процедур валідації є необхідними для уникнення перенавчання та забезпечення узагальнювальної здатності моделей.

У третьому розділі реалізовано класифікаційні моделі в середовищі Oracle Machine Learning та проведено їх порівняльний аналіз. На основі обраних метрик якості продемонстровано відмінності в ефективності моделей залежно від їхньої структури, складності та чутливості до характеристик даних. Отримані результати підтверджують, що жодна модель не є універсально оптимальною, а вибір конкретного алгоритму має здійснюватися з урахуванням специфіки маркетингової задачі, обсягу та якості даних, а також вимог до інтерпретованості результатів. Загалом результати дослідження свідчать про високу практичну цінність використання класифікаційних моделей Oracle Machine Learning у маркетингу. Запропонований підхід до побудови та оцінювання моделей може бути використаний як методологічна основа для впровадження аналітичних рішень у діяльність підприємств, а також для подальших наукових досліджень у сфері прикладного машинного навчання та маркетингової аналітики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Баранов О., Костенко О., Дубняк М., Головко О. Цифрова трансформація суспільства: правові аспекти та сучасні проблеми: монографія, 2-ге вид.; ДНУ «Інститут інформації, безпеки і права Національної академії правових наук України». Київ; Одеса: Фенікс, 2024. 184 с.
2. Білоус А. Вступ до Machine Learning: знайомство з моделями. DOU. 2025. URL: https://dou.ua/lenta/articles/introduction-machine-learning-1/
3. Бондар В. О., Кравець О. І. Моделі та методи обробки природної мови. Журнал «Інформаційні технології», 2022. 32 с.
4. Бортник К. Я., Ольшевський О. В., Кирилюк А. Л. Машинне навчання, як основа для розвитку технологій майбутнього. Науковий журнал «Комп'ютерно-інтегровані технології: освіта, наука, виробництво» Луцьк. Вип. № 27. 2017. с. 85-88
5. Волнянский А. Штучний інтелект і машинне навчання в маркетингу: застосування, кейси, статистика. Colobridge. 2025. URL: https://blog.colobridge.net/uk/2025/07/ai-ml-for-marketing-ua/
6. Васильєва М., Васильєв А. Диверсифікація як механізм забезпечення розвитку підприємства в умовах нестабільності ринку. Адаптивне управління: теорія і практика. № 15. 2023. С. 66 – 78
7. Грід-системи та технології хмарних обчислень. Навчальний посібник для здобувачів освітнього рівня «магістр» спеціальностей: F3 «Комп'ютерні науки», F6 «Інформаційні системи та технології» / Г. В. Шимчук, О. Б. Назаревич, Я. В. Литвиненко, В. А. Готович, В. В. Никитюк., І. О. Боднарчук. Тернопіль : ФОП Паляниця В. А., 2025. 396 с.
8. Дарчук В. Г. Аналіз та планування маркетингової діяльності. Навчальний посібник. Київ: ДУТ, 2019. 146 с.

9. Краус К. М., Краус Н. М., Манжура О. В. Електронна комерція та інтернет-торгівля: навчально-методичний посібник. Київ: Аграр Медіа Груп, 2021. 454 с.
10. Котлер Ф. Основи маркетингу. Київ: Видавничий дім «Діалектика», 2020. 800 с.
11. Кузьминчук Н. В., Писаревська Г. І., Ляшенко В. В. Формування маркетингової стратегії як елементу маркетингової діяльності підприємства. Економіка і суспільство. Вип. № 56. 2023. URL: file:///C:/Users/Metropolitan/Downloads/3066-%D0%A2%D0%B5%D0%BA%D1%81%D1%82%20%D1%81%D1%82%D0%B0%D1%82%D1%82%D1%96-2973-1-10-20231206.pdf
12. Лекційний курс із навчальної дисципліни «Інноваційна діяльність» для студентів освітнього рівня бакалавр усіх форм навчання спеціальності 076 «Підприємництво, торгівля та біржова діяльність» / укладач Радинський С.В. Тернопіль, ТНТУ ім.І.Пулюя, 2022. 122 с.
13. Луханіна К. А. Глобальні тренди Інтрнет-маркетингу: сучасний вимір та майбутні трансформації. Економіка та суспільство. 2024. Вип. 60. 2024. с. 1-9
14. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання. Вінниця : ВНТУ, 2024. 258 с. URL: https://pdf.lib.vntu.edu.ua/books/2024/Mokin_2024_263.pdf.
15. Мокрицька О. В., Мочернюк Ю. М. Використання алгоритмів машинного навчання для автоматизації процесу модерації контенту в групових чатах месенджерів. Науковий вісник НЛТУ України. 2024, т. 34, № 7. с. 52–59
16. Писаренко В. В., Багорка М. О. Стратегічний маркетинг: навч.посіб. Дніпро: Видавець, 2019. 240 с.

17. Праведна М. Поведінка споживачів в епоху цифрових технологій та її вплив на маркетингову діяльність. Революція креативних індустрій: як маркетинг і менеджмент створюють майбутнє: тези доповідей І Міжнародної науково-практичної інтернет-конференції (24-25 жовтня 2024 р.). Хмельницький: ХНУ, 2024. С. 59-62
18. Роїк О. М., Азарова А. О., Небава М. І. Основи стратегічного менеджменту. Навчальний посібник. Вінниця: ВНТУ, 2007. 213 с.
19. Сємєхін І. В. Захист прав споживачів на електронному ринку. Формування механізмів управління якістю та підвищення конкурентоспроможності підприємств: XV Міжнародна науково-практична інтернет-конференція: тези доповідей. Дніпро: Університет імені Альфреда Нобеля, 2024. с.233-235
20. Ситник В. Ф., Ситник Н. В. Дерева рішень в системах дейтамайнінгу. Вчені записки, 2006. Вип. № 9. с. 442-454
21. Субботін С. Нейро-нечіткий мережевий синтез та спрощення на прецедентах у задачах діагностики та розпізнавання образів. Оптична пам'ять та нейронні мережі (Інформаційна оптика). 2013. Вип. 22, № 2. с. 97 – 103
22. Федько В. В. Дослідження ефективності застосування технології Machine Learning Services в задачах з прогнозування. Збірник наукових праць Харківського національного університету Повітряних Сил, Вип. 2(68). 2021. С. 116-121
23. Федотова-Півень І. М. Операційні системи: навчальний посібник; Черкаський державний технологічний університет. Харків: ТОВ «ДІСА ПЛЮС», 2019. 216 с.
24. Харченко В. О. Основи машинного навчання: навч. посіб. Суми: Сумський державний університет, 2023. 264 с.
25. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підручник. Київ. 2010. 821 с.

26. Що таке машинне навчання, його алгоритми | Як achine Learning (L) впливає на розвиток бізнесу. Kyivstar Business Hub – корпоративний блог для бізнесу. URL: https://hub.kyivstar.ua/articles/shho-take-mashinne-navchannya-ta-yak-vonovplivaye-na-rozvitok-biznesu
27. Almulla H, Gay G. Learning how to search: Generating exception-triggering tests through adaptive fitness function selection. In IEEE Interna-tional Conference on Software Testing, Validation and Verification (ICST), 2020; p. 63–73.
28. Ashari A., Paryudi I., Min Tjoa A. Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. International Journal of Advanced Computer Science and Applications, Vol. 4, №. 11, 2013. p. 33-39
29. Oracle Machine Learning. Concepts. OML4SQL. Exponential Smoothing. 2023. URL: https://docs.oracle.com/en/database/oracle/machinelearning/oml4sql/21/dmcon/expnential-smoothing.html#GUID-65C7E533-E403-4F71- A5FE-EC034745904F
30. Machine learning facilitated business intelligence (Part I): Neural networks learning algorithms and applications /W.A. Khan, S.H. Chung, M.U. Awan, X. Wen // Industrial Management & Data Systems. 2019. №. 120(1). p. 164-195
31. Fontes A, Gay G. Using machine learning to generate test oracles: A systematic literature review. In International Workshop on Test Oracles,Association for Computing Machinery: New York, NY, USA, TORACLE, 2021; p. 1–10
32. Pan A. A Gentle Introduction to Machine Learning Models. Weights & Baises. 2011. URL: https://wandb.ai/wandb_fc/gentle-intros/reports/A-Gentle-Introduction-to-Machine-Learning-Models--VmlldzoyOTUxNjQw
33. Redych O., Boichuk R. Comparative Analysis of Holt-Winters Algorithms on the Oracle Machine Learning Platform. Proceedings of International

Conference on Applied Innovation in IT. 2024. Volume 12, Issue 1, pp. 71-77

34. Witten I. H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2016. 167 p.

ДОДАТОК 1

Варіант дерева рішень для оцінки кредитного ризику [20, с. 450]

