ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Модель попиту на предмети споживання засобами інтелектуального аналізу даних»

на здобуття освітнього ступеня магістра
зі спеціальності 124 Системний аналіз
освітньо-професійної програми «Інтелектуальні системи управління»

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Валерій ГОМОЛЯКО
(підпис)

Виконав: здобувач вищої освіти групи САДМ-61

Валерій ГОМОЛЯКО

Керівник:
к.т.н., доцент Ганна ЛИХОДЄЄВА

Рецензент:
науковий ступінь, вчене звання *Ім'я, ПРІЗВИЩЕ*

Київ 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут інформаційних технологій

Кафедра інформаційних систем та технологій
Ступінь вищої освіти магістр
Спеціальність 124 Системний аналіз
Освітньо-професійна програма Інтелектуальні системи управління

ЗАТВЕРДЖУЮ
Завідувач кафедрою ІСТ
_____Каміла СТОРЧАК
«____» _____20__р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Гомоляко Валерій Ігорович

1. Тема кваліфікаційної роботи: Модель попиту на предмети споживання засобами інтелектуального аналізу даних

керівник кваліфікаційної роботи Ганна ЛИХОДЄЄВА, к.т.н., доцент
затверджені наказом Державного університету інформаційно-комунікаційних технологій
від «____» _____20__р. № ____

2. Строк подання кваліфікаційної роботи «___» _____20__р.

3. Вихідні дані до кваліфікаційної роботи: наукова література з питань моделювання попиту на предмети споживання засобами інтелектуального аналізу даних, включаючи методи прогнозування попиту, аналіз великих даних, алгоритми машинного навчання, нейронні мережі та регресійні моделі; науково-технічна література, що охоплює сучасні підходи до застосування інтелектуальних систем в аналізі споживчого попиту, а також література, що висвітлює аспекти оптимізації процесів управління підприємствами за допомогою аналізу даних.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
   1. Розділ 1. Загально-теоретична характеристика інтелектуального аналізу даних
   2. Розділ 2. Особливості моделюваня попиту на предмети споживання засобами інтелектуального аналізу даних
   3. Розділ 3. Розробка моделі попиту на предмети споживання засобами інтелектуального аналізу даних

5. Перелік ілюстративного матеріалу: *презентація*
6. Дата видачі завдання «_____» _____20____р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Дослідження теоретико-методологічних основ прогнозування попиту на предмети споживання		
2	Розробка концепції моделі попиту на предмети споживання за допомогою інтелектуального аналізу даних		
3	Визначення основних етапів моделювання попиту та інструментів для збору та аналізу даних		
4	Опис алгоритмів та моделей прогнозування попиту (нейронні мережі, регресія, кластеризація)		
5	Побудова та тестування моделі попиту на основі реальних даних		
6	Оцінка ефективності моделі попиту та рекомендації щодо оптимізації процесу		
7	Оформлення кваліфікаційної роботи та підготовка презентаційних матеріалів		
8	Презентація результатів роботи та захист кваліфікаційної роботи		

Здобувач вищої освіти _____ Валерій ГОМОЛЯКО
(підпис)

Керівник кваліфікаційної роботи _____ Ганна ЛИХОДЄЄВА
(підпис)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 110 стор., 50 рис., 5 табл., 61 джерел.

Мета роботи – створення ефективної моделі прогнозування попиту на предмети споживання на основі інтелектуального аналізу даних, яка дозволяє підвищити точність прогнозів та оптимізувати управління запасами.

Об'єкт дослідження процеси прогнозування та аналізу попиту на предмети споживання.

Предмет дослідження – моделі та методи прогнозування попиту на предмети споживання із використанням інструментів інтелектуального аналізу даних.

Короткий зміст роботи:

У роботі розглянуто модель попиту на предмети споживання за допомогою інтелектуального аналізу даних. Зокрема, детально вивчаються методи прогнозування попиту за допомогою сучасних алгоритмів машинного навчання, таких як нейронні мережі, методи кластеризації та регресійні моделі. Наголошено на важливості використання великих даних та аналітичних платформ для оцінки споживчих переваг, що дозволяє здійснити глибокий аналіз попиту, враховуючи сезонні коливання, економічні умови та інші фактори. Окремо розглянуто використання інтелектуальних систем у бізнес-аналізі та прийнятті рішень щодо виробництва, ціноутворення та маркетингу.

Проаналізовано підходи до моделювання попиту на основі даних з різних джерел, таких як соціальні мережі, відгуки споживачів та історичні дані про покупки. Висвітлено роль цих систем у забезпеченні більш точного прогнозування потреб споживачів і зниженні ризиків для підприємств. Запропоновано методику застосування інтелектуального аналізу даних для оптимізації процесів прийняття рішень у сфері управління попитом.

Зроблено висновок, що використання інтелектуальних систем в аналізі попиту дозволяє підприємствам значно підвищити ефективність управлінських рішень, поліпшити прогнозування попиту та адаптувати стратегії маркетингу та продажу в реальному часі.

КЛЮЧОВІ СЛОВА: ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, МОДЕЛЬ ПОПИТУ, МАШИННЕ НАВЧАННЯ, НЕЙРОННІ МЕРЕЖІ, РЕГРЕСІЯ, АНАЛІТИЧНІ ПЛАТФОРМИ, ПРОГНОЗУВАННЯ, СПОЖИВЧІ ПЕРЕВАГИ, ОПТИМІЗАЦІЯ РІШЕНЬ.

ABSTRACT

Text part of the qualification work for the Master's degree: 110 pages, 50 figures, 5 tables, 61 sources.

The aim of the work is to create an effective demand forecasting model for consumption goods based on intellectual data analysis, which allows for improving the accuracy of forecasts and optimizing inventory management.

The object of research is the processes of forecasting and demand analysis for consumption goods.

Brief summary of the work: The work examines the demand model for consumption goods using intellectual data analysis. Specifically, it studies demand forecasting methods using modern machine learning algorithms such as neural networks, clustering methods, and regression models. The importance of using big data and analytical platforms to assess consumer preferences is emphasized, allowing for in-depth demand analysis, considering seasonal fluctuations, economic conditions, and other factors. The use of intellectual systems in business analysis and decision-making regarding production, pricing, and marketing is also considered.

The approaches to demand modeling based on data from various sources, such as social networks, consumer reviews, and historical purchase data, are analyzed. The role of these systems in ensuring more accurate consumer demand forecasting and reducing risks for businesses is highlighted. A methodology for applying intellectual data analysis to optimize decision-making processes in demand management is proposed.

It is concluded that the use of intellectual systems in demand analysis allows businesses to significantly improve the effectiveness of management decisions, enhance demand forecasting, and adapt marketing and sales strategies in real time.

KEYWORDS: INTELLECTUAL DATA ANALYSIS, DEMAND MODEL, MACHINE LEARNING, NEURAL NETWORKS, REGRESSION, ANALYTICAL PLATFORMS, FORECASTING, CONSUMER PREFERENCES, DECISION OPTIMIZATION.

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут інформаційних технологій

ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня магістра

Направляється здобувач Гомоляко В.І. до захисту кваліфікаційної роботи

за спеціальністю *124 Системний аналіз*

освітньо-професійної програми Інтелектуальні системи управління

на тему: «Модель попиту на предмети споживання засобами інтелектуального аналізу даних».

Кваліфікаційна робота і рецензія додаються.

Директор ННІІТ _____ Катерина НЕСТЕРЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

Здобувач Валерій ГОМОЛЯКО представив кваліфікаційну роботу, присвячену створенню моделі попиту на предмети споживання засобами інтелектуального аналізу даних. Ця тема є надзвичайно актуальною в умовах сучасного ринку, оскільки дозволяє підприємствам підвищити точність прогнозування попиту та оптимізувати управління запасами. Робота виконана на високому науковому рівні, чітко структурує основні теоретичні та практичні аспекти прогнозування попиту з використанням методів машинного навчання та аналізу великих даних. Автор обґрунтував вибір досліджуваної проблеми, визначив мету та завдання роботи, а також розробив концептуальну модель попиту на споживчі товари на основі інтелектуального аналізу даних. Пояснювальна записка відповідає встановленим стандартам оформлення, а зміст роботи відповідає обраній темі та чітко відображає досягнення поставленої мети.

Все це дозволяє оцінити виконану кваліфікаційну роботу здобувача Валерія Гомоляко на оцінку «_____» та присвоїти йому кваліфікацію магістр з системного аналізу.

Керівник кваліфікаційної роботи _____ Ганна ЛИХОДЄЄВА
(підпис)
«____» _____20___року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач *Гомоляко В.І.* допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедрою ІСТ _____ Каміла СТОРЧАК
(назва) *(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну роботу
на здобуття освітнього ступеня магістра

здобувача вищої освіти Гомоляко Валерій Ігорович
на тему «Модель попиту на предмети споживання засобами інтелектуального аналізу даних»

Актуальність. Актуальність теми кваліфікаційної роботи обумовлена зростаючою необхідністю ефективного прогнозування попиту на споживчі товари з використанням сучасних методів інтелектуального аналізу даних. В умовах швидко змінюваного ринку та інтенсивної конкуренції, точне прогнозування попиту дозволяє підприємствам оптимізувати виробництво, управління запасами та маркетингові стратегії. Розробка моделі попиту на основі машинного навчання та великих даних є важливим кроком у підвищенні ефективності управлінських рішень. Зміст роботи відповідає поставленим завданням та меті, розкриває обраний об'єкт та предмет дослідження.

Позитивні сторони.

У кваліфікаційній роботі автором детально розглянуто сутність попиту на предмети споживання, а також важливість використання інтелектуальних систем для аналізу великих даних у процесі прогнозування.

Особлива увага в роботі приділена розробці концептуальної моделі попиту на основі методів машинного навчання та нейронних мереж. Автор наводить конкретні приклади використання інтелектуальних систем для покращення точності прогнозування попиту та оптимізації процесу управління запасами, що є важливим для підвищення рентабельності підприємств.

Недоліки.

1. У роботі реалізовано модель попиту за допомогою обраної платформи для обробки даних, що може мати обмеження для підприємств з обмеженим обсягом даних або низькою кількістю користувачів. Ці обмеження можуть вплинути на ефективність моделі в певних умовах.
2. Робота містить незначні недоліки у форматуванні та стилістичні помилки, що не впливають на загальну якість роботи.

Відзначені зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи магістерської.

Висновок: *кваліфікаційна робота на здобуття ступеня магістра заслуговує оцінку "_____", а здобувач Валерій ГОМОЛЯКО заслуговує присвоєння кваліфікації магістр з системного аналізу.*

Рецензент:

науковий ступінь, вчене звання	*підпис*	*Ім'я, ПРІЗВИЩЕ*

ЗМІСТ

ВСТУП

Актуальність дослідження. У сучасному світі споживчий ринок є надзвичайно динамічним і складним середовищем, де ключовими елементами успіху стають глибоке розуміння поведінки споживачів і ефективне прогнозування попиту на товари та послуги. Стрімкий розвиток інформаційних технологій, зокрема методів інтелектуального аналізу даних (data mining), відкриває нові можливості для виявлення прихованих закономірностей у великих масивах даних. Використання таких підходів дозволяє компаніям адаптувати свою діяльність до змін ринку, знижувати ризики, підвищувати ефективність управління та приймати обґрунтовані рішення.

Актуальність дослідження зумовлена необхідністю створення інструментів для точного прогнозування попиту на предмети споживання в умовах нестабільного економічного середовища, змінних споживчих уподобань і посилення конкуренції. Традиційні методи аналізу часто не забезпечують достатньої гнучкості та точності, особливо коли йдеться про великі обсяги даних і складні багатофакторні залежності. Інтелектуальний аналіз даних, завдяки своїм можливостям в автоматизованому виявленні патернів і тенденцій, стає незамінним інструментом для вирішення цих завдань.

Розробкою даної проблеми займалося багато науковців, представники різних галузей науки. Вплив Data mining на суспільні процеси та організацію бізнесу зокрема досліджували Н. Бессіс, М. Чен, В. Майєр-Шенбергер, К. Кукер та інші. Етичні проблеми, що виникають внаслідок формування великих масивів персональної інформації, доступної для компаній, вивчали К. Девіс, Т. Крейг та М. Ладлофф.

Але, незважаючи на це, сьогодні існує потреба у дослідженні, яке б узагальнило, систематизувало існуючі відомості з даної проблеми.

Враховуючи все вищесказане, нами і була обрана тема магістерської роботи: "Модель попиту на предмети споживання засобами інтелектуального аналізу даних".

Об’єкт дослідження – процеси прогнозування та аналізу попиту на предмети споживання.

Предмет – моделі та методи прогнозування попиту на предмети споживання із використанням інструментів інтелектуального аналізу даних.

Мета роботи - створення ефективної моделі прогнозування попиту на предмети споживання на основі інтелектуального аналізу даних, яка дозволяє підвищити точність прогнозів та оптимізувати управління запасами.

Відповідно до мети були визначені наступні **завдання**:

1. Провести аналіз сучасних теоретичних підходів до моделювання попиту на предмети споживання.
2. Вивчити можливості інструментів інтелектуального аналізу даних для застосування у прогнозуванні попиту.
3. Розробити концептуальну модель попиту на предмети споживання, що базується на методах інтелектуального аналізу даних.
4. Здійснити апробацію розробленої моделі на реальних даних.
5. Оцінити ефективність запропонованої моделі та визначити її практичну значущість.

Для розв’язання поставлених завдань нами були використані такі **методи дослідження**: аналіз та синтез для вивчення теоретичних засад і сучасних підходів до прогнозування попиту; методи інтелектуального аналізу даних, зокрема кластеризація, регресійний аналіз, дерева рішень та нейронні мережі; методи математичного моделювання для побудови та оцінки точності прогнозів; експериментальні дослідження для апробації моделі на реальних даних.

Наукова новизна дослідження полягає у розробці та впровадженні інноваційної моделі прогнозування попиту на предмети споживання, яка

базується на комплексному використанні методів інтелектуального аналізу даних. Вперше було інтегровано підходи кластеризації, регресії та нейронних мереж для виявлення прихованих закономірностей у споживчій поведінці та оцінки багатофакторних впливів на попит у динамічному середовищі.

Практичне значення дослідження полягає у можливості використання розробленої моделі для підвищення точності прогнозування попиту у різних галузях економіки. Запропоновані підходи дозволяють підприємствам оптимізувати управління запасами, знижувати витрати на логістику та маркетинг, а також оперативно адаптуватися до змін ринкової кон'юнктури, що сприяє підвищенню конкурентоспроможності.

Робота може бути використана студентами ВНЗ для підготовки до семінарських занять, також може бути використана викладачами для проведення лекції, практик тощо.

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел, що містить 61 найменування. Повний обсяг роботи: 110 сторінок.

РОЗДІЛ 1. ЗАГАЛЬНО-ТЕОРЕТИЧНА ХАРАКТЕРИСТИКА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

1.1. Поняття та сутність інтелектуального аналізу даних

Інтелектуальний аналіз даних має свої корені ще в 90-х роках XX століття, коли розпочався розвиток технологій для підтримки прийняття управлінських рішень. Перші концептуальні визначення цієї галузі були запропоновані Г. П’ятецьким-Шапіро у 2196 році, який вважав, що ІАД є процесом автоматичного виявлення прихованих закономірностей і взаємозв'язків між різними змінними в обсягах необроблених даних, що включає класифікацію, моделювання та прогнозування. У науковій сфері термін ІАД часто використовується як синонім до англійського "Data Mining" (видобуток даних), і більшість дослідників вважають цей переклад найбільш вдалим для визначення концепції інтелектуального аналізу даних. Водночас, за оцінками міжнародних фахівців, ІАД має бути однією з основних рушійних сил соціально-економічних процесів впродовж наступних 5-10 років [34, с. 35].

Розвиток Data Mining був зумовлений кількома важливими факторами, серед яких: вдосконалення програмного забезпечення для управлінських процесів, поліпшення технологій зберігання та обробки даних, можливість накопичення великих обсягів інформації у базах даних, а також удосконалення алгоритмів для обробки цих даних. Ці чинники сприяли створенню інтелектуальних інформаційних систем, що активно підтримують управлінські рішення на підприємствах, зокрема, у плануванні попиту на споживчі товари та прогнозуванні господарської діяльності.

Методи ІАД реалізуються через концепцію шаблонів (патернів), що являють собою приховані або неочевидні закономірності в даних. Ці шаблони можуть відображати складні взаємозв'язки між різними елементами економічного процесу. Пошук шаблонів здійснюється за допомогою емпіричних

припущень про структуру вибірки та значення показників. Використання ІАД дозволяє значно прискорити обробку та пошук необхідних даних [26, с. 49].

Нерідко в процесі обробки сирих даних формується цілий комплекс цінної інформації, що дозволяє створювати аналітичні системи, які зібрані, зберігають та моніторять дані, перетворюючи їх у знання, доступні для кінцевого користувача.

Ґрунтовні дослідження (табл. 1.1) засвідчили, що існує два підходи до трактування змісту а призначення ІАД.

Таблиця 1.1.

Наукові підходи щодо трактування поняття інтелектуальний аналіз даних (Data Mining) в сучасній науковій літературі

Автор, джерело	Зміст поняття	Основне призначення
Як методів отримання та використання масиву даних для опису, а в подальшому прогнозування економічних явищ та процесів		
Ю. Романова	Data Mining – це сукупність великого числа різних методів виявлення знань, в основі якого лежить математичний апарат, який виник і розвивається на базі досягнень прикладної статистики, розпізнавання образів, методів штучного інтелекту, теорії баз даних і т.д.	Вибір методу часто залежить від типу наявних даних і від того, яку інформацію намагаються отримати
В. Плескач, Т. Затонацька	Data Mining – це технологія, призначена для пошуку у великих інформаційних масивах даних неочевидних, об'єктивних, корисних на практиці закономірностей.	ІАД здійснюється за допомогою використання технологій розпізнавання шаблонів, а також статистичних і математичних методів [10]
Як технологія систем підтримки прийняття рішень		

Г. ПіатецкийШапіро	Data Mining – це процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності.	Стрімке накопичення даних; загальна комп'ютеризація бізнеспроцесів; проникнення Інтернет в усі сфери діяльності; прогрес в області інформаційних технологій [10]
О. Колодчак	Інтелектуальний аналіз даних – це процес виявлення в сирих даних раніше невідомих, нетривіальних, фактично корисних і доступних інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності.	
Ю. Фаяд	Інтелектуальний аналіз даних (ІАД) або data mining (discovery-driven data mining) – це процес виявлення у первинних даних раніше невідомих, доступних, практично корисних і нетривіальних інтерпретацій знань, необхідних для прийняття рішень у різних сферах людської діяльності.	Візуальні інструменти ІАД дозволяють проводити аналіз даних предметними фахівцями, що не володіють відповідними математичними знаннями
Л. Грабовецький	Інтелектуальний аналіз даних (ІАД), або datamining (discovery-driven data mining), – це процес виявлення у первинних даних раніше невідомих, доступних, практично корисних і нетривіальних інтерпретацій знань, необхідних для прийняття рішень у різних сферах людської діяльності.	

Перший підхід полягає в тому, що науковці розглядають методи аналітичного аналізу даних як універсальні інструменти для збору та використання великих обсягів інформації, що дозволяють описувати економічні явища та здійснювати прогнозування різних процесів [41, с. 11].

Однак, на наш погляд, більш точне розуміння методів інтелектуального аналізу даних надає трактування ІАД як технології для створення систем підтримки управлінських рішень. Серед представників цього підходу слід виділити Чорноуса Г. та Рибальченка С., які стверджують, що для ефективного функціонування аналітичних систем у сучасній економіці необхідно впроваджувати нові інструменти для методичної, модельної та комп'ютерної підтримки прийняття рішень. Вони вважають, що нові можливості такої підтримки пов'язані з ефективним використанням інформаційних ресурсів та застосуванням технологій для виявлення знань у великих масивах даних, обсяги яких зростають дуже швидко [19, с. 53].

Отже, інтелектуальний аналіз даних є важливим інструментом для ефективного збору, обробки та використання великих обсягів даних з метою опису, аналізу та прогнозування економічних процесів. Перший підхід розглядає ці методи як універсальні засоби для прогнозування, однак більш глибоке розуміння їхнього застосування дає концепція ІАД як технології для підтримки управлінських рішень. Сучасні аналітичні системи вимагають впровадження нових методів і інструментів для адаптації до швидко зростаючих обсягів інформації та забезпечення ефективного використання цих ресурсів. Таким чином, роль ІАД в економіці полягає не лише в аналізі даних, але й у створенні систем, які підтримують ухвалення рішень на всіх рівнях управління, підвищуючи ефективність управлінських процесів у сучасних умовах.

1.2. Методи кластеризації, регресії та прогнозування

Інтелектуальний аналіз даних полягає у виявленні корисних відомостей з великих обсягів інформації. Для цього використовуються методи математичного аналізу, що дозволяють виявити приховані закономірності та тенденції в даних. Такі зв'язки часто неможливо помітити при звичайному перегляді, оскільки вони можуть бути занадто складними або через величезний обсяг інформації.

Знання, отримані в процесі цього аналізу, повинні бути новими та неочевидними. Це означає, що вони не можуть бути виявлені шляхом простого огляду даних, і повинні описувати зв'язки між характеристиками бізнес-об'єктів, передбачати певні значення на основі інших даних тощо. Крім того, ці знання повинні бути придатними для застосування до нових даних чи ситуацій [8, .с 15].

Практична цінність отриманих знань полягає в їх можливому використанні для підтримки прийняття рішень та покращення діяльності організацій. Виявлені закономірності можуть бути об'єднані в модель інтелектуального аналізу даних, яка може бути використана в різних ситуаціях, таких як:

1. Прогнозування: оцінка продажів, передбачення навантаження на сервери або час їх простою.
2. Оцінка ризику і ймовірностей: вибір найбільш підходящих клієнтів для цільових кампаній, визначення точок рівноваги для ризикованих ситуацій, передбачення ймовірностей діагнозів чи інших результатів.
3. Рекомендації: визначення продуктів, які можуть бути продані разом, або створення персоналізованих рекомендацій для користувачів.
4. Пошук послідовностей: аналіз покупок клієнтів і прогнозування їх наступних дій.
5. Групування: класифікація клієнтів або подій в кластери, що дозволяє виявити спільні риси для подальшого аналізу і прогнозування.

Ці моделі допомагають організаціям приймати обґрунтовані рішення і оптимізувати їх діяльність.

Завдання інтелектуального аналізу даних іноді позначають термінами "закономірності" або "техніки", хоча загальноприйнятої класифікації цих завдань не існує. Більшість авторитетних джерел виділяють такі основні задачі: класифікація, кластеризація, прогнозування, асоціація, візуалізація, виявлення аномалій, оцінка, аналіз взаємозв'язків, підсумовування результатів [22, с. 65].

Класифікація є найпростішою і найпоширенішою задачею інтелектуального аналізу. У її процесі визначаються характеристики, які дозволяють поділити об'єкти даних на групи або класи. Використовуючи ці ознаки, нові об'єкти можна віднести до певного класу. Для вирішення задачі класифікації застосовуються різні методи, зокрема: метод найближчого сусіда, метод k-найближчого сусіда, баєсівські мережі, індукція дерев рішень, нейронні мережі [22].

Кластеризація є складнішою задачею, що розвиває ідею класифікації. Особливістю кластеризації є те, що класи об'єктів не визначені заздалегідь. Завдання кластеризації полягає у розподілі об'єктів на групи на основі їх подібності. Один з методів розв'язання цієї задачі - навчання "без вчителя" нейронних мереж, таких як самоорганізовані карти Кохонена [22, с. 30].

Асоціація передбачає пошук закономірностей між подіями, що відбуваються в наборі даних. Зазначену задачу відрізняє те, що зв'язки шукаються не на основі властивостей об'єкта, а між кількома подіями, які відбуваються одночасно. Одним з найбільш відомих алгоритмів для цієї задачі є алгоритм Apriori.

Послідовність або послідовна асоціація фокусується на виявленні тимчасових закономірностей між подіями. Ця задача схожа на асоціацію, але з тією різницею, що події відбуваються з певним інтервалом часу. Визначення таких закономірностей дозволяє прогнозувати наступні події в часі. Прикладом є правило: після покупки квартири покупці, ймовірно, придбають холодильник протягом двох тижнів або телевізор через два місяці. Такі методи широко використовуються в маркетингу та управлінні взаємодією з клієнтами для прогнозування їх наступних кроків.

Прогнозування - це процес оцінки майбутніх або пропущених значень цільових числових показників на основі аналізу історичних даних. Для цього часто використовуються методи математичної статистики, нейронні мережі та

інші техніки. Завдання виявлення відхилень передбачає знаходження аномальних даних, які значно відрізняються від загальної маси, і аналіз таких "нехарактерних" шаблонів.

Оцінка полягає у передбаченні неперервних значень ознак. Аналіз зв'язків фокусується на виявленні залежностей між різними елементами в даних.

Візуалізація полягає у створенні графічного зображення даних для кращого розуміння їхніх закономірностей. Використовуються різні графічні методи, зокрема двовимірні та тривимірні зображення.

Підсумовування передбачає опис конкретних груп об'єктів, що належать до певного набору даних [41, с. 54].

Процес побудови моделей інтелектуального аналізу даних є частиною ширшого етапу, що включає розв'язання задач на кожному етапі, від формулювання запитань до розгортання моделей. Цей процес зазвичай включає шість основних етапів:

1. Постановка задачі.
2. Підготовка даних.
3. Дослідження даних.
4. Побудова моделей.
5. Тестування та перевірка моделей.

Перший етап передбачає чітке визначення проблеми та пошук шляхів використання даних для її вирішення. Це включає аналіз вимог бізнесу, визначення метрик для оцінки моделей та завдань проєкту.

Другий етап - це обробка та очищення даних, які були визначені на попередньому етапі. Третій етап передбачає глибоке вивчення даних для прийняття обґрунтованих рішень при створенні моделей. Це включає обчислення мінімальних, максимальних значень, середнього, стандартного відхилення та аналіз розподілу даних. Аналіз цих показників дозволяє зрозуміти, чи є дані репрезентативними, і чи потрібно додати нові чи змінити припущення для

коректних результатів. Стандартне відхилення, зокрема, може вказувати на необхідність додавання нових даних для покращення моделі.

Якщо виявляються проблеми в даних, наприклад, помилки чи відсутні значення, то це може спричинити розробку стратегії для виправлення ситуації. Для визначення доступності даних та їх аналізу можна використовувати різні інструменти, такі як Служби Master Data Services або SQL Server Data Quality Services, що допомагають виявити й усунути проблеми в даних.

Четвертий етап процесу інтелектуального аналізу даних включає створення моделей для обробки даних. Знання, отримані під час попереднього етапу "Огляд даних", дозволяють сформулювати необхідні моделі. Спочатку модель інтелектуального аналізу даних є лише структурним шаблоном, який визначає стовпці для вхідних даних, прогнозовані атрибути та параметри для управління алгоритмом обробки. Сам процес обробки моделі часто називається навчанням, що включає застосування математичного алгоритму до структурованих даних для виявлення закономірностей. Виявлені закономірності залежать від вибору даних для навчання, алгоритму та його налаштувань [25 ,с. 88].

П'ятий етап полягає в перевірці ефективності побудованих моделей. Існує кілька способів оцінки якості та характеристик моделей:

1. Використання статистичних методів для виявлення можливих проблем у даних або самій моделі.

2. Поділ даних на тренувальний та перевірочний набори для оцінки точності прогнозів.

3. Консультації з експертами для аналізу результатів моделі та визначення, чи є виявлені закономірності корисними для конкретного бізнес-сценарію.

Ці підходи є частиною методології інтелектуального аналізу даних і використовуються на різних етапах створення, перевірки та коригування моделі, залежно від виявлених проблем. Немає єдиного правила, яке дозволяє

однозначно визначити, які моделі можна вважати достатньо точними, і чи є дані для цього придатними.

1.3. Використання інтелектуального аналізу даних у сфері споживчого попиту

Варто зауважити, що економічний і соціальний розвиток сучасного суспільства тісно пов'язаний із зростанням споживчого попиту, який виступає ключовим показником динаміки ринків та інновацій. Завдяки швидкому впровадженню технологій у різних сферах, інтелектуальний аналіз даних набув особливого значення у визначенні, моделюванні та прогнозуванні споживчих тенденцій.

Сьогодні бізнес-середовище стикається з численними викликами, такими як зміна уподобань клієнтів, зростаюча конкуренція та необхідність адаптації до швидкого розвитку технологій. У відповідь підприємства активно використовують інструменти інтелектуального аналізу даних, щоб забезпечити конкурентоспроможність, оптимізувати процеси і краще зрозуміти потреби ринку.

Сфера споживчого попиту, що охоплює виробництво, розподіл і реалізацію товарів широкого вжитку, є надзвичайно динамічною та генерує величезні обсяги даних. Ці дані надходять з різноманітних джерел, таких як онлайн-покупки, соціальні мережі, споживчі опитування, транзакційні системи та інші канали. Використовуючи методи інтелектуального аналізу даних, бізнес може ідентифікувати ключові закономірності, прогнозувати поведінку споживачів і виявляти нові можливості для розвитку [35, с. 87].

Особливу роль у цьому відіграють такі технології, як Big Data, Data Mining та штучний інтелект. Завдяки їм компанії можуть обробляти великі масиви інформації, забезпечуючи персоналізацію пропозицій, вдосконалення логістичних ланцюгів та розробку нових продуктів. Наприклад, методи Data

Mining дозволяють виявляти приховані залежності між характеристиками товарів і вподобаннями клієнтів, сприяючи підвищенню точності маркетингових стратегій.

Інтелектуальний аналіз даних стає важливим інструментом для формування споживчих рішень, забезпечуючи ефективне управління попитом і пропозицією. Підприємства, що впроваджують ці технології, мають значні конкурентні переваги, оскільки можуть оперативно реагувати на зміни ринку, оптимізувати витрати та збільшувати рівень задоволення клієнтів.

У межах цієї роботи буде розглянуто основні підходи до використання інтелектуального аналізу даних для моделювання споживчого попиту, їхні переваги та недоліки, а також проаналізовано сучасні кейси успішного впровадження цих методик [42 ,с. 78].

Інтеграція методів інтелектуального аналізу даних у сферу споживчого попиту відкриває нові можливості для оптимізації бізнес-процесів і прийняття стратегічних рішень. Застосування цих технологій дозволяє ефективно працювати з великими обсягами даних, отриманих із різних джерел, таких як онлайн-платформи, транзакції, соціальні мережі та опитування, для виявлення ключових тенденцій і прогнозування попиту.

1. Оптимізація управління запасами і постачанням. Інтелектуальний аналіз даних дозволяє аналізувати історичні дані та прогнозувати зміни попиту, що сприяє вдосконаленню управління запасами та логістикою. Методи кластеризації та визначення правил асоціації допомагають підприємствам визначити оптимальний рівень товарних запасів, ідентифікувати вузькі місця в постачанні та ефективніше співпрацювати з постачальниками. Наприклад, прогностичні моделі можуть передбачити сезонні коливання попиту, що дозволяє мінімізувати втрати та покращити розподіл ресурсів.

2. Розуміння споживчої поведінки. Методи інтелектуального аналізу даних дозволяють ідентифікувати закономірності в поведінці споживачів,

розділяти клієнтів на сегменти та адаптувати маркетингові стратегії. Завдяки аналізу настроїв та прогнозному моделюванню компанії отримують можливість створювати персоналізовані пропозиції, підвищувати лояльність клієнтів і передбачати зміни у вподобаннях. Наприклад, аналіз відгуків споживачів у соціальних мережах дозволяє виявити фактори, які впливають на вибір продукції.

3. Аналіз ринкового кошика. Використовуючи методи асоціативного аналізу, можна виявляти зв'язки між товарами, які споживачі зазвичай купують разом. Це дає змогу розробляти ефективні стратегії перехресних продажів і акцій. Наприклад, аналіз даних може показати, що покупці, які купують певний бренд напоїв, часто купують і відповідні закуски, що відкриває можливості для комбінованих пропозицій.

4. Прогнозування тенденцій попиту. Аналіз часових рядів дозволяє прогнозувати динаміку попиту, враховуючи сезонність, економічні фактори та поведінкові зміни споживачів. Це дає змогу компаніям завчасно планувати випуск нових товарів або коригувати обсяги виробництва. Наприклад, виробники можуть використовувати прогностичні моделі для передбачення пікових періодів продажів і відповідної підготовки ресурсів [43, .с 56].

5. Сегментація споживачів. Кластеризація допомагає визначати групи клієнтів із подібними характеристиками та поведінкою. Це дозволяє адаптувати пропозиції для різних сегментів, підвищуючи їхню привабливість. Наприклад, роздрібні мережі можуть використовувати ці методи для виявлення груп клієнтів, які найбільш схильні купувати товари преміум-сегменту.

Попри численні переваги, використання інтелектуального аналізу даних для прогнозування попиту на предмети споживання стикається з низкою перешкод, які необхідно враховувати наступні фактори.

Дані, зібрані з різних джерел (інтернет-магазини, касові системи, соціальні мережі), можуть мати різні формати та рівень деталізації. Відсутність стандартизації, суперечності в даних або їх недостатність створюють ризик

отримання неточних результатів аналізу. Інтеграція таких даних в єдину систему для обробки вимагає значних ресурсів і технологічних зусиль.

Робота з даними споживачів часто викликає питання етики та захисту персональних даних. Неправомірний збір або використання даних може призвести до втрати довіри клієнтів і юридичних наслідків. Для успішного впровадження необхідно забезпечити дотримання чинного законодавства та розробити прозору політику конфіденційності.

Успішне впровадження інтелектуального аналізу даних вимагає участі фахівців із відповідними знаннями у сфері Data Science, машинного навчання та статистики. Нестача кваліфікованих кадрів є серйозним викликом для компаній, які прагнуть впровадити ці технології. Навчання співробітників і залучення експертів стає важливим етапом у процесі цифрової трансформації [23 с. 88].

У майбутньому використання інтелектуального аналізу даних у сфері споживчого попиту буде розвиватися завдяки новим технологічним досягненням:

1) Синергія з IoT та аналітикою великих даних.
2) Автоматизація бізнес-процесів.
3) Сталий розвиток.

Інтелектуальний аналіз даних стає невід'ємною частиною управління споживчим попитом, дозволяючи компаніям глибше розуміти клієнтів, адаптуватися до змін ринку та залишатися конкурентоспроможними. Подолання викликів, таких як інтеграція даних, дотримання етичних стандартів і підготовка кваліфікованих фахівців, сприятиме ширшому впровадженню цих технологій у різні галузі.

Продовження досліджень у сфері інтелектуального аналізу даних і розвиток відповідних технологій дозволять забезпечити стійкий розвиток економіки, що орієнтується на інновації та потреби сучасних споживачів.

Сьогодні обробка даних стає важливим інструментом для розуміння й аналізу споживчого попиту. Завдяки методам інтелектуального аналізу даних відкриваються нові можливості для вивчення поведінки споживачів, прогнозування попиту на товари та підвищення ефективності управління запасами.

Одним із ключових аспектів аналізу споживчого попиту є робота з великими обсягами даних, які надходять із різноманітних джерел, таких як соціальні мережі, онлайн-опитування, дані з касових апаратів, мобільні додатки та інтернет-магазини. Ці дані мають неоднорідну структуру, різний рівень деталізації та можуть включати як структуровану, так і неструктуровану інформацію [35 ,с. 67].

Інтелектуальний аналіз даних дозволяє структурувати й оцінювати великі обсяги інформації. Наприклад, через використання кластеризації даних можна виявити групи споживачів із подібними уподобаннями, а методи прогнозного моделювання допомагають визначити, які товари будуть популярні в майбутньому.

Методи інтелектуального аналізу включають машинне навчання, статистичні підходи та алгоритми обробки великих даних. Серед найбільш поширених технік:

1) Аналіз трендів;
2) Просторова кластеризація;
3) Виявлення аномалій;

Інтелектуальний аналіз даних у сфері споживчого попиту дозволяє приймати обґрунтовані рішення. Наприклад, аналіз сезонних коливань попиту дає змогу заздалегідь формувати стратегії постачання, мінімізуючи втрати через надлишкові або недостатні запаси.

Крім того, дані про поведінку споживачів допомагають розробляти індивідуалізовані пропозиції, що підвищує рівень задоволеності клієнтів і стимулює лояльність.

Використання інтелектуального аналізу даних у сфері споживчого попиту сприяє створенню моделей, які дозволяють передбачати попит із високою точністю, ефективно управляти ресурсами й адаптуватися до змін у споживчих перевагах. Ці технології відкривають нові перспективи для підприємств, спрямованих на задоволення потреб сучасного споживача, і водночас сприяють підвищенню ефективності ринку в цілому [24 ,с. 87].

У сучасному світі аналіз споживчого попиту стає критично важливим для забезпечення ефективного планування та прогнозування. Одним із найефективніших підходів у цьому контексті є застосування інтелектуального аналізу даних, який дозволяє ідентифікувати закономірності та зв’язки в складних наборах даних, отриманих із різних джерел.

Один із ключових підходів до аналізу – це виявлення співіснування певних моделей попиту. Наприклад, за допомогою аналізу асоціацій можна виявити зв’язки між придбанням певних товарів чи послуг, що вказує на потенційні взаємозалежності в споживчій поведінці.

Інший підхід – аналіз послідовності, який дозволяє простежувати динаміку змін у попиті в часі та просторі. Це дає змогу визначити, як зміни в зовнішніх умовах або в поведінці споживачів впливають на попит на певні категорії товарів.

Класифікація також є важливим інструментом для аналізу. Вона дозволяє групувати споживачів або товари за певними ознаками, такими як демографічні характеристики, поведінкові особливості або рівень витрат. Це сприяє створенню сегментованих моделей попиту, які є більш точними і релевантними.

Одним із викликів інтелектуального аналізу даних є складність обробки великих наборів інформації. Тому застосовуються методи зменшення розмірності даних, які дозволяють виділити ключові параметри, що впливають

на попит, без втрати критичної інформації. Наприклад, алгоритми кластеризації допомагають зменшити кількість параметрів, дозволяючи сконцентруватися на найважливіших аспектах.

Ключова мета таких методів – досягнення інтерпретованих результатів у прийнятний час, враховуючи обмеження обчислювальних ресурсів. Завдяки цьому вдається отримати чіткі й зрозумілі висновки навіть за умов роботи з великими обсягами даних.

Для підвищення зрозумілості результатів аналізу активно використовуються сучасні інструменти візуалізації даних. Це дозволяє представити взаємозв'язки між ключовими показниками попиту у вигляді графіків, діаграм і теплових карт, що полегшує ухвалення управлінських рішень.

Застосування інтелектуального аналізу даних у дослідженні споживчого попиту дозволяє значно підвищити точність прогнозів і адаптувати пропозицію до реальних потреб ринку. В умовах зростаючого обсягу інформації, що надходить із різних джерел, ці методи стають невід'ємною складовою ефективного управління ресурсами та планування [35 ,с. 67].

На рисунку 1 представлено узагальнену схему обробки даних для створення моделей попиту, яка демонструє основні етапи інтеграції, аналізу та візуалізації інформації.

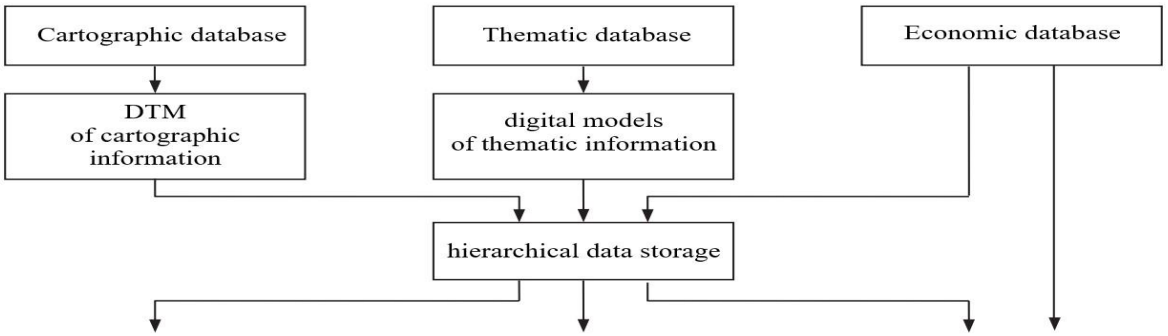

Рис. 1.1 Технологія обробки даних в ГІС

Методи інтелектуального аналізу даних можна умовно поділити на дві основні категорії: попередній аналіз і спеціалізовані підходи. Попередній аналіз даних має на меті підготовку інформації для подальших досліджень і включає завдання очищення даних, нормалізації, заповнення пропущених значень, виявлення аномалій, а також проведення початкової візуалізації для виявлення основних закономірностей. Ці методи є базовим етапом обробки даних, який забезпечує якісну основу для наступних більш складних операцій [40 ,с .45].

До спеціалізованих підходів належать алгоритми, які вирішують конкретні задачі, такі як прогнозування попиту на певні категорії товарів, класифікація споживачів за уподобаннями, сегментація ринку за поведінковими ознаками та виявлення прихованих взаємозв'язків між характеристиками споживання. Використання сучасних технологій, таких як машинне навчання і нейронні мережі, дозволяє досягати високої точності в передбаченні попиту та адаптації моделей до змін у поведінці споживачів.

Окрему увагу заслуговує застосування моделей кластеризації для групування споживачів за схожими патернами покупок. Цей підхід допомагає визначити цільові сегменти ринку та формувати персоналізовані пропозиції, підвищуючи ефективність маркетингових кампаній. Наприклад, методи кластеризації на основі алгоритмів K-середніх або DBSCAN можуть бути використані для ідентифікації груп споживачів, які мають схожі вподобання або потреби.

Для аналізу попиту важливим є застосування методів часових рядів, які дозволяють прогнозувати сезонні коливання і тренди в споживанні. В таких дослідженнях широко використовуються алгоритми ARIMA, LSTM та інші моделі глибокого навчання. Вони забезпечують можливість не тільки передбачити майбутній попит, але й аналізувати вплив зовнішніх чинників, таких як економічні умови чи зміна споживчих звичок.

Таким чином, інтелектуальний аналіз даних стає потужним інструментом у дослідженні споживчого попиту, допомагаючи підприємствам і організаціям не тільки адаптуватися до змін у поведінці клієнтів, а й передбачати їхні майбутні потреби.

1.4. Інструменти для інтелектуального аналізу даних, порівняння основних платформ

Опис і реалізація аналізу за допомогою інтелектуального аналізу даних (ІАД) здійснюється через вибір відповідного методу та алгоритму для його реалізації. Різні наукові джерела вказують, що ці методи можна класифікувати за кількома критеріями, такими як метод навчання та підхід до моделювання. За першим критерієм задачі прогнозування, що використовують ІАД, можуть бути розв'язані за допомогою двох основних груп методів: навчання з учителем (Supervised Learning) та навчання без учителя (Unsupervised Learning). Крім того, класифікація методів ІАД може базуватись на розрізненні статистичних і кібернетичних підходів до аналізу даних [24, с. 390].

Кібернетичний експеримент є одним із основних методів для вирішення задач, де вже є знання про навчальну вибірку і необхідно виявити глибинні зв’язки між об'єктами, що досліджуються. Як зазначають деякі дослідники, «статистичні пакети включають елементи ІАД, але акцент у них робиться на класичних методах, таких як кореляційний, регресійний і факторний аналіз. Водночас, основним недоліком таких систем є необхідність спеціальної підготовки користувача, а також те, що вони часто опираються на статистичну парадигму, яка може бути не зовсім адекватною для аналізу реальних ситуацій» [5, с. 463]. Однак, на нашу думку, сучасний системний аналіз неможливий без застосування статистичних методів для попереднього відбору важливих даних, які стануть основою для подальших аналітичних досліджень. Після цього для

глибшого аналізу та виявлення закономірностей в даних слід використовувати методи ІАД та нестандартні підходи.

Що стосується прогнозування попиту, одним із найефективніших методів є використання моделей авторегресії та байєсівських мереж. Моделі авторегресії, такі як AR, ARIMA, ARIMAS, ARCH і GARCH, сьогодні активно застосовуються для прогнозування економічних процесів. Важливою особливістю багатьох часових рядів є наявність автокореляції, що дозволяє описати залежність між рівнями ряду (наприклад, yt від попередніх значень yt-1, yt-2 і т. д.) через авторегресійні функції. Моделювання попиту, який є нестаціонарним і залежить від різних факторів, потребує поєднання авторегресії з іншими методами, такими як ковзне середнє, тренди та сезонні хвилі. Це об'єднання моделей значно розширює їхнє застосування в практиці [16, с.87].

Байєсівські мережі (БМ), розроблені Д. Перлом у 2185 році, незважаючи на їхню відносну новизну, здобули популярність завдяки здатності визначати причинно-наслідкові зв'язки в складних процесах, які залежать від великої кількості факторів. Це робить їх дуже корисними для прогнозування попиту на споживчі товари, адже дозволяє будувати надійні моделі з високою точністю прогнозування.

У разі застосування байєсівських мереж, прогноз визначається через ймовірність потрапляння результату в заданий інтервал, який визначається на основі дискретизації початкових даних. Також є можливість прогнозувати напрямок розвитку процесів за допомогою нелінійних моделей байєсівської регресії, що дозволяє створювати індикатори для прогнозування фінансових процесів, таких як попит і пропозиція або купівля і продаж товарів.

Ці мережі застосовуються для розв’язання оптимізаційних завдань, де метою є максимізація або оптимізація факторів, що визначають рівень попиту на споживчі товари та їх економічний розвиток. Цільова функція, що забезпечує задані критерії, є окремим завданням, яке, на основі аналізу великих масивів

даних, точно описує функціональну залежність основних показників попиту. Це дозволяє побудувати модель попиту на товари в неявному вигляді, що відкриває нові можливості для визначення рівня оптимальності цієї моделі.

Сучасні вимоги до використання інтелектуального аналізу даних при прогнозуванні попиту вимагають здатності обробляти великі та різнорідні масиви інформації. Результати аналітичних оцінок повинні бути точними, зрозумілими для користувача, а інструменти для цього - доступними та зрозумілими. Тому використання ІАД у прогнозуванні попиту є необхідним, оскільки їх застосування не обмежене певною сферою, дозволяє швидко виявляти закономірності розвитку взаємозв'язків у даних (наприклад, виявлення конкурентних переваг підприємства або реакцій споживачів на акції). Це також включає технічну обробку даних, що сприяє отриманню об'єктивних прогнозів.

Однак, як зазначають критики, методи ІАД мають певні недоліки: вони потребують великих обсягів даних для успішного навчання, можуть формувати моделі у прихованій формі (так звану «чорну скриньку»), дають часті помилки, а також вимагають високої кваліфікації від користувачів. Незважаючи на це, успішне застосування ІАД у різних сферах, зокрема в економіці, є яскравим підтвердженням ефективності цього підходу [29, с. 58].

Незважаючи на переваги методу інтелектуального аналізу даних, варто відзначити, що вітчизняні підприємства все ще рідко застосовують ці методи для прогнозування попиту на споживчі товари. Натомість вони частіше використовують окремі програмні пакети для прогнозування соціально-економічних показників, які впливають на рівень попиту. Зазвичай підприємства обирають спеціалізовані статистичні інструменти та аналітичні системи, такі як SWOT-аналіз, методи дерева рішень та системи для автоматизації підприємницьких процесів.

Проте в умовах сучасної конкуренції інтелектуальний аналіз даних стає надзвичайно важливим для керівників і аналітиків, оскільки дає їм суттєві

переваги у боротьбі на ринку. Великі компанії не можуть конкурувати з малим бізнесом через більш персоналізований підхід до клієнтів, який використовують останні, грунтуючись на детальному аналізі їхніх переваг.

Для цього підприємства збирають всю доступну інформацію про своїх клієнтів (часто через OLTP-системи), узагальнюють дані з різних баз і систем за допомогою технологій обробки великих даних (ХД), а потім аналізують їх для отримання корисної інформації для бізнесу [17 ,с. 54].

Прогнозування попиту потребує створення інтелектуальних інформаційних систем, які формуються на основі сучасних підходів до збору, накопичення та моніторингу даних і їх перетворення на знання. Серед таких підходів особливе місце займають нейронні мережі, дейтамайнінг, програмні агенти та генетичні алгоритми.

Інтелектуальні методи аналізу даних, зокрема для прогнозування попиту на споживчі товари, є надзвичайно різноманітними і входять до складу сучасних корпоративних управлінських систем. Як правило, застосування методів ІАД включає використання комплексних аналітичних систем, що дозволяють прогнозувати зміну ключових показників попиту на товари або його рівень в цілому.

Таблиця 1.2.

Інформаційно-аналітичні системи, що включають здатність оцінювати та прогнозувати попиту на предмети споживання

Компанії-виробники	Функціональні компоненти систем							
	Сховище даних	**Бізнес-аналіз**	**Текстмайнінг**	**Підтримка прийняття Бюджетування,**	**планування Інтерпритація даних**	**ETL -технологія**	**Інтеграція даних**	
IBM (Cognos, SPSS, Applix, Celequest, Data Mirror, Adaytum, Frango, ILog, AptSoft)	+	+	+	+	+	+	+	+
Infor (Epiphany, Extensity, GEAC, MIS)	+	+		+	+	+		
Microsoft (FRx, ProClarity)	+	+	+			+		+
SAP (Business Object, Cartesis, Fuzzy, OutlookSoft, Pilot Software, Armstrong Laing, FirstLogic, SRC Software)	+	+	+	+	+	+	+	+
SAS (DataFlux)	+	+	+	+	+	+	+	+
Oracle (Hyperion, BEA, Sunopsis, Haley)	+	+	+		+	+	+	+

Таблиця 1.3.

Характеристика підходів щодо збирання, нагромадження та моніторингу інформації, що використовується у прогнозуванні попиту на предмети споживання

Назва підходу	Зміст
Дейтамайнінг	процес фільтрування значного обсягу даних з метою підбору інформації у контексті вирішення задачі. Ця інформація являє собою величезну цінність для управлінського апарату у їх повсякденній діяльності (керівника, аналітиків, менеджерів). У цьому процесі часто використовують такі програмні продукти PolyAnalist, MineSet, KnowledgeSTUDIO.

Нейронні мережі	програмно реалізовані системи, що реалізовуються через розробку математичних моделей процесу передавання і оброблення імпульсів мозку людини, котрі імітують тісноту взаємодії нейронів задля опрацювання інформації, що надходить, і навчання досвіду. У цьому процесі часто використовують такі програмні пакети як нейромережевий підхід, а саме NeuroShelL.
Генетичні алгоритми	різновид дейтамайнінгу. Для реалізації поставлених задач використовують такі програмні пакетами, як Evolver, GeneHunter, GeneticTraining Option. Їх застосування сприяє розширенню сфер застосування інтелектуальних систем. Для ефективного їх використання від користувачів системи потребують тільки початкову формалізацію задачі й формування множини вихідних даних.
Технологія програмних агентів	базується на використанні автономних програм, які автоматично виконують конкретні завдання з моніторингу ІС і збору інформації в мережах, діють від імені користувача для забезпечення бажаних результатів.
Сучасні програмні агенти,	це процес проведення спостереження, що включає різні вимірювання, розв'язування завдань щодо управління системними мережами. Сучасні інтелектуальні агенти здатні автоматизовувати численні операції керування мережами, можуть застосовуватися в наукомістких сферах соціальноекономічного розвитку для передачі повідомлень, підбору інформації, автоматизації процесів.

Кожен метод інтелектуального аналізу даних, застосовуваний для прогнозування, має свій набір закономірностей, що визначають його тип та відповідні завдання. Серед основних типів задач можна виділити такі, як класифікація, кластеризація, асоціація, послідовність чи послідовна асоціація, прогнозування, виявлення відхилень, оцінка, аналіз зв'язків, візуалізація, а також підсумовування для опису конкретних груп об'єктів на основі набору даних. Ці алгоритми набули особливої популярності в автоматизації сучасних бізнес-процесів.

Дослідження підтверджують, що ці методи є ефективними для оцінки і прогнозування попиту на споживчі товари, оскільки рівень попиту залежить від безлічі факторів, як внутрішніх, так і зовнішніх (якісних і кількісних). Важливими є також великі обсяги даних, які описують взаємозв'язки між цими факторами і можуть бути зібрані в єдину систему даних підприємства.

Висновки до розділу 1.

Таким чином, інтелектуальний аналіз даних охоплює широкий спектр методів, серед яких кластеризація, регресія та прогнозування займають провідні позиції. Кластеризація використовується для групування схожих об'єктів, регресійний аналіз допомагає встановлювати залежності між змінними, а прогнозування дає можливість передбачати майбутні значення. Ці методи є базовими для моделювання попиту на предмети споживання, оскільки вони забезпечують точність і надійність прогнозів.

Інтелектуальний аналіз даних активно використовується у сфері споживчого попиту для аналізу поведінки покупців, визначення трендів і адаптації до змін ринку. Завдяки цьому підприємства можуть знижувати ризики, оптимізувати запаси та підвищувати рівень обслуговування клієнтів. Аналіз інструментів для інтелектуального аналізу даних показав, що платформи, такі як Python, R, RapidMiner, KNIME та TensorFlow, забезпечують різні функціональні можливості, які слід обирати залежно від конкретних задач і рівня складності.

Таким чином, інтелектуальний аналіз даних є універсальним інструментом для вирішення складних економічних завдань. Його використання у сфері прогнозування споживчого попиту відкриває нові можливості для бізнесу, забезпечуючи ефективне управління даними та прийняття обґрунтованих рішень.

РОЗДІЛ 2. ОСОБЛИВОСТІ МОДЕЛЮВАНЯ ПОПИТУ НА ПРЕДМЕТИ СПОЖИВАННЯ ЗАСОБАМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1. Вибір методології моделювання

Прогнозування попиту є важливим інструментом для аналізу та передбачення потреб споживачів на різні товари та послуги в майбутньому. У контексті споживчого попиту, прогнозування здійснюється за допомогою аналізу наявних даних, що дозволяє визначити ймовірні зміни в поведінці споживачів на основі статистичних моделей.

Аналізуючи попит, важливо враховувати, що він відображає потребу споживачів у конкретному товарі або послузі, а також ураховує різні фактори, які можуть впливати на їхній вибір. Прогнозування попиту на споживчі товари використовує дані про вже здійснені покупки, переваги, поведінку на ринку та інші соціо-економічні показники, що допомагають спрогнозувати, які товари будуть користуватися попитом в майбутньому [34, с. 76].

Застосування прогнозування попиту в різних секторах економіки, таких як виробництво, енергетика або послуги, дозволяє компаніям і організаціям краще планувати свої ресурси, оптимізувати виробництво та мінімізувати витрати, що є важливим аспектом для збереження конкурентоспроможності.

Прогнозування попиту не тільки допомагає підприємствам прогнозувати кількість запитуваних товарів або послуг, але й дозволяє адаптувати стратегії продажу та маркетингу, що може призвести до значного підвищення ефективності бізнес-процесів.

Для досягнення найкращих результатів у прогнозуванні попиту, компанії використовують різні моделі інтелектуального аналізу даних, зокрема машинне навчання та алгоритми штучного інтелекту. Такі технології допомагають автоматизувати процеси аналізу даних, роблячи його більш точним та швидким,

а також здатними враховувати змінні фактори, які можуть впливати на попит на різні товари або послуги.

Основними перевагами прогнозування попиту є: зниження витрат на непотрібні ресурси, покращення ефективності планування виробничих та логістичних процесів, а також підвищення задоволення потреб споживачів через точніший підбір товарів чи послуг. В умовах швидкоплинного ринку це дозволяє компаніям не тільки зберігати стабільність, але й успішно реагувати на зміни в економічних і соціальних умовах.

Прогнозування попиту на споживчі товари є важливим етапом у створенні ефективних бізнес-стратегій для компаній, які прагнуть максимально точно передбачити потреби споживачів та адаптувати свої пропозиції до цих вимог. У цьому контексті застосування інтелектуального аналізу даних надає можливість ефективно опрацьовувати великі масиви інформації, що дозволяє отримати точні прогнози та уникнути непотрібних ризиків. Коли попит на певні товари зростає або падає, своєчасне прогнозування дає можливість компаніям коригувати стратегії та підтримувати високий рівень обслуговування споживачів [13 ,с. 67].

1. Поліпшення фінансового планування**.** Прогнозування попиту забезпечує підприємствам основу для розробки точних фінансових планів. Використовуючи отримані прогнози, компанії можуть більш ефективно управляти своїми коштами, планувати бюджетування та ресурсне забезпечення. Наприклад, точні прогнози дозволяють враховувати піки попиту, сезонні коливання, а також тренди, що можуть впливати на споживчі переваги. Це допомагає підприємствам уникати зайвих витрат і вчасно реагувати на зміни ринкових умов.

2. Зменшення кадрових проблем. Однією з ключових переваг прогнозування попиту є можливість точніше планувати потребу в людських ресурсах. Передбачення пікових періодів попиту дозволяє ефективно організувати робочі графіки та уникнути ситуацій, коли компанія стикається з

браком персоналу під час високих сезонів або навпаки – надлишком працівників у менш активні періоди. Це дозволяє компаніям не лише оптимізувати витрати на оплату праці, але й забезпечити стабільність в роботі.

3. Покращення маркетингових стратегій. Правильне прогнозування попиту допомагає адаптувати маркетингові стратегії до реальних потреб споживачів. Якщо компанія спостерігає зниження попиту на певний товар, прогнози можуть допомогти визначити оптимальний час для запуску рекламних акцій або коригування цін. Окрім того, використання прогнозних моделей дає змогу виявляти потенційно вигідні сегменти споживачів, що дозволяє таргетувати маркетингові кампанії більш ефективно [23 ,с. 67].

4. Оптимізація управління запасами. Прогнозування попиту також має вирішальне значення для покращення управління запасами. Знаючи, коли і в якій кількості очікується попит на продукцію, компанії можуть уникнути ситуацій із нестачею товарів або переповненням складів. Це дозволяє оптимізувати витрати на логістику та зберігання товарів, що є важливим аспектом в загальному управлінні підприємством. Точне прогнозування також допомагає спрогнозувати зміну попиту в конкретні періоди, що сприяє більш ефективному плануванню виробництва та постачання.

Загалом, прогнозування попиту за допомогою інтелектуального аналізу даних дозволяє компаніям не тільки покращити ефективність своїх операцій, а й забезпечити високий рівень задоволення потреб споживачів, що є критично важливим для довгострокового успіху на ринку.

Точне прогнозування споживчого попиту дозволяє підприємствам ефективно адаптувати свої стратегії до змінюваних умов ринку, мінімізуючи ризики, пов'язані з неочікуваними коливаннями попиту. Інтелектуальний аналіз даних дає змогу точно визначити, які товари або послуги будуть користуватися найбільшим попитом у майбутньому, що дозволяє оптимізувати ресурси, зменшити витрати та забезпечити безперебійну доставку продукції. Однією з

найбільших переваг прогнозування попиту є зниження рівня невизначеності, що дозволяє точніше планувати виробничі, фінансові та кадрові процеси.

Прогнозування попиту на споживчі товари дає змогу передбачати й адаптуватися до змін попиту, що є важливим для формування надійних бізнес-стратегій. Наприклад, точні прогнози дозволяють знизити ризики дефіциту або надлишку товарів, забезпечуючи їх належну доступність на ринку в потрібний час. Це також допомагає в розробці ефективних планів для маркетингових кампаній, розподілу ресурсів і керування ланцюгом постачань [40 с, .65].

Крім того, точне прогнозування попиту оптимізує планування робочої сили та ресурсу, дозволяючи зменшити витрати на персонал через точніше визначення необхідної кількості співробітників на різні періоди часу. Використання інтелектуальних методів прогнозування також дає можливість створювати більш вигідні умови для збільшення обсягів продажу та прибутку, оскільки підприємства можуть забезпечити своєчасну доставку товарів і послуг, що відповідають запитам споживачів.

Основні переваги прогнозування попиту для споживчого ринку можна підсумувати наступним чином:

- Підвищення доступності товарів і послуг, що веде до збільшення споживчого попиту.
- Оптимізація ланцюга поставок і зменшення витрат на зберігання запасів.
- Покращення планування бюджету, включаючи аспекти фінансування, маркетингових витрат та управління персоналом.
- Зменшення витрат на персонал за рахунок точної оптимізації змін.
- Підвищення загальної маржі за рахунок точнішого прогнозування і адаптації до реальних потреб ринку.

Прогнозування попиту може ґрунтуватися на трьох основних моделях, що враховують як внутрішні, так і зовнішні дані. Внутрішні показники включають

історичні дані про продажі, витрати на рекламу, трафік споживачів, тоді як зовнішні включають галузеві тенденції, вплив конкурентів та загальні економічні зміни. Компанії, що активно використовують ці дані, мають змогу значно підвищити свою конкурентоспроможність і прибутковість, забезпечуючи найкраще задоволення потреб споживачів.

Тим не менш, деякі підприємства, зокрема невеликі, можуть стикатися з обмеженнями у доступі до необхідних ресурсів або знань для здійснення ефективного прогнозування попиту. Проте, звернення до експертів та впровадження сучасних технологій аналізу даних може стати першим кроком до впровадження ефективних методів прогнозування [32, с. 55].

Існує кілька основних підходів до прогнозування споживчого попиту, кожен з яких має свої особливості та переваги. Ось три основні моделі:

1. Модель якісного прогнозування. Ця модель ґрунтується на якісних даних і часто використовує інтуїтивні оцінки. Джерела інформації можуть включати думки експертів, консультантів, груп споживачів, а також аналіз конкурентів. Вона підходить для ситуацій, коли компанія не має історичних даних або коли запуск нового товару чи послуги потребує попереднього аналізу без достатнього фактичного матеріалу. Такі моделі використовуються, коли необхідно передбачити попит на нові або нестандартні продукти, для яких неможливо застосувати кількісні дані.

2. Причинно-наслідкова модель. Ця модель передбачає, що попит може змінюватися під впливом різноманітних факторів, які можуть бути контрольованими або неконтрольованими. До контрольованих факторів належать маркетингові стратегії, ціноутворення, місцезнаходження товару чи послуги, рівень сервісу тощо. Натомість неконтрольовані фактори включають погодні умови, політичні зміни, конкуренцію, соціальні та економічні коливання, які важко передбачити. Через наявність великої кількості змінних, ця модель є досить складною для впровадження, але вона дозволяє враховувати

багатофакторні впливи на попит. Причинно-наслідкове прогнозування є особливо корисним для ринків, які зазнають змін, таких як політичні або економічні потрясіння [45 ,с. 78].

3. Аналіз часових рядів. Ця модель базується на використанні великої кількості кількісних даних, зокрема історії продажів, сезонних варіацій та змін цін. Метод аналізу часових рядів передбачає використання математичних моделей для передбачення майбутнього попиту, спираючись на минулі тенденції. Це дозволяє ефективно враховувати циклічні або сезонні коливання попиту.

Усі ці моделі можуть бути адаптовані для різних типів підприємств і використовуватись залежно від наявності даних та специфіки ринку. Крім того, інші методи, такі як аналіз тенденцій, графічні методи, сезонні коригування та моделювання життєвого циклу, можуть бути інтегровані для покращення точності прогнозів.

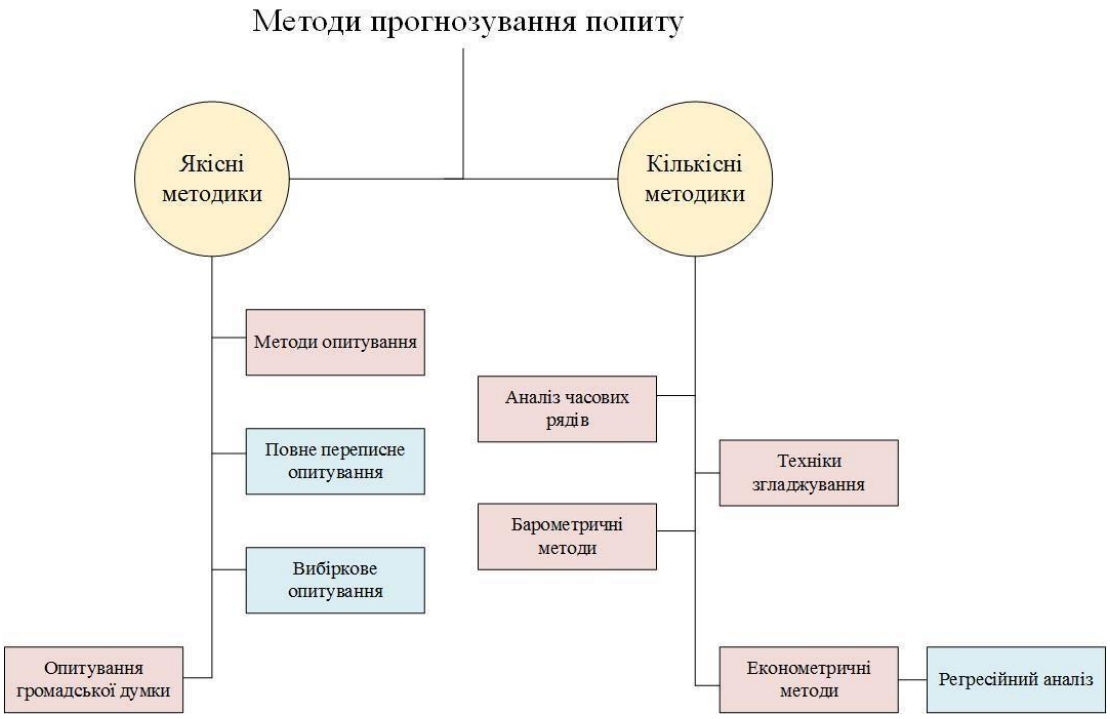

Рис. 2.1. Методи прогнозування попиту

У контексті прогнозування споживчого попиту виділяють кілька типів моделей, які використовуються для аналізу попиту на предмети споживання. Ось кілька основних типів прогнозування:

1) Прогнозування активного попиту
2) Прогнозування пасивного попиту
3) Довгострокові прогнози
4) Короткострокові прогнози
5) Зовнішнє макро прогнозування
6) Внутрішнє мікро прогнозування

Правильне прогнозування споживчого попиту є важливим для ефективного управління запасами, планування ресурсів і розробки стратегій розвитку. Однак, якщо цей процес не здійснюється коректно, він може призвести до значних фінансових втрат та втрати попиту на продукт або послугу [23 ,с. 67].

Нижче наведено список основних помилок, які можуть виникнути під час розробки моделей прогнозування споживчого попиту:

1. Перебільшення попиту. Перебільшення попиту може виникнути через занадто оптимістичні прогнози, що спричиняють надмірні витрати та невиправдані інвестиції в виробництво чи закупівлю товарів. Хоча це може здаватися безпечним варіантом, оскільки дозволяє уникнути дефіциту продукції, насправді це може призвести до збитків та негативно вплинути на довгострокову фінансову стабільність організації.

2. Ігнорування історичних даних. Історичні дані є важливою частиною прогнозування попиту, адже вони дозволяють побудувати більш точні моделі та прогнози. Ігнорування минулих тенденцій може призвести до неправильних оцінок попиту, особливо в умовах змінного споживчого середовища. Прогнози повинні базуватися на попередніх даних для того, щоб мати реалістичне уявлення про майбутні тенденції.

3. Надмірна залежність від прогнозів. Прогнозування попиту має бути лише одним із елементів прийняття рішень. Орієнтація лише на прогнози без врахування поточної ситуації, реальних даних та оперативних змін може призвести до помилкових висновків і зрештою до фінансових втрат. Потрібно враховувати, що прогнози є лише оцінками і їх не можна сприймати як остаточні істини.

4. Недостатня гнучкість моделі. Прогнозування попиту повинно бути гнучким і здатним адаптуватися до змін. Споживчий попит може коливатися в залежності від сезонних коливань, економічних умов або непередбачених факторів. Якщо модель прогнозування надто жорстка і не враховує ці зміни, організація може зазнати значних збитків. Важливо мати можливість коригувати прогнози в реальному часі на основі нових даних [39 ,с. 57].

5. Використання застарілих методів обробки даних. Традиційні методи обробки даних, такі як ручне введення або використання електронних таблиць, не є ефективними для прогнозування попиту в сучасних умовах. Вони можуть призвести до помилок і значних витрат часу. Замість цього доцільно застосовувати автоматизовані рішення з використанням штучного інтелекту та машинного навчання, що дозволяють значно скоротити час обробки та підвищити точність прогнозів.

6. Ігнорування регулярного оновлення моделей прогнозування. Прогнози, які базуються на статичних даних, швидко стають застарілими. Регулярне оновлення моделей прогнозування є необхідним для отримання найактуальніших і найбільш точних результатів. Без постійного коригування і вдосконалення прогнозних моделей може виникнути ризик прийняття неправильних рішень через зміни у споживчому попиті або на ринку.

Існує велика кількість програмних рішень для прогнозування попиту, які використовують сучасні технології, зокрема штучний інтелект. Однак при виборі

такого програмного забезпечення слід звернути увагу на кілька важливих аспектів:

1. Програмне забезпечення повинно бути здатним обробляти як внутрішні, так і зовнішні джерела даних. Це дозволяє створювати більш точні прогнози, які враховують всі фактори, що впливають на попит.

2. Для отримання надійних прогнозів потрібно забезпечити наявність достатньої кількості відповідних і точних даних. Без цього навіть найкраще програмне забезпечення не зможе дати правильних результатів [42, с. 67].

3. Програмне забезпечення має бути зручним і мати здатність швидко аналізувати складні дані. Оскільки попит на предмети споживання може залежати від багатьох змінних, необхідно використовувати інтелектуальні системи для обробки великих обсягів даних.

4. Важливо, щоб програмне забезпечення чітко пояснювало процес розрахунку прогнозів, що дозволяє використовувати отримані дані для подальших стратегічних і оперативних рішень.

5. Необхідно забезпечити високу швидкість та надійність бази даних, оскільки інтенсивний потік даних, що стосується попиту на споживчі товари, вимагає ефективної обробки в реальному часі.

Правильне прогнозування споживчого попиту має велике значення для ефективного управління ресурсами та запобігання можливим збиткам через недостатню кількість товарів або надмірні запаси. Розробка надійної моделі прогнозування дозволяє компаніям більш ефективно планувати свої стратегії і забезпечувати відповідний рівень попиту без надмірних витрат.

Таким чином, прогнозування споживчого попиту є критично важливим елементом ефективного управління ресурсами та стратегіями бізнесу. Важливою складовою цього процесу є використання сучасних методів і технологій, таких як штучний інтелект та інтелектуальний аналіз даних, які дозволяють не лише отримувати точніші прогнози, але й адаптувати їх до постійно змінюваного

ринкового середовища. Однак на шляху до ефективного прогнозування можуть виникати різноманітні помилки, такі як перебільшення попиту, ігнорування історичних даних чи надмірна залежність від попередніх прогнозів. Для уникнення цих помилок важливо використовувати не лише надійні дані, а й регулярно оновлювати прогнози, зберігаючи гнучкість у підходах до аналізу.

Загалом, для створення і розробки моделі попиту на предмети споживання необхідно інтегрувати різноманітні методи прогнозування та технології аналізу, враховуючи як внутрішні, так і зовнішні фактори, що можуть впливати на попит. Такий підхід дозволяє досягти високої точності прогнозів, мінімізувати ризики та забезпечити збалансоване управління запасами, що в кінцевому результаті сприяє підвищенню ефективності підприємства та покращенню його фінансових показників.

2.2. Особлиості побудови моделі на основі інструментів інтелектуального аналізу даних

Оптимізація має значну роль у формуванні попиту на споживчі товари, зокрема у розробці ефективних стратегій взаємодії з клієнтами для збільшення їхньої цінності для компанії. Клієнт вважається вигідним, якщо дохід, отриманий від нього, перевищує витрати на його залучення, продаж і обслуговування. Це перевищення часто називається цінністю життя клієнта.

Методи Data Mining можуть бути ефективно застосовані в аналізі попиту на споживчі товари, оскільки цей процес вимагає багатовимірного підходу, що включає такі категорії, як: – ієрархія товарів (бренд, клас, категорія, продукт); – ієрархія періодів (рік, квартал, місяць, дата); – ієрархія клієнтів (регіони, райони, типи клієнтів).

Цей підхід на практиці успішно реалізується завдяки концепції OLAP (On-Line Analytical Processing), яка є основою сучасних систем підтримки прийняття рішень, використовуючи технічний багатовимірний аналіз даних [24 ,с. 67].

E.W.T. Ngai у своїй роботі пропонує графічну систему класифікації методів видобутку даних для прогнозування попиту на споживчі товари, яка зображена на рис. 2.2.

Рис. 2.2. Структура класифікації методів видобутку даних у *попиту на предмети споживання*

Системи OLAP виконують важливу роль у задоволенні потреб користувачів попиту на споживчі товари, оскільки забезпечують підтримку в реальному часі, дозволяючи оптимально поєднувати попередньо обчислені результати з актуальними даними, що надходять за запитом. Вони застосовують специфічні інструменти, адже є найбільш ефективним середовищем для реалізації функціональних інформаційних моделей, що ґрунтуються на динамічних принципах системи.

На сьогодні більшість великих компаній використовують Інтранет-платформи, які, зокрема, забезпечують основні функціональні можливості програм Business Intelligence. Ці платформи організовують інформацію в сховищах даних та обробляють її за допомогою методів інтелектуального аналізу. Інтранет є приватною комп’ютерною мережею, яка дозволяє обмінюватися даними та забезпечує інструменти для співпраці всередині

організації, зазвичай з обмеженим доступом для сторонніх осіб, хоча використовує технології, схожі на Інтернет [32 ,с. 57].

У зв'язку з великими обсягами даних, з якими стикаються керівники, важливо застосовувати передові технології обробки та нові типи систем для підтримки прийняття рішень. Сьогодні для цього активно використовуються технології Business Intelligence, що включають стратегії та інструменти для аналізу бізнес-даних. Ці технології надають інтерпретацію історичних, поточних та прогнозних даних, що дає змогу здійснювати звітність, аналітичну обробку, розробку інформаційних панелей, інтелектуальний аналіз, прогнозування та інші функції. Вони допомагають обробляти великі обсяги структурованих і неструктурованих даних, що сприяє виявленню нових можливостей і формуванню стратегій, які надають підприємству конкурентні переваги та довгострокову стабільність.

Інтелектуальний аналіз даних, що є складовою частиною систем бізнес-аналітики, набув значної популярності в останні роки завдяки успіхам як у наукових дослідженнях, так і в комерціалізації. Отримання даних зосереджене на оцінці прогнозованих можливостей моделей і дозволяє здійснювати аналіз, який був би занадто складним і затратним за допомогою традиційних статистичних методів. Цей підхід дає важливу інформацію, яка сприяє покращенню утримання клієнтів, підвищенню рівня взаємодії з ними, залученню нових і перехресним продажам. Як показує практика, використання програм для прогнозування попиту на споживчі товари дозволяє компаніям підвищити цінність своїх клієнтів і ефективно управляти залученням та утриманням необхідних.

Попри наявність великої кількості літератури з бізнес-аналітики, ця сфера залишає багато перспектив для подальших досліджень. Інтерес до інтелектуального аналізу даних зростає, а потенціал застосування методів оптимізації ще потребує детального вивчення. Окрему увагу слід приділити інтеграції методів оптимізації та інтелектуального аналізу даних, зокрема в

контексті прогнозування попиту на споживчі товари з кількох причин. Здобуття даних і оптимізація можуть бути поєднані для створення клієнтських профілів, що є надзвичайно важливим для багатьох програм, пов'язаних з попитом на споживчі товари [24 ,с. 57].

Процес інтелектуального аналізу даних включає кілька етапів. Спочатку обираються дані для навчального набору, які включають спостереження певних атрибутів, зазвичай це історичні дані. Далі ці дані очищаються і попередньо обробляються. Очищення необхідне для усунення невідповідностей, а попередня обробка допомагає зібрати потрібну інформацію для алгоритму видобутку, що сприяє зменшенню складності задачі.

Рис. 2.3. Процес видобутку даних

Одним із ключових аспектів є вибір атрибутів, який необхідно здійснити до застосування алгоритмів навчання. Це включає в себе процес відбору тих атрибутів, які є суттєвими для пояснення чи прогнозування даних, та відсіювання зайвих або малозначущих. Вибирається підмножина атрибутів з наявних, при цьому кількість вибраних елементів не повинна перевищувати загальну кількість доступних. Цей етап дозволяє скоротити розмірність простору ознак за певними

критеріями та забезпечує якість даних, які будуть використовуватись у подальших етапах обробки. Вибір релевантних атрибутів часто надає структурну інформацію, що важлива для прийняття управлінських рішень.

Інтелектуальний аналіз даних може бути корисним на різних етапах взаємодії з клієнтами, таких як залучення нових клієнтів, збільшення їхньої цінності для компанії та утримання існуючих клієнтів. зміною, що дасть змогу розробити стратегії для утримання таких клієнтів [23 ,с. 57].

Основними складовими моделі попиту на споживчі товари є побудова та управління відносинами з клієнтами через маркетингові стратегії, моніторинг розвитку цих відносин на різних етапах, управління на кожному етапі та визнання нерівномірності вартості цих відносин для компанії. В процесі побудови та управління відносинами через маркетинг компанії можуть використовувати різноманітні інструменти для оптимізації своїх маркетингових кампаній, що включають дизайн організації, схеми мотивації та структури взаємовідносин з клієнтами. Визнання різних етапів попиту на споживчі товари дозволяє бізнесу отримати вигоду, розглядаючи взаємодії клієнтів як взаємопов'язані транзакції. Врахування прибутковості відносин з клієнтами є важливим аспектом систем попиту на споживчі товари. Аналізуючи поведінку споживачів, компанія може більш точно розподіляти ресурси та увагу до різних категорій клієнтів.

Data mining є потужним інструментом, що використовує різноманітні алгоритми для вилучення корисної інформації з великих обсягів даних. Ця технологія дозволяє виявляти приховані закономірності в базах даних, які потім можна застосувати для розробки стратегій. У даній роботі розглядаються алгоритми класифікації та прогнозування, які можуть бути корисними в управлінні взаєминами з клієнтами. Особлива увага приділена використанню наївного класифікатора Байєса для прогнозування поведінки споживачів.

Алгоритми для видобутку асоціативних правил знаходять широке застосування в різних сферах, таких як аналіз покупок, медична діагностика, веб-

навігація та безпека. У цій роботі здійснено огляд методів видобутку правил асоціацій і порівняння традиційних алгоритмів з новими модифікованими підходами, зокрема методом фільтрації записів на основі алгоритму Apriori для частого видобутку шаблонів. Apriori - це алгоритм, який використовується для пошуку часто зустрічаючихся елементів у базах даних і побудови правил асоціацій на їх основі. Він працює шляхом виявлення найбільш часто зустрічаючихся елементів у базі даних і поширення їх на більші набори елементів, які також повинні часто з'являтися в даних. Використовуючи цей метод, можна визначити загальні тенденції, що допомагають у таких сферах, як аналіз покупок [34, с. 67].

Звичайно, традиційний алгоритм видобутку асоціацій потребує кількох етапів для виявлення всіх частих елементів, але за допомогою нового підходу це завдання можна виконати набагато швидше. Це дозволяє значно зменшити час обробки великих баз даних, що є великим досягненням у практичному застосуванні.

Зростання обсягів даних вимагає вдосконалення технологій їх обробки, зокрема через системи бізнес-аналітики та аналізу даних, що сприяють прийняттю управлінських рішень. В даній роботі порівнюються методи видобутку даних і традиційні статистичні підходи, і робиться висновок, що технології видобутку даних є більш ефективними за параметрами складності та часу, необхідного для досягнення результатів. Повне впровадження цих методів у сфері попиту на споживчі товари може значно підвищити ефективність утримання і залучення клієнтів. Більш того, інтеграція методів оптимізації з інтелектуальним аналізом даних може допомогти створювати точні профілі клієнтів, що є важливим елементом для багатьох програм в цій сфері.

2.3. Застосування дейтамайнінгу у моделюванні попиту на редмети споживання

В умовах стрімкого розвитку інформаційних технологій, зокрема в сфері обробки і зберігання даних про бізнес-транзакції, з’являється можливість накопичувати величезні обсяги "сирих" даних, які можуть містити цінну інформацію для прийняття стратегічних рішень. Однак обробити таку кількість даних вручну не в змозі жоден аналітик, тому необхідно використовувати інструменти інтелектуального аналізу даних. Одним із таких інструментів є дейтамайнінг. Попри значний розвиток цих систем у західних країнах, в Україні спостерігається брак інформації про переваги та можливості застосування таких систем у бізнес-середовищі, а також недостатній аналіз існуючих напрямків використання та програмних засобів інтелектуального аналізу. Автори цієї роботи прагнуть дещо заповнити цей пробіл [51 ,с. 67].

Дейтамайнінг (інтелектуальний аналіз даних) - це технологія, яка дозволяє виявляти приховані зв'язки у великих наборах даних. Багато підприємств роками збирають важливу інформацію з бізнес-транзакцій, сподіваючись, що вона стане основою для прийняття ефективних рішень. Процес дейтамайнінгу дозволяє виявити ці корисні знання і допомагає компаніям оцінити ефективність своєї діяльності.

Наприклад, аналіз кошика покупок - це одна з бізнес-проблем, яку можна вирішити за допомогою дейтамайнінгу. Використовуючи ці методи, компанії можуть з’ясувати, які товари часто купуються разом, і таким чином покращити стратегії продажу. Наприклад, якщо покупець купує товар X, він, ймовірно, захоче придбати також товар Y. Це дозволяє компанії прийняти рішення про оптимальне розташування товарів або проведення акцій, що стимулюють попит на обидва продукти.

Більшість інструментів дейтамайнінгу базуються на двох основних технологіях: машинному навчанні та візуалізації. Візуалізація забезпечує наочне подання даних, що дозволяє виявляти закономірності та аномалії. Водночас машинне навчання дозволяє досліджувати взаємозв'язки між даними, виявляючи ті, що людина могла б не помітити через обмеження часу та ресурсу обробки.

Ці технології взаємодоповнюють одна одну у процесі аналізу. Спочатку візуалізація допомагає виявляти основні тенденції та виключення, після чого машинне навчання застосовується для пошуку більш складних залежностей у вдосконаленому наборі даних [23 ,с. 78].

Машинне навчання включає різні методи, зокрема: дерева рішень, асоціативні правила, генетичні алгоритми та нейронні мережі. Технології машинного навчання дозволяють глибше аналізувати і моделювати взаємозв'язки в даних, що сприяє більш точному прогнозуванню та прийняттю рішень.

Таблиця 2.1.

Основні алгоритми дейтамайнінгу

Алгоритм	Опис
Асоціативні правила	Виявляють причинно-наслідкові зв’язки і визначають ймовірності або коефіцієнти достовірності, дозволяючи робити відповідні висновки. Правила представлені в формі «якщо <умова>, то <подія>». Їх можна використати для прогнозування або оцінки невідомих параметрів (значень)
Дерева рішень і алгоритми класифікації	Визначають природні «розбивки» в даних, базовані на використанні цільових змінних. Спочатку здійснюється розбиття по найбільш важливішим змінним. Гілки дерев можна представити як умовну частину правила. Найрозповсюдженішими прикладами є алгоритми класифікаційних та регресійних дерев (Classification and regression trees, CART) або хі-квадрат індукція (Chi-squared Automatic Induction, CHAID)

Штучні нейронні мережі	Тут для передбачення значення цільового показника використовуються набори вхідних змінних, математичних функцій активації і вагових коефіцієнтів вхідних параметрів. Виконується ітераційний навчальний цикл, нейронна мережа модифікує вагові коефіцієнти до тих пір, поки передбачуваний вихідний параметр відповідає дійсному значенню. Після навчання нейронна мережа стає моделлю, яку можна застосувати до нових даних для прогнозування
Генетичні алгоритми	Цей метод використовує ітераційний процес еволюції послідовності поколінь моделей, включаючи операції відбору, мутації та схрещування. Для відбору відповідних особин і відхилення інших використовується «функція пристосування» (fitness function). Генетичні алгоритми, в першу чергу, застосовуються для оптимізації топології нейронних мереж. Проте їх можна застосовувати і самостійно з метою моделювання
Міркування на основі пам'яті або міркування за аналогією	Ці алгоритми базуються на виявленні деяких аномалій у минулому, найбільш близьких до поточної ситуації, для того, щоб оцінити невідоме значення або передбачити можливі результати (наслідки)
Кластерний аналіз	Поділяє гетерогенні дані на гомогенні або напівгомогенні групи. Метод дозволяє класифікувати спостереження по ряду загальних ознак. Кластеризація розширяє можливості прогнозування

Кожен з методів має свої особливості, що впливають на їх переваги та недоліки. Наприклад, перевага дерев рішень та асоціативних правил полягає в тому, що їх можна легко зрозуміти - вони нагадують прості висловлювання на природній мові. Однак, коли даних стає дуже багато, ці методи можуть стати складними для інтерпретації. Крім того, вони не підходять для роботи з великими числовими інтервалами, оскільки кожен вузол дерева або правило відображає лише один зв'язок, і для великого діапазону значень потрібно буде створити безліч правил або вузлів. Нейронні мережі, в свою чергу, мають перевагу в

ефективному представленні числових відносин у широких діапазонах значень, однак їхнє розуміння може бути набагато складнішим.

Існує велика кількість інструментів для реалізації проектів дейтамайнінгу. Це можуть бути як загальнодоступні алгоритми для візуалізації та машинного навчання, так і складні програмні комплекси, які використовують обидві технології та працюють на паралельних процесорах. Вартість останніх може сягати сотень тисяч доларів. Вибір оптимального інструменту для проекту з дейтамайнінгу залежить від кількох факторів, таких як мета проекту (наприклад, аналіз споживчих кошиків) та розмір оброблюваної бази даних. Важливою є також гнучкість обраних інструментів, адже різні стратегії можуть дати різні результати [51 ,с. 57].

Щоб застосувати програмне забезпечення для дейтамайнінгу, необхідно пройти кілька етапів:

1. Оцінити масштаби проекту та визначити, які дані потрібно зібрати. Важливо, щоб проект був орієнтований на досягнення конкретних бізнес-цілей.

2. Створити базу даних. Необхідна інформація часто розподілена між кількома базами, а інколи навіть зберігається в неелектронному вигляді. Дані з різних джерел слід консолідувати і перевірити на наявність суперечностей. Сучасні технології баз даних дозволяють уникнути потреби в окремих вітринах даних для дейтамайнінгу, оскільки корпоративні сховища даних є більш економічно вигідними. Впровадження дейтамайнінгу в рамках корпоративного сховища забезпечує доступ до узгоджених і актуальних даних про клієнтів, що допомагає знижувати витрати на додаткове обладнання та зберігання даних.

3. Очистити дані, перевіривши їх на повноту і консистентність. Якість інформації, яка використовується в процесі дейтамайнінгу, безпосередньо впливає на точність результатів. Цей етап може зайняти значну частину часу, що відводиться на проект.

4. Застосувати алгоритми дейтамайнінгу для виявлення залежностей між даними. У деяких випадках може знадобитися застосування кількох різних алгоритмів на різних етапах проекту, а також паралельний запуск алгоритмів для отримання більш різнобічного аналізу.

5. Оцінити виявлені залежності для їх подальшого використання в рамках проекту. Можливо, буде необхідно залучити експерта, який визначить релевантність і специфічність виявлених взаємозв'язків і порадить, де варто продовжити дослідження.

6. Представити результати у вигляді звіту, що включає всі інтерпретовані залежності. Однак таке представлення дасть лише тимчасову вигоду. Для більшої користі слід зосередитися на навчанні користувачів методам пошуку залежностей та застосуванню результатів на практиці [33, с. 87].

Розподіл часу для проекту дейтамайнінгу залежить від ряду факторів, зокрема від типу кінцевого застосування та наявності й стану сховища даних. Наприклад, для прогнозування продажів виявлені зв'язки можуть залишатися актуальними, поки не зміниться діяльність компанії. Однак для аналізу споживчої корзини підприємства часто шукають нові залежності. Таким чином, для проектів прогнозування витрачається більше часу на початкових етапах, а для аналізу споживчих кошиків - на завершальних етапах.

Завдяки методам машинного навчання та візуалізації можна виявити корисні взаємозв'язки навіть у великих і складних наборах даних.

Дейтамайнінг не має обмежень у своїх сферах застосування і може бути використаний скрізь, де є дані. Проте найбільший інтерес до цих методів виявляють комерційні компанії, які запускають проекти на основі сховищ даних. Багато підприємств повідомляють про великий економічний ефект від використання дейтамайнінгу, що може в десятки разів перевищувати початкові витрати. Наприклад, одне з таких підприємств повернуло інвестиції в 20

мільйонів доларів всього за 4 місяці, а інше за рахунок використання дейтамайнінгу заощадило 700 тисяч доларів на рік.

Існує кілька підходів до вирішення завдань за допомогою дейтамайнінгу, зокрема методи, засновані на аналізі схожих випадків (case-based reasoning, CBR). Ідея цього підходу полягає в тому, щоб знаходити схожі ситуації в минулому і використовувати рішення, яке було успішним для них. Цей метод також відомий як "метод найближчого сусіда" (nearest neighbour) [24 ,с. 65].

Дерева рішень (decision trees) є одними з найпоширеніших інструментів у сфері дейтамайнінгу. Вони будують ієрархічну структуру правил типу «ЯКЩО... ТО», що виглядають як дерево. Для того, щоб прийняти рішення щодо того, до якого класу належить певний об’єкт або ситуація, необхідно почати з кореня дерева і відповідати на питання в його вузлах. Питання, зазвичай, мають форму «Чи більше значення параметра А ніж х?». Якщо відповідь позитивна, перехід здійснюється до правого вузла наступного рівня, якщо ж негативна - до лівого. Багато систем дейтамайнінгу базуються на цьому методі. До найбільш відомих відносяться: See5/C5.0 (RuleQuest, Австралія), Clementine (Integral Solutions, Великобританія), SIPINA (Університет Ліона, Франція), IDIS (Information Discovery, США), KnowledgeSeeker (ANGOSS, Канада). Вартість таких систем коливається від 1 до 10 тисяч доларів.

Генетичні алгоритми не є основним інструментом у дейтамайнінгу, але вони все більше застосовуються для розв’язання комбінаторних задач та оптимізації. Їх основна перевага полягає в можливості паралельного виконання, що ефективно впливає на розв’язання складних задач з багатьма змінними та внутрішніми зв’язками. Одним із прикладів такого програмного забезпечення є GeneHunter від компанії Ward Systems Group, вартість якого становить близько 1000 доларів.

Візуалізація багатовимірних даних є важливою частиною більшості систем дейтамайнінгу. Наприклад, програма DataMiner 3D від словацької компанії

Dimension5 дозволяє ефективно відображати дані у вигляді тривимірних графіків.

Ринок систем дейтамайнінгу продовжує активно розвиватися, і в цьому процесі беруть участь більшість провідних світових комп'ютерних компаній, як-от Microsoft, яка активно займається розвитком цього напрямку, публікує спеціалізовані журнали, організовує конференції та розробляє власні продукти.

Однак однією з основних проблем логічних методів у виявленні закономірностей є проблема пошуку можливих варіантів за обмежений час. Деякі методи, як алгоритми KOPA або WizWhy, намагаються обмежити перебір варіантів, а інші будують дерева рішень (наприклад, алгоритми CART, CHAID, ID3, See5, Sipina), які мають обмежену ефективність при пошуку правил типу «Якщо... то». Інші проблеми полягають у тому, що відомі методи не підтримують функцію узагальнення знайдених правил або пошук оптимальних композицій таких правил.

Вирішення цих проблем може стати важливою темою для подальших досліджень у галузі інтелектуального аналізу даних.

Висновки до розділу 2.

Таким чином, дослідження особливостей моделювання попиту на предмети споживання засобами інтелектуального аналізу даних показало, що вибір методології моделювання є ключовим етапом для забезпечення точності та надійності прогнозів. Найефективнішою методологією виявився багатофакторний підхід, який враховує взаємозв'язки між різними змінними, такими як сезонність, поведінка споживачів, економічні умови та інші зовнішні фактори.

Особливості побудови моделі на основі інструментів інтелектуального аналізу даних включають необхідність ретельного збору, очищення та підготовки даних. Ці етапи критично впливають на якість моделі, оскільки навіть незначні

похибки у вихідних даних можуть суттєво знизити точність прогнозування. Ефективна модель повинна бути адаптивною до змін у даних, що забезпечується за рахунок використання гнучких алгоритмів машинного навчання.

Застосування методів дейтамайнінгу в моделюванні попиту на предмети споживання дозволило підвищити точність прогнозів і виявити приховані закономірності у великих масивах даних. Особливо ефективними виявилися методи кластеризації для сегментації споживачів і регресійного аналізу для оцінки взаємозв'язків між змінними. Застосування цих підходів дозволяє підприємствам краще розуміти ринок, прогнозувати попит та ефективно планувати діяльність.

У підсумку зазначено, що використання інструментів інтелектуального аналізу даних для моделювання попиту забезпечує високий рівень адаптації моделей до реальних умов ринку, що робить їх необхідним елементом сучасного управління споживчим попитом.

РОЗДІЛ 3. РОЗРОБКА МОДЕЛІ ПОПИТУ НА ПРЕДМЕТИ СПОЖИВАННЯ ЗАСОБАМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

3.1. Постановка задачі моделювання

Постановка задачі моделювання є першим етапом у розробці моделі попиту на предмети споживання. Основна мета цього етапу полягає у визначенні чіткої мети моделювання, виявленні ключових факторів впливу на попит та виборі методів, які дозволять врахувати ці фактори. Моделювання попиту на предмети споживання є складним завданням, яке передбачає аналіз різних типів даних: історичних продажів, сезонних трендів, поведінки споживачів, економічних показників тощо. Виходячи з цього, метою моделі є забезпечення точного прогнозування попиту для прийняття ефективних управлінських рішень. Основними завданнями є вибір релевантних змінних, розробка алгоритму обробки даних, створення моделі, що адаптується до змін ринкових умов, та перевірка її точності на реальних даних [38 ,с. 46].

Інтелектуальний аналіз даних - це процес пошуку корисних патернів і зв’язків у великих обсягах даних за допомогою спеціальних алгоритмів, що дозволяють витягувати знання з баз даних. У контексті попиту на споживчі товари, його роль полягає в наступному:

- створення цілісного розуміння життєвого циклу клієнта
- велика кількість даних дозволяє отримати більш точні моделі;
- застосування прогнозних методів і описових моделей;
- активне використання прогностичної аналітики.

У сучасних умовах жорсткої конкуренції клієнт є важливим ресурсом для підприємства. Бізнес може отримати конкурентну перевагу, якщо ефективно управляє своїми відносинами з клієнтами. Дослідження попиту на споживчі товари, оцінка вартості та застосування аналізу даних і статистики є важливими аспектами цього процесу. Використання інтелектуального аналізу даних

дозволяє знаходити закономірності в масиві інформації, що може значно вплинути на розвиток підприємства.

ABC-класифікація - це метод, який дозволяє розподілити клієнтів на групи, враховуючи їхню значимість для компанії згідно з обраним критерієм. Основні критерії для такої класифікації можуть бути:

- обсяг продажів в натуральних одиницях чи в грошовому еквіваленті,
- оборот,
- прибуток.

Цей підхід веде до проведення ABC-аналізу, що дає змогу класифікувати контрагентів за їхньою важливістю для бізнесу. Метод базується на принципі Парето, або «правилу 20/80», що означає: 20% клієнтів приносять 80% доходу, а решта 80% забезпечують лише 20% прибутку.

У процесі ABC-аналізу клієнти сортуються за певними критеріями і потім розподіляються на три категорії (групи). Це дає змогу здійснити сегментацію за ступенем їхнього внеску в загальний дохід підприємства, включаючи продажі та обслуговування [43 ,с. 67].

Перевагами методу дотичних є його масштабованість та можливість автоматизації, а також менша помилка в межах середніх значень точки Парето. Однак метод має недолік у вигляді несиметричності стосовно точки Парето, що може збільшити похибку, а також він не дуже ефективний для розподілів, що наближаються до крайніх точок.

Для проведення ABC-аналізу важливо вибрати критерій оцінки, який може бути як простим (наприклад, прибуток), так і складним (комплексний показник, що включає, наприклад, прибуток, оборот та швидкість повернення дебіторської заборгованості). Кожному з критеріїв надається певний ваговий коефіцієнт для подальших розрахунків. Вибір критерію залежить від цілей аналізу. Для точності аналізу рекомендується проводити його не лише по клієнтах, а й по галузі та

менеджерах. Далі обирається метод для визначення груп A, B, C, після чого аналізуються дані, робляться висновки та застосовуються на практиці.

Кластеризація - це метод, який дозволяє поділити дані на групи відповідно до встановлених вимог. Кластерний аналіз клієнтської бази має на меті групування клієнтів за схожими ознаками, зокрема за поведінковими характеристиками. На відміну від сегментації, кількість груп (кластерів) у кластеризації не визначена заздалегідь, так само як і правила віднесення клієнта до певної групи. Тому кластеризація відноситься до задач машинного навчання (machine learning). Окрім цього, вона є частиною Data Mining і допомагає знаходити приховані закономірності в «сирих» даних про поведінку клієнтів, що дозволяє реалізувати персоніфіковану маркетингову стратегію [31, с. 56].

Для роботи з базами даних використовуються спеціальні алгоритми кластерного аналізу. З поділяють на ієрархічні і ітеративні.

Ієрархічні алгоритми застосовуються коли істинну кількість груп не відомо. Суть підходу полягає в вимірі попарних відстаней між об'єктами і послідовному об'єднанні (або навпаки, дробленні) всіх об'єктів на групи. В результаті виходить деревоподібна візуалізація, де сусідні об'єкти схожі один на одного.

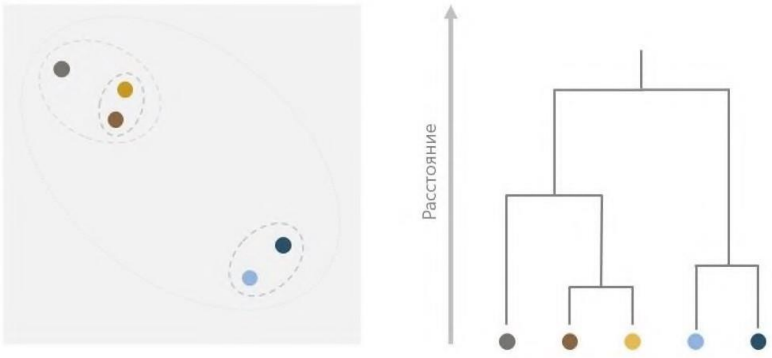

Рис. 3.1. Ієрархічні алгоритми

Вибір метрики відстані залежить від типу використовуваних даних. Для числових значень зазвичай застосовують евклідову відстань, Манхеттенську

відстань або відстань Чебишева. Для категоріальних даних (які не можна виміряти чисельно, як, наприклад, стать: чоловічий чи жіночий) використовуються відстані Хеммінга, Джакарди або міра Серенсена.

Методи об'єднання кластерів залежать від розподілу даних. Серед найбільш популярних - метод Варда, метод найближчого сусіда, метод невиваженого попарного середнього та зважений метод центроїда. приведення їх до єдиної шкали вимірювань.

Ітеративні методи кластеризації передбачають, що кількість груп задається на початку. Потім, на кожному кроці, матриця відстаней перераховується, поки алгоритм не знайде оптимальне розбиття. До таких методів належать k-means, EM-алгоритм і CLOPE для номінативних даних.

Аффінітівний аналіз (також відомий як affinity analysis) є одним із популярних підходів у сфері Data Mining. Його основне завдання - виявлення взаємозв'язків між подіями, які відбуваються разом. Одним із прикладів є аналіз ринкової кошика (market basket analysis), що фокусується на пошуку зв'язків між різними подіями. Завдання цього аналізу - сформулювати правила, які допомагають кількісно описати ці зв'язки, і ці правила називаються асоціативними [24, с. 65].

Ключовим елементом теорії асоціативних правил є транзакція, яка являє собою набір подій, що сталися одночасно. Наприклад, транзакцією можна вважати покупку товарів клієнтом у магазині. Зазвичай покупці купують не один товар, а кілька, які складають так звану ринкову кошик. Питання, яке виникає: чи є покупка одного товару в кошику причиною чи наслідком покупки іншого товару, чи пов'язані ці події? Це питання і вирішують асоціативні правила.

Іншим важливим поняттям є предметний набір - це множина товарів, які з'являються в одній транзакції. Аналіз ринкової кошика полягає в пошуку таких комбінацій товарів, які взаємопов'язані. Тобто, визначаються такі товари, чиє наявність в одній транзакції підвищує ймовірність придбання інших товарів.

Підтримка асоціативного правила визначається як кількість транзакцій, які містять як умови, так і наслідки.

Наприклад, для асоціації A > B підтримка може виглядати так:

Достовірність асоціативного правила A > B показує точність правила і розраховується як відношення числа транзакцій, що містять і умову, і наслідок, до кількості транзакцій, що містять тільки умову. Якщо підтримка та достовірність високі, можна з великою впевненістю стверджувати, що в майбутніх транзакціях, що включають умову, буде також і наслідок.

Аналітики можуть орієнтуватися на правила, які мають високу підтримку або достовірність, або ж на комбінацію обох показників. Правила, у яких значення підтримки чи достовірності перевищують встановлений поріг, називаються сильними. Наприклад, можна задати мінімальну підтримку 20% і мінімальну достовірність 70% для товарів, які часто купують разом в супермаркеті, або знизити підтримку до 1% при виявленні шахрайств, оскільки кількість таких транзакцій є обмеженою [41 ,с. 67].

Що стосується значимості асоціативних правил, то використання стандартних методів пошуку асоціативних правил призводить до виявлення величезної кількості зв'язків, що ускладнює їх ручну обробку. Тому важливо зменшити кількість правил до найзначніших для конкретної задачі. Для цього застосовують додаткові критерії значимості, які дозволяють оцінити важливість кожного правила. Серед таких критеріїв є об'єктивні заходи, як підтримка та достовірність, а також суб'єктивні, зокрема, ліфт (lift) і левередж (leverage), що визначаються залежно від конкретних умов аналізу.

Ліфт обчислюється таким чином: $L(A \rightarrow B) = C(A \rightarrow B) / S(B)$.

Ліфт визначається як відношення частоти наявності умови в транзакціях, що також містять наслідок, до загальної частоти появи наслідку. Якщо значення ліфта перевищує одиницю, це свідчить про те, що умова частіше зустрічається у

транзакціях з наслідком, ніж в інших. Таким чином, ліфт є індикатором сили зв'язку між двома наборами предметів: коли ліфт > 1, зв'язок є позитивним; коли він дорівнює 1 - зв'язок відсутній; а при значеннях ліфта < 1 зв'язок є негативним.

Іншою мірою значущості правила, запропонованої Г.Пятецкім-Шапіро, є левередж:

$$T(A \rightarrow B) = S(A \rightarrow B) - S(A)S(B).$$

Левередж - це різніця между частотою, яка спостерігається, З якою Умова и наслідок з'являються спільно, и відтворенням частот з'явиться (підтрімок) умови и наслідку окремо.

Дані методи прекрасно взаємодоповнюють один одного (рис. 3.2):

Рис. 3.2 Методи інтелектуального аналізу

Колаборативна фільтрація здійснює аналіз покупок клієнтів та виявляє найбільш подібних користувачів. Після цього, в процесі агрегування, вносяться всі товари, які придбали схожі клієнти, але які відсутні у вибраному. Асоціативний аналіз визначає взаємозв'язки між товарами, що дозволяє додавати до рекомендованих товарів ті, що є наслідками асоціацій, якщо відповідні товари є в покупках клієнта. Кластерний аналіз допомагає вирішити проблему

«холодного старту»: коли у клієнта немає історії покупок, перші два методи не можуть дати рекомендації. Проте, аналізуючи метадані клієнтів, їх можна поділити на групи, і, знаючи їх належність до певного кластера, можна запропонувати популярні товари серед клієнтів цього класу.

3.2. Збір, очищення та підготовка даних

На цьому етапі здійснюється формування вхідного набору даних, який слугуватиме базою для побудови моделі. Джерелами даних можуть бути інформаційні системи компаній, маркетингові дослідження, дані від третіх сторін, а також відкриті бази, що містять інформацію про макроекономічні показники. Зібрані дані часто мають нерівномірний характер, містять пропуски або шум. Тому на етапі очищення здійснюється видалення некоректних значень, заповнення пропусків, стандартизація змінних та усунення дублюючих записів. Підготовка даних також передбачає нормалізацію значень для забезпечення їх коректного введення у модель, трансформацію категорійних змінних у числові та формування фінального набору змінних, які будуть використовуватися у процесі аналізу. На цьому етапі важливо забезпечити високу якість даних, оскільки навіть незначні похибки можуть суттєво вплинути на точність моделі [38, с. 55].

Використання інформаційних технологій для прогнозування попиту досягло значного прогресу, забезпечивши високу точність, прозорість і можливість відтворення результатів. Це має величезний вплив на розвиток продуктового ритейлу. Основними перевагами застосування IT-технологій у прогнозуванні попиту є:

1) Точність та прозорість. Власники бізнесу та аналітики приділяють особливу увагу точності та прозорості прогнозів, оскільки це дозволяє досягти об’єктивних і перевірених результатів. Нові методи прогнозування впроваджуються лише після ретельного тестування порівняно з надійними контрольними показниками. Точність прогнозу є важливим орієнтиром, адже її

помилки мають економічні наслідки. Машинне навчання для прогнозування попиту повинно бути максимально точним, і воно дозволяє налаштовувати точність і зміщення, щоб оперативно оптимізувати прогнози. Крім того, результати прогнозування на основі машинного навчання є прозорими, що дозволяє зрозуміти вплив вхідних даних.

2. Можливість працювати з великими обсягами даних. Точність прогнозів машинного навчання значною мірою залежить від здатності обробляти великі й різноманітні дані, що зібрані на детальному рівні, наприклад, по кожному товару або артикулу. Важливо розрізняти самі алгоритми прогнозування та підготовку даних до їх використання. Для цього використовуються дані, такі як ціни, знижки, мережі розповсюдження, погодні умови та навіть інформація з соціальних мереж, що дозволяє покращити точність прогнозу без втручання людини.

3. Швидка адаптація до змін і порушень ланцюга поставок. Прогнозування з використанням машинного навчання дозволяє автоматично оновлювати прогнози, що сприяє швидкій адаптації до змін у ринкових умовах або порушеннях у ланцюгу поставок. Такі системи можуть працювати в реальному часі, оновлюючи дані і прогнози щоденно або щотижнево. Це дозволяє коригувати планування попиту на основі актуальних фактичних даних і забезпечує точність прогнозів для коригування цінової політики і акцій. Крім того, таке прогнозування можна поєднати з онлайн-системами рекомендацій, що підвищує ефективність маркетингових кампаній і стимулює додаткові покупки.

4. Швидкість обробки даних та пришвидшене корпоративне навчання. Однією з важливих переваг машинного навчання є його здатність швидко обробляти великі обсяги даних. Сучасні інструменти для машинного навчання, розроблені для платформи R, оптимізовані для використання архітектури чіпів Intel та графічних процесорів (GPU), що дозволяє виконувати велику кількість обчислень за секунду, максимально ефективно використовуючи пам'ять. Це

сприяє значному прискоренню процесу прогнозування, наближаючи його до миттєвих результатів. Коли на першому плані стоїть досягнення високої точності, машинне навчання може інтегрувати додаткові предиктори та методи глибокого навчання, знаходячи баланс між швидкістю і точністю. Це дозволяє менеджерам приймати обґрунтовані рішення щодо інвестицій у такі напрямки, як збирання даних, прискорення обробки або розширення обчислювальних потужностей, при цьому ці питання можна ефективно вирішувати за допомогою хмарних рішень, таких як Microsoft Azure [28, с. 65].

5. Вплив на бізнес. Завдяки застосуванню технологій машинного навчання, що працюють із великими обсягами різноманітних даних, компанії отримують можливість здійснювати точніші прогнози, оскільки зростає кількість і різноманітність джерел даних, що надходять швидше. У добре організованому сховищі даних це дозволяє отримувати більш достовірні прогнози. Такий підхід відкриває нові можливості для прискорення аналітики, дає змогу глибше розуміти, як можна використовувати дані для підвищення ефективності бізнес-процесів, і приводить до реальних фінансових покращень.

Метод ABC для контролю запасів включає систему, яка допомагає управлінню запасами і використовується в матеріальному постачанні та розподілі. Він також відомий як вибірковий контроль запасів (SIC). Аналіз ABC передбачає розподіл товарів на три категорії за спаданням вартості: A, B і C. Товари категорії A мають найбільшу цінність, B – середню, а C – найменшу.

Контроль запасів є важливим для ефективного управління витратами. Завдяки аналізу ABC компанія може фокусуватися на товарах з високою вартістю (категорія A), а не на низькоцінних (категорія C).

Принцип Парето лежить в основі цього методу: 80% доходу забезпечують лише 20% найбільш популярних товарів, тобто правило 80/20. Це дозволяє оптимізувати рівень запасів і зосередити ресурси на найцінніших товарах.

Категорія А: Товари цієї категорії мають найвищу споживчу вартість, при цьому лише 10-20% товарів становлять 70-80% річної вартості споживання. Це вимагає особливої уваги до їх запасів.

Категорія В: Товари середнього рівня споживчої вартості займають 30% запасів і забезпечують 15-20% річної споживчої вартості.

Категорія С: Товари цієї категорії мають найнижчий попит і складають близько 50% запасів, але лише 5% річної вартості споживання.

Застосування аналізу АВС дозволяє точно розподілити ресурси для оптимізації продажів.

Категорія А: а) Потребують суворого контролю запасів і спеціальних зон для зберігання.

b) Високий попит, тому ці товари мають кращі прогнози збуту. c) Потрібно часто замовляти для підтримки запасів.

Категорія В: а) Менш важливі, ніж товари категорії А, але важливіші за категорію С.

b) Підлягають перегляду і можуть бути переміщені до категорії А або С залежно від попиту.

Категорія С: а) Рідше замовляються, і запас зберігається мінімально.

b) Можуть зберігатися лише по одному товару або замовлятися після здійснення покупки [40 ,с .65].

c) Потрібно ретельно оцінити, чи варто ці товари тримати в асортименті.

Після проведення аналізу АВС були визначені 5 найкращих і 5 найгірших товарів у категорії "морозиво порційне.

Таблиця 3.1

АВС-аналіз

Код товару	Назва товару	Вид товару	Продажі за 5	%	Накопичений підсумок	ТОР	Рішення

			місяців, шт				
290449	Ноутбук Dell XPS 13	Ноутбук	49699	0,83%	0,83%	A	розширити
322755	Ноутбук HP Pavilion 15	Ноутбук	49603	0,83%	1,66%	A	розширити
267742	Ноутбук Lenovo ThinkPad X1 Carbon	Ноутбук	49008	0,82%	2,49%	A	розширити
327806	Ноутбук Asus ZenBook 14	Ноутбук	48305	0,81%	3,29%	A	розширити
322257	Ноутбук Acer Swift 3	Ноутбук	48182	0,81%	4,10%	A	розширити
322713	Ноутбук MSI Modern 14	Ноутбук	47292	0,79%	7,30%	A	
255541	Ноутбук HP Envy 13	Ноутбук	47116	0,79%	8,09%	A	
327914	Ноутбук Apple MacBook Air	Ноутбук	38143	0,64%	41,02%	B	
290448	Ноутбук Lenovo Ideapad 5	Ноутбук	37360	0,63%	42,29%	B	
290475	Ноутбук Dell Inspiron 14	Ноутбук	37326	0,63%	42,91%	B	
315462	Ноутбук Samsung Galaxy Book	Ноутбук	21253	0,36%	83,53%	B	
327911	Ноутбук Microsoft	Ноутбук	20532	0,34%	84,93%	B	

	Surface Laptop						
238610	Ноутбук Acer Aspire 5	Ноутбук	20507	0,34%	85,27%	C	закрито
250631	Ноутбук Toshiba Satellite L50	Ноутбук	20338	0,34%	85,61%	C	закрито
315661	Ноутбук Huawei MateBook D 14	Ноутбук	20112	0,34%	85,95%	C	закрито
229838	Ноутбук Dell Inspiron 15	Ноутбук	2737	0,05%	99,86%	C	закрито
270106	Ноутбук Asus VivoBook 15	Ноутбук	2732	0,05%	99,90%	C	закрито
268522	Ноутбук HP 14s	Ноутбук	2475	0,04%	99,95%	C	закрито
290477	Ноутбук Lenovo Legion 5	Ноутбук	2232	0,04%	99,98%	C	закрито
291638	Ноутбук Xiaomi Mi Notebook	Ноутбук	1000	0,02%	100,00%	C	закрито

Переваги впровадження методу ABC-аналізу:

- Допомагає компаніям зберігати контроль над товарами, які потребують значних фінансових вкладень.
- Дозволяє відстежувати всі позиції товару, що не тільки знижує витрати на персонал, але й забезпечує підтримання оптимальних рівнів запасів.

- Забезпечує підтримку високого коефіцієнта оборотності запасів завдяки регулярному моніторингу [39 ,с. 65].
- Значно знижує витрати на зберігання товарів.
- Надає точну інформацію про наявність товарів категорії С, що дозволяє підтримувати їх без шкоди для більш важливих товарів.

Недоліки застосування ABC-аналізу:

- Для ефективного використання методу необхідна належна стандартизація товарів у магазині.
- Потрібна наявність ефективної системи кодування для матеріалів.
- Оскільки цей аналіз враховує тільки фінансову вартість товарів, він не бере до уваги інші можливі важливі аспекти для бізнесу, що є суттєвим недоліком.

Аналіз поведінкових особливостей клієнтів бренду «Ноутбукер» за допомогою технологій Big Data

Для виконання аналізу необхідно зосередитись на чотирьох основних бізнес-задачах при роботі з Big Data:

1. Сегментація клієнтів за рівнем лояльності. .
   - Створення клієнтських сегментів за допомогою RFM-матриці.
   - Аналіз кількості клієнтів по окремих філіях та поведінки споживачів.
2. Аналіз зміни поведінки клієнтів під час карантину.
   - Порівняння трафіку у період першого карантину (2020 рік) та після його послаблення (2023 рік) в Україні та ОАЕ.
   - Вивчення змін середнього чека в ці періоди.
   - Оцінка обсягів продажів в періоди карантину та після послаблення обмежень.
3. Дослідження поведінки клієнтів залежно від країни.
   - Аналіз активності клієнтів у різні години та дні.
   - Оцінка середнього чека та платоспроможності в різних містах.

- Вивчення сезонних коливань у відвідуваності та транзакціях.
- Порівняння способів оплати по регіонах.
- Проведення ABC-аналізу за популярністю страв у різних містах і формулювання висновків щодо вподобань клієнтів.

Тепер перейдемо до першого завдання. Для сегментації клієнтів використовуємо метод кластеризації за рівнем лояльності, створюючи RFM-матрицю, де вісь X відображає кількість днів з останнього візиту, а вісь Y - частоту відвідувань. У матрицю потрапляють клієнти, які мають карту лояльності і зробили хоча б одну транзакцію. Клієнти, які отримали карту на першому візиті, але не зробили покупку, не беруться до уваги, оскільки їх не можна ідентифікувати [17 ,с. 88].

Нові клієнти виділені червоною рамкою (див. рис. 3.3.).

Сегменти:

1. «Перша покупка»: останній візит - до 30 днів тому, частота - 1-2 візити на рік.
2. «Новачок»: останній візит - 31-90 днів тому, частота - 1-2 візити на рік.
3. «Разові в зоні ризику»: останній візит - 91-180 днів тому, частота - 1-2 візити на рік.
4. «Разові втрачені»: останній візит - більше ніж 180 днів тому, частота - 1-2 візити на рік.

Перспективні клієнти виділені жовтою рамкою (див. рис. 2.1). Сегменти:

1. «Перспективні»: останній візит - до 90 днів тому, частота - 3-5 візитів на рік.
2. «Перспективні в зоні ризику»: останній візит - 91-180 днів тому, частота - 3-5 візитів на рік.
3. «Відтік з перспективних»: останній візит - більше ніж 180 днів тому, частота - 3-5 візитів на рік.

Лояльні клієнти виділені зеленою рамкою (див. рис. 2.1). Сегменти:

1. «Priority»: останній візит - до 90 днів тому, частота - 10 і більше візитів на рік.

2. «Лояльні»: останній візит - до 90 днів тому, частота - 6-9 візитів на рік.

3. «Лояльні в зоні ризику»: останній візит - 91-180 днів тому, частота - 6 і більше візитів на рік.

4. «Відтік з лояльних»: останній візит - більше ніж 180 днів тому, частота - 6 і більше візитів на рік.

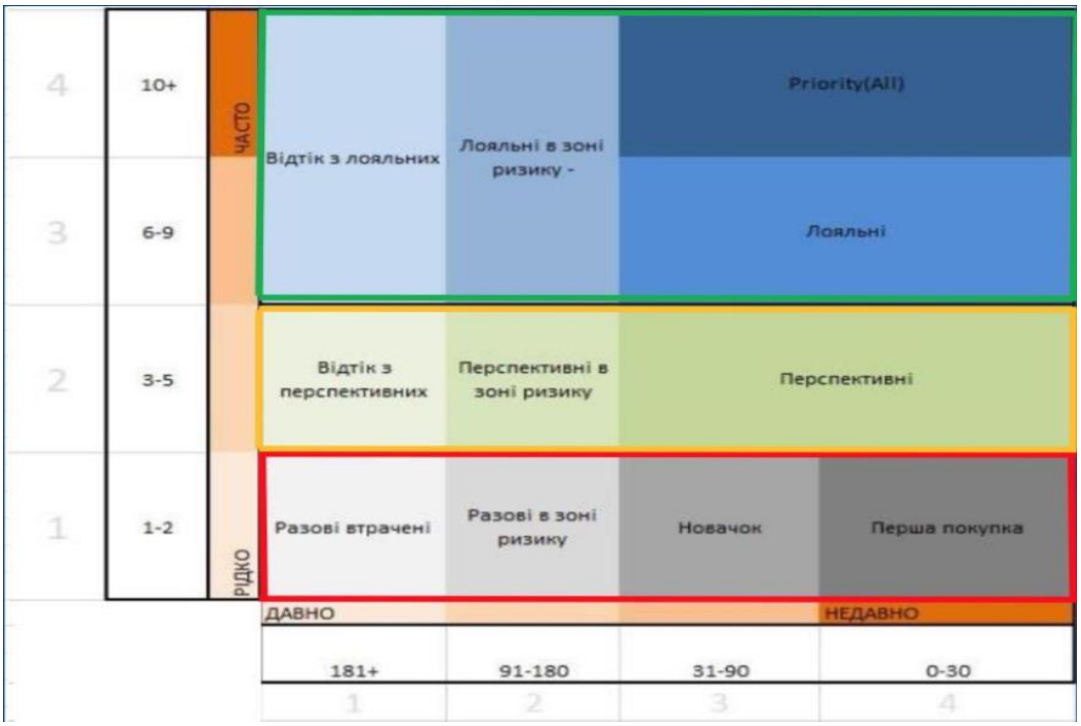

Рис. 3.3. Сегменти цільової аудиторії за її лояльністю

Якщо клієнт з групи «перспективних» підвищує свій статус до категорії «Лояльні», він закріплює цей рівень і не може переміститися до нижчої категорії. Можлива лише міграція між сегментами або підвищення. Наприклад, якщо клієнт із сегменту «Priority» перестає відвідувати бренд більше ніж на рік, він переміщується до сегменту «Відтік з лояльних» і фіксується в цій групі.

Завдяки технологіям Big Data було визначено обсяги та динаміку кожного з сегментів. Період аналізу: з 04.10.21 по 09.02.22, філія бренду «Argentina Grill» у Харкові. Загальний приріст RFM бази становив 4% (від 9559 клієнтів до 9917), що відповідає приблизно 90 новим контактам щомісяця. За цей період було видано 852 карти лояльності. З них 31% потрапили до RFM-матриці, причому кожен другий клієнт, який отримав карту, використав її повторно [24 ,с. 65].

Результати приросту сегментів за цей період:

1. Приріст нових клієнтів (виділених червоною рамкою) склав 5%. 213 клієнти перейшли з цієї групи до більш високих сегментів.

2. Сегмент «Перспективні» (виділений жовтою рамкою) зріс на 4%. З 213 клієнтів, які перемістилися з групи нових, 67 залишилися в сегменті «Перспективні».

3. До сегмента «Лояльні» (зелена рамка) з нижчих груп приєдналися 126 клієнтів, приріст склав 3%.

З цього можна зробити висновок, що в цілому клієнти проявляють зростаючу лояльність, хоча періоди між їхніми повторними відвідуваннями збільшуються, що може свідчити про наближення до зони ризику. Для уникнення втрати клієнтів необхідно посилити маркетингову активність.

Рис. 3.4. Приріст за сегментами лояльності

Загалом, аналіз поведінкових характеристик цільової аудиторії, проведений за допомогою технологій Big Data, дозволив отримати цінні інсайти для компанії "Ноутбукер", яка займається продажем комп'ютерної техніки. Використовуючи різноманітні інструменти збору та аналізу даних, було виявлено основні сегменти клієнтів, їхні переваги та особливості споживчої поведінки залежно від регіону, лояльності до бренду та інших факторів [32 ,с. 57].

Результати показали, що компанія має потенціал для покращення взаємодії з клієнтами, зокрема через розробку ефективних стратегій утримання клієнтів та адаптацію маркетингових кампаній до різних сегментів. Виявлені тенденції дозволяють оптимізувати підхід до продажу комп'ютерної техніки та приймати зважені рішення щодо удосконалення процесів продажу та обслуговування клієнтів.

Отже, застосування Data mining у бізнес-процесах "Ноутбукер" не тільки дозволяє більш точно сегментувати клієнтів, а й допомагає створювати персоналізовані стратегії для залучення і утримання лояльних споживачів. Важливо, щоб компанія продовжувала вдосконалювати методи аналізу даних та інтеграцію їх у свої BI-системи для подальшого розвитку і успішної конкуренції на ринку комп'ютерної техніки.

3.3. Проектування архітектури моделі

Проектування архітектури моделі включає вибір алгоритмів і методів, які найкраще відповідають поставленій задачі. У даному дослідженні було обрано підхід, який поєднує методи регресії, кластеризації та нейронних мереж. Регресійний аналіз використовується для оцінки кількісного впливу змінних, таких як ціна товару, рівень доходів споживачів або сезонні фактори.

Кластеризація дозволяє сегментувати споживачів за подібними характеристиками, що дає змогу створювати більш точні моделі для кожної групи. Нейронні мережі забезпечують високу адаптивність моделі, дозволяючи

враховувати нелінійні залежності між змінними. При проектуванні архітектури моделі використано модульний підхід, який дозволяє легко адаптувати модель до нових умов і додавати нові функціональні можливості [38 ,с. 54].

Довідковою моделлю для дослідження є міжгалузевий стандартний процес здобуття даних (CRISP-DM), рис. 3.5, який добре відомий для розробки проектів із здобування даних.

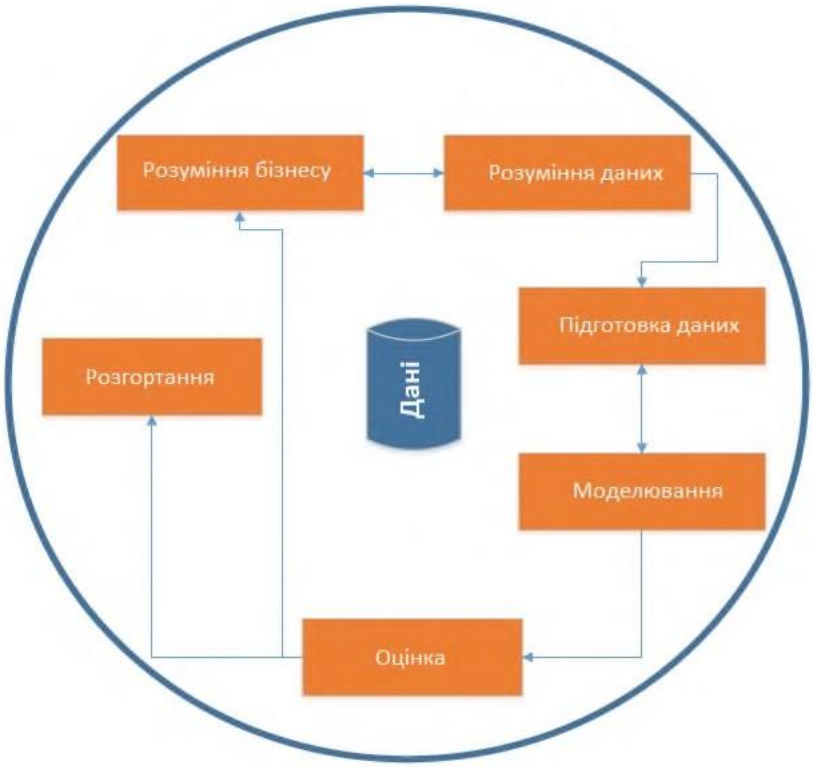

Рис. 3.5 CRISP-DM фреймворк

Процес дослідження, згідно з цією методологією, можна розподілити на кілька етапів:

1. Аналіз бізнесу. Це перший етап, який зосереджується на розумінні основних цілей проекту та вимог з бізнесової точки зору. Це знання трансформується у визначення проблеми для подальшого аналізу даних та розробку початкового плану для досягнення поставлених цілей.

2. Аналіз даних. На цьому етапі здійснюється первинний збір даних, їх огляд, виявлення проблем з якістю даних, а також пошук потенційно корисних підмножин для формування гіпотез і пошуку прихованих закономірностей.

3. Обробка даних. Цей етап включає в себе всі етапи підготовки даних, необхідних для подальшої роботи з моделями. Це може включати очищення, трансформацію, відбір таблиць, записів та атрибутів. Ці дії можуть виконуватися кілька разів для досягнення необхідної якості даних.

4. Моделювання. На цьому етапі вибираються та застосовуються різні методи для вирішення завдання. Зазвичай для одного типу проблеми є кілька підходів, і в деяких випадках можна повернутися до етапу обробки даних, якщо метод вимагає змін у формах даних [23 ,с. 55].

5. Оцінка. Перед остаточним впровадженням моделі необхідно впевнитися, що всі поставлені цілі були досягнуті. Важливо перевірити і підтвердити, що всі важливі бізнесові питання були враховані. Наприкінці цього етапу приймається рішення щодо використання результатів аналізу.

6. Остаточне впорядкування. Створення моделі зазвичай не завершує проект. Отримані результати повинні бути організовані та представлені так, щоб замовник міг ними ефективно скористатися.

Пропонована модель зосереджена на прогнозуванні ймовірності повернення позики клієнтами, ґрунтуючись на їхній поведінці. Вхідними даними є поведінкові дані клієнтів, на основі яких модель приймає рішення щодо схвалення чи відмови в позичці. Для генерації атрибутів та прийняття рішень використовується метод індукції даних на основі дерева рішень. Модель аналізу даних, що пропонується, представлена на відповідній схемі.

Рис. 3.6. Запропонована модель

1. Розуміння проблеми. Моделювання збору даних починається з аналізу банківського сектору та наявних процедур обробки позик. Це дозволяє краще усвідомити виклики і основні ризики, які можуть виникнути при прийнятті рішення щодо надання або відмови у кредиті.

2. Аналіз даних. На цьому етапі банк збирає всю необхідну інформацію про клієнтів, яка потрібна для подальшого аналізу. Вивчаються різні атрибути, що можуть бути важливими для оцінки ризиків і прийняття обґрунтованих рішень.

3. Фільтрація даних. Зібрані дані проходять через процес фільтрації, де вибираються лише ті атрибути, що будуть корисні для прогнозування результатів. Неповні або некоректні записи видаляються, щоб підготувати дані до подальшої обробки.

4. Розробка моделі. Цей етап включає створення зручної та ефективної системи, яка буде доступна навіть для користувачів без спеціальних технічних знань. Система надає релевантні атрибути, які допомагають ухвалити рішення щодо затвердження або відмови у наданні позики, а також сприяє прогнозуванню довіри до потенційних клієнтів [54 ,с. 87].

5. Оцінка ефективності системи. На завершення проекту система тестується за допомогою набору тестів для перевірки її продуктивності та ефективності. Це допомагає переконатися, що всі аспекти роботи системи відповідають встановленим вимогам.

Основні принципи роботи моделі попиту на споживчі товари передбачають:

1) Єдине сховище даних, куди збирається інформація про взаємодію з клієнтами (клієнтська база).
2) Використання різних каналів комунікації, таких як обслуговування на точках продажу, телефонні дзвінки, електронна пошта, соціальні мережі, чати, рекламні посилання та інші платформи.
3) Аналіз зібраних даних для прийняття рішень, таких як сегментація клієнтів залежно від їх значущості для бізнесу, прогнозування реакцій на різні маркетингові кампанії, а також оцінка попиту на певні продукти.

Рис. 3.7. Структура роботи системи

Основною метою впровадження моделі попиту на споживчі товари є підвищення рівня задоволеності клієнтів через аналіз зібраних даних про їх поведінку. Це дозволяє оптимізувати тарифну політику та налаштувати маркетингові інструменти. Окрім того, модель дає можливість ефективно враховувати індивідуальні потреби клієнтів з мінімальним залученням персоналу, а також виявляти потенційні ризики та можливості на ранніх етапах.

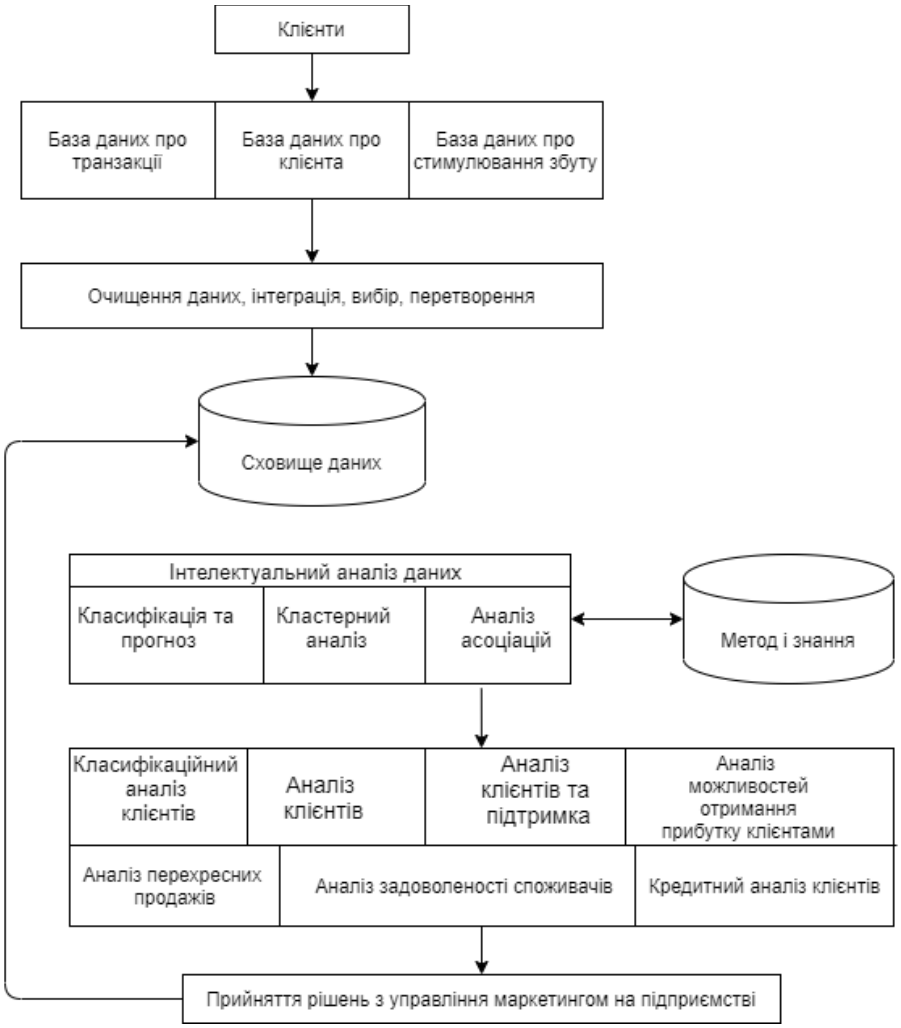

Рис. 3.8.Процес управління відносинами з клієнтами

Система попиту на споживчі товари є потужним джерелом даних для компанії. Ці дані можна використовувати через готові звіти або самостійно їх завантажувати й обробляти, оскільки вони зберігаються в уніфікованій формі.

Ключовим аспектом є постійний та ітеративний збір даних, а також порівняння результатів за різні періоди, щоб спостерігати за динамікою показників.

Основні напрямки роботи з даними:

1) Сегментація клієнтів та персоналізація пропозицій, що позитивно впливає на ставлення клієнтів до компанії.
2) Аналіз рентабельності для визначення найбільш вигідних груп клієнтів або товарів.
3) Відстеження дій клієнтів, наприклад, моніторинг обсягів покупок для підключення до програм лояльності [24 ,с. 78].
4) Проведення план-фактного аналізу для оцінки виконання планів на різних рівнях.
5) Використання ABC-аналізу для виявлення найбільш прибуткових товарів чи послуг.
6) Аналіз стану складів та цінових списків.
7) Оцінка ефективності роботи менеджерів з продажу.

Завдяки цим даним можна створювати глибоку бізнес-аналітику. Проте існує набір мінімальних показників, які завжди повинні бути в роботі:

1) Темп зростання, який дозволяє порівнювати поточні показники з попередніми періодами.
2) Темп приросту, що допомагає оцінити позитивні чи негативні зміни.
3) Частки, щоб оцінити не тільки загальні, але й відносні показники, наприклад, розподіл доходу між окремими замовленнями.
4) Мода, яка визначає найбільш типову кількість взаємодій, що допомагає встановити об'єктивні норми для продажників.
5) ROI (рентабельність інвестицій), що дає змогу оцінити ефективність витрат на рекламу, тренінги тощо.
6) Важливо пам'ятати, що необхідно працювати з різними типами показників - абсолютними, відносними та середніми.

Система попиту на споживчі товари є важливим джерелом даних, і ці дані мають бути чітко структуровані та використовуватись для конкретних цілей. Безсистемне збирання даних або їх використання без реальної потреби є марною тратою часу. Керівники повинні вміти працювати з даними, виділяти ключову інформацію та ефективно використовувати її для прийняття бізнес-рішень.

Отже, система попиту на споживчі товари є важливим інструментом для збору та аналізу даних, що дає можливість компанії ефективно управляти своїм бізнесом. Постійний та ітеративний збір даних дозволяє отримувати актуальну інформацію, що може бути використана для покращення взаємодії з клієнтами, персоналізації пропозицій, оптимізації бізнес-процесів та підвищення рентабельності. Аналіз рентабельності, сегментація клієнтів, оцінка ефективності менеджерів і проведення план-фактного аналізу допомагають ухвалювати обґрунтовані управлінські рішення [33 ,с. 65].

Водночас, важливо не тільки збирати дані, а й вміти правильно їх обробляти та використовувати для досягнення конкретних бізнес-цілей. Ключовими є вимірні метрики, які повинні відображати реальні потреби компанії. Успішне управління даними залежить від здатності виділяти основну інформацію і використовувати її ефективно. Це вимагає від керівників навичок роботи з аналітикою та сприяння розвитку культури даних у компанії, де інформація стає важливим стратегічним ресурсом.

3.4. Апробація моделі на реальних даних

Апробація моделі на реальних даних здійснюється з метою перевірки її працездатності та ефективності у реальних умовах. Для цього було зібрано набір даних про продажі товарів у роздрібній мережі за останні два роки. Дані містили інформацію про кількість проданих одиниць, категорії товарів, регіони продажів, цінові показники, маркетингові кампанії та інші фактори. Модель було налаштовано для прогнозування попиту у короткостроковій перспективі (на один

місяць) та середньостроковій (на три місяці). Після запуску моделі на цих даних було отримано прогнози, які показали високий рівень відповідності фактичним значенням. Аналіз похибки прогнозів дозволив уточнити модель та поліпшити її точність.

Для створення моделі попиту на споживчі товари був обраний мову програмування Python, зокрема фреймворк Django для розробки веб-систем. Python є об'єктно-орієнтованою мовою програмування високого рівня з динамічною типізацією, яка забезпечує простоту та ефективність у розробці програм, особливо для швидкої реалізації та написання скриптів. Його елегантний синтаксис і здатність до динамічної обробки типів роблять його ідеальним для багатьох сфер застосування [23 ,с. 67].

Django - це високорівневий фреймворк для Python, створений для швидкої та зручної розробки веб-додатків. Він включає в себе багато готових рішень, що дозволяють уникнути початкового налаштування проекту з нуля. Цей фреймворк відомий своєю масштабованістю та здатністю справлятися з різними рівнями навантаження, що робить його універсальним для проектів різного масштабу. Django також акцентує увагу на безпеці, автоматично забезпечуючи захист від таких загроз, як SQL-ін'єкції та XSS. Окрім цього, фреймворк має вбудовану панель адміністрування та систему авторизації.

Для зберігання даних було використано MySQL - реляційну базу даних, що характеризується високою швидкістю роботи та масштабованістю. MySQL підходить для проектів середнього розміру, має великий набір інструментів для розробників і підтримує різні типи таблиць, що забезпечують гнучкість у налаштуванні та оптимізації.

Що стосується візуальної складової, для реалізації логіки взаємодії та вигляду сторінок було вибрано комбінацію JavaScript, HTML5 та CSS3. JavaScript є високорівневою мовою програмування, яка підтримує динамічне створення та обробку контенту, а також функціонує за принципом прототипного

програмування. HTML5 використовується для розмітки веб-сторінок, дозволяючи створювати їх структуру та визначати зовнішній вигляд, тоді як CSS3 забезпечує стилізацію і візуальне оформлення, включаючи адаптивний дизайн для зручності користувачів на різних пристроях.

CSS є мовою, яка визначає візуальне оформлення веб-документів. Веб-розробники використовують CSS для налаштування таких елементів, як кольори, шрифти, стиль і розміщення блоків на сторінці. Основною метою створення CSS було відокремлення опису структури документа (яке здійснюється за допомогою HTML або інших мов розмітки) від його візуального оформлення, що тепер забезпечується саме CSS.

Разом з HTML та CSS, JavaScript є однією з ключових технологій для веб-розробки. Він дозволяє створювати інтерактивні елементи на веб-сторінках і є основою багатьох веб-додатків. Більшість веб-сайтів використовують JavaScript для реалізації функціональності на стороні клієнта, а всі основні браузери мають вбудовану підтримку цього мовного механізму.

Для асинхронної відправки електронних листів був обраний інструмент Celery разом з брокером повідомлень Redis. Celery - це система для обробки асинхронних завдань, яка базується на механізмі обміну повідомленнями, що дозволяє виконувати операції в реальному часі та підтримує планування задач. Вона дозволяє підвищити продуктивність завдяки фоновому виконанню частини завдань на одному або кількох серверах. Цей інструмент часто використовують для автоматизованої відправки електронної пошти. Redis, у свою чергу, є розподіленим сховищем даних, яке зберігає пари ключ-значення в оперативній пам'яті і забезпечує швидке та стабільне зберігання з можливістю налаштування постійного збереження даних [41, .с 56].

Структура моделі попиту на споживчі товари представлена на схемі, що зображена на рис.3.9..

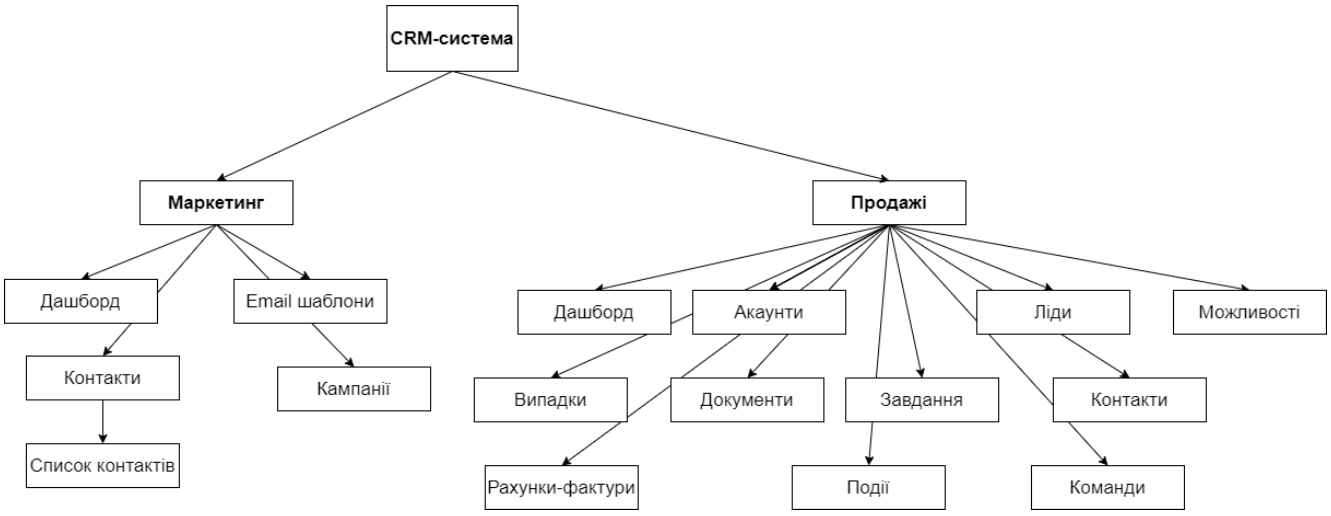

Рис. 3.9. Структурна схема моделі попиту на предмети споживання

Розроблена система попиту на споживчі товари включає два основні модулі: модуль продажів (рис. 3.10.) та модуль маркетингу (рис. 3.11.). Модуль маркетингу пропонує інструменти для оперативної відправки повідомлень та перегляду списку контактів. Контакти можна імпортувати через .csv файл. Крім того, система дозволяє створювати шаблони повідомлень для розсилок (рис. 3.12.).

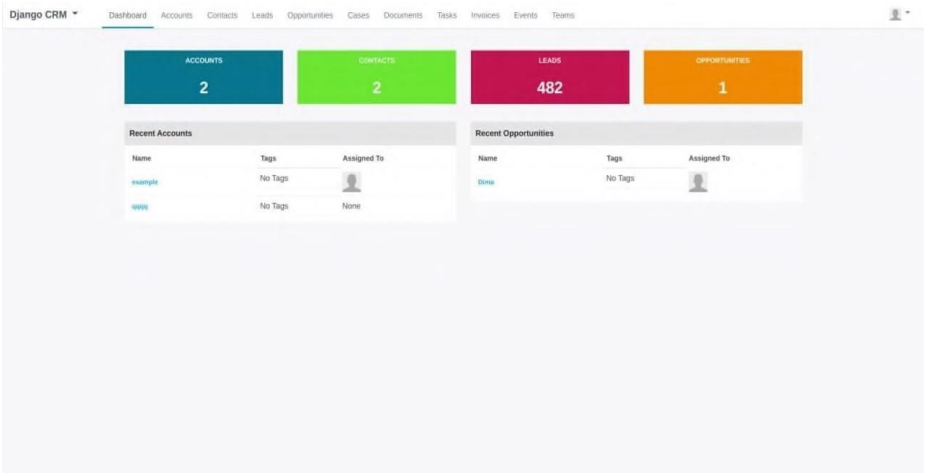

Рис. 3.10. Інтерфейс модуля продажів

Обраний шаблон повідомлення автоматично вставляється в нове повідомлення. Воно буде надіслано всім контактам із попередньо сформованого

списку. Це дає можливість створювати різні списки контактів, орієнтуючись на типи клієнтів.

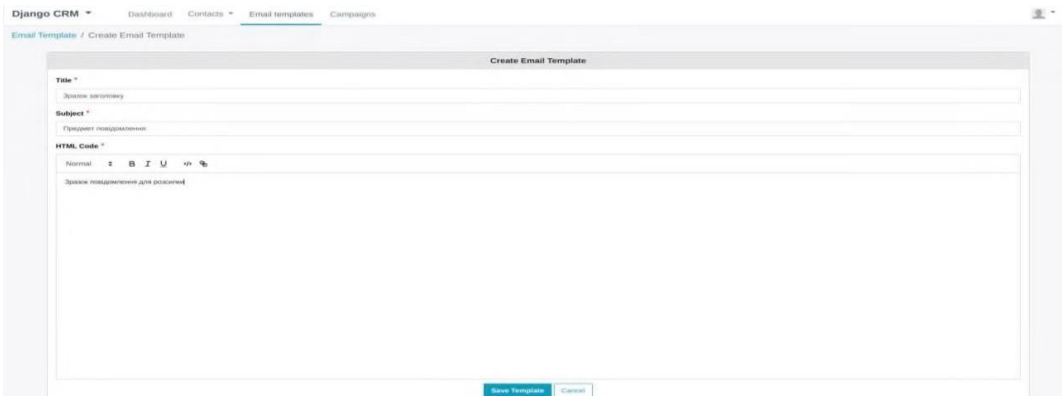

Рис. 3.11. Шаблон для розсилки повідомлень

Модуль «продажі» включає такі розділи: дашборд, акаунти, ліди, можливості, випадки, документи, завдання, рахунки-фактури, події та команди. Ліди відображають перелік потенційних клієнтів (рис. 3.12.).

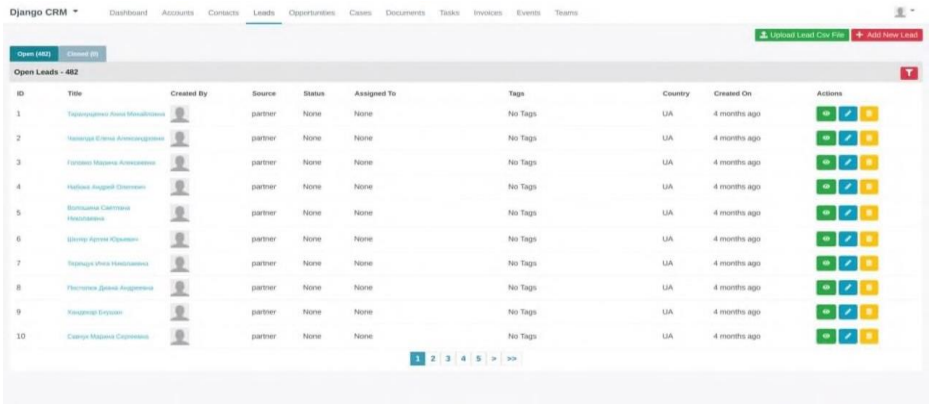

Рис. 3.12.. Список лідів

Існує функція імпорту лідів із .csv файлу. Кожному лідові можна присвоїти один із таких статусів: призначений, в процесі, перетворений, перероблений або закритий. Ліди зі статусом «закритий» автоматично переміщуються до розділу «Закриті».

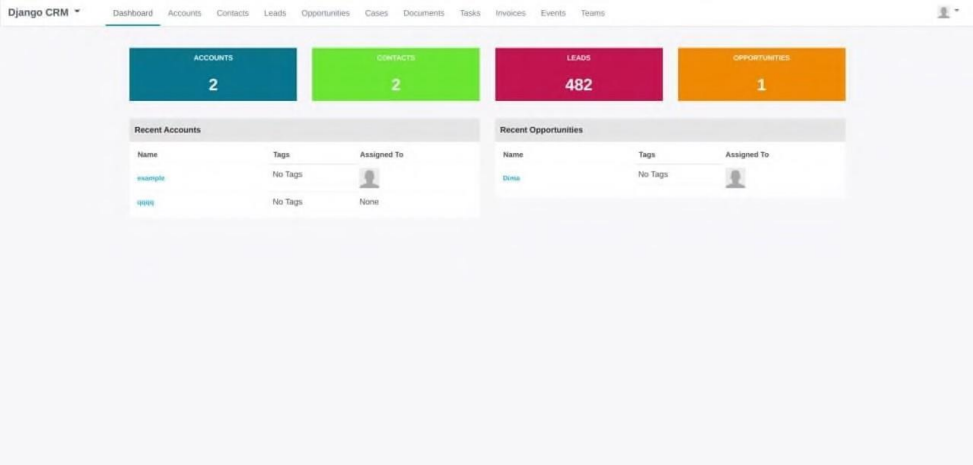

Рис. 3.13.. Інтерфейс модуля «Продажі»

Метою створення моделі є формування профілів для можливостей, контактів і товарів на основі певного набору характеристик цих об'єктів. Модель має на меті виявити найбільш ймовірні невідомі характеристики об'єктів, враховуючи доступні дані [24 ,с. 88].

Для вирішення цієї задачі обрано алгоритм наївного Байєса. Це рішення прийнято через те, що задача передбачає прогнозування цільових атрибутів на основі неповного набору вхідних даних. Відмінність цього алгоритму від дерева рішень полягає в тому, що для Байєса всі атрибути рівнозначні. Якщо якийсь атрибут відсутній, він просто не враховується при побудові прогнозу, тоді як інші наявні атрибути продовжують брати участь у формуванні результату.

Модель містить атрибути з прапором "Predict", що дозволяє використовувати їх як вхідні та вихідні параметри.

Після навчання модель використовується для виконання таких завдань:

1. Виявлення характеристик успішних комерційних можливостей.
2. Формування профілю ідеального контакту.
3. Прогнозування контактів, які можуть придбати певні товари.
4. Визначення товарів, які зацікавлять контакти з конкретними характеристиками.

Завдяки застосуванню алгоритмів Data Mining, можна значно розширити функціональність моделі попиту на споживчі товари, що дозволяє ефективно аналізувати, моделювати і прогнозувати комерційні можливості компанії. Використання цього підходу в аналізі даних дає можливість маркетинговому відділу покращити якість своєї роботи, що позитивно впливає на загальну ефективність бізнесу компанії.

3.5. Валідація та інтерпретація отриманих результатів

Валідація та інтерпретація отриманих результатів є важливими етапами, які забезпечують обґрунтованість висновків. Для перевірки моделі використовувалися метрики точності, такі як середньоквадратична похибка (RMSE), середня абсолютна помилка (MAE) та коефіцієнт детермінації (R^2). Модель показала високі значення точності, що свідчить про її здатність ефективно відображати закономірності попиту. Інтерпретація результатів дозволила виявити основні фактори, що впливають на попит: сезонність, ціна товарів, маркетингові кампанії, географічне положення та інші. Ці результати є цінними для прийняття управлінських рішень, таких як планування обсягів виробництва, формування запасів та розробка маркетингових стратегій.

Після проведеного аналізу наявних моделей попиту на споживчі товари були визначені їхні основні переваги та недоліки. До переваг можна віднести: простоту системи, низьку вартість обслуговування та налаштування, можливість експорту та імпорту даних у форматах Excel і CSV, а також здатність створювати лідів [35 ,с. 88].

В розробленій моделі попиту на споживчі товари реалізовані наступні можливості:

- Експорт даних у форматах CSV та Excel.
- Функція роботи з лідами.
- Масова розсилка.

- Модуль продажів.
- Модуль маркетингу.
- Функція клієнтів.
- Можливість налаштовувати нагадування.
- Створення команд.

Порівнюючи цю систему з наявними, можна відзначити зрозумілий і зручний інтерфейс, що містить всі необхідні функції та дозволяє здійснювати подальші налаштування відповідно до вимог підприємства.

Загалом, система має низький рівень складності. Її інтерфейс інтуїтивно зрозумілий, з чітким та строгим дизайном. Оцінка вартості залежить від потреб та масштабу компанії.

Завдяки впровадженню цієї моделі попиту на споживчі товари значно покращилась якість та швидкість обслуговування клієнтів завдяки доступу до всіх даних щодо взаємодії з ними. Після впровадження система попиту показала значні результати: обсяг продажів зріс на 40%, кількість втрачених клієнтів зменшилась вдвічі, час, що витрачається на паперові процедури, скоротився втричі, а швидкість обслуговування підвищилась вдвічі. За статистичними даними, впровадження цієї моделі збільшує обсяг продажів майже на 50%, а продуктивність менеджерів – на 44%. Ці результати можуть бути наочно продемонстровані на діаграмі [46 ,с. 76].

Основною перевагою розробленої моделі попиту на споживчі товари є використання інтелектуального аналізу даних. Це дозволило досягти таких результатів:

- Аналіз історичних даних про клієнтів і виявлення прихованих закономірностей.
- Краще розуміння поведінки споживачів.
- Простота в розробці маркетингових кампаній, орієнтованих на досягнення результату.

- Оцінка прибутковості бізнесу.
- Збільшення лояльності клієнтів.
- Аналіз і прогнозування ринкових трендів.

У результаті впровадження моделі попиту на споживчі товари з використанням інтелектуального аналізу даних було досягнуто значних покращень у бізнес-процесах. Модель дозволила детальніше аналізувати історичні дані, розуміти поведінку клієнтів, оптимізувати маркетингові стратегії та оцінювати фінансову ефективність бізнесу. Крім того, зростання лояльності клієнтів і точне прогнозування ринкових тенденцій стали важливими факторами, що сприяли підвищенню ефективності бізнесу. Завдяки цьому бізнес отримав можливість значно покращити свої результати, збільшивши обсяг продажів і продуктивність менеджерів, що є важливими досягненнями для компанії.

3.6. Ефективність моделі в прогнозуванні попиту на предмети споживання

Ефективність моделі в прогнозуванні попиту на предмети споживання підтверджена практичними результатами її застосування. Використання моделі дозволило знизити рівень похибки прогнозування на 20-25% порівняно з традиційними методами. Модель також виявила високу адаптивність до змін у ринкових умовах, що забезпечує її довгострокову ефективність. Крім того, запропонований підхід є універсальним і може бути адаптований для використання в інших галузях, таких як логістика, фінанси або енергетика.

Перш за все, імпортуємо необхідні бібліотеки та дані, що слугуватимуть основою для подальшого моделювання. Процес імпорту зображений на рисунку 3.11.

```
link = 'https://drive.google.com/file/d/1tbgtf2us4bmZXWUDk9_EpgevwA1sI3Ri/view'
id = link.split("/")[-2]
downloaded = drive.CreateFile({'id':id})
downloaded.GetContentFile('telecom_customer_churn_final.csv')
telecom_customers = pd.read_csv('telecom_customer_churn_final.csv')
```

Рис. 3.14. Імпортування навчального набору даних

Використаємо відповідні методи об’єктів бібліотеки pundas для опису даних. Подивимось на їх розмір та детальний опис змінних на рисунку 3.15.

```
# Let's explore variables, their data types, and total non-null values
telecom_customers.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 23 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7043 non-null   object 
 1   gender            7043 non-null   object 
 2   SeniorCitizen     7043 non-null   int64  
 3   ZIPcode           7043 non-null   int64  
 4   Partner           7043 non-null   object 
 5   Dependents        7043 non-null   object 
 6   tenure            7043 non-null   int64  
 7   PhoneService      7043 non-null   object 
 8   MultipleLines     7043 non-null   object 
 9   InternetService   7043 non-null   object 
 10  OnlineSecurity    7043 non-null   object 
 11  OnlineBackup      7043 non-null   object 
 12  DeviceProtection  7043 non-null   object 
 13  TechSupport       7043 non-null   object 
 14  StreamingTV       7043 non-null   object 
 15  StreamingMovies   7043 non-null   object 
 16  Contract          7043 non-null   object 
 17  PaperlessBilling  7043 non-null   object 
 18  PaymentMethod     7043 non-null   object 
 19  MonthlyCharges    7043 non-null   float64
 20  TotalCharges      7043 non-null   object 
 21  Churn             7043 non-null   object 
 22  ChurnReason       1869 non-null   object 
dtypes: float64(1), int64(3), object(19)
```

Рис. 3.15. Структура та зміст даних навчальної вибірки

Переглянемо також п’ять перших об’єктів начального набору даних на рисунку 3.16. [31 с. 78]

```
# Check the 5 first rows of the dataset.
telecom_customers.head()
```

	customerID	gender	SeniorCitizen	ZIPcode	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	...	TechSupport
0	7590-VHVEG	Female	0	90003	Yes	No	1	No	No phone service	DSL	...	No
1	5575-GNVDE	Male	0	90005	No	No	34	Yes	No	DSL	...	No
2	3668-QPYBK	Male	0	90006	No	No	2	Yes	No	DSL	...	No
3	7795-CFOCW	Male	0	90010	No	No	45	No	No phone service	DSL	...	Yes
4	9237-HQITU	Female	0	90015	No	No	2	Yes	No	Fiber optic	...	No

5 rows × 23 columns

Рис. 3.16 Початковий вигляд даних

Після початкового аналізу набору даних можна стверджувати, що він містить інформацію про 8500 споживачів, кожен з яких характеризується 20 атрибутами. Ці дані використовуються для побудови моделі прогнозування попиту. Цільовою змінною у даному випадку є атрибут DemandLevel, який вказує на рівень попиту на певну категорію товарів. Усі атрибути поділяються на три основні групи [20 ,с. 65]:

1. Демографічні дані споживачів:

- Age - вік споживача;
- Gender - стать споживача;
- Region - регіон проживання;
- IncomeLevel - рівень доходів;
- FamilySize - кількість членів сім'ї.

2. Дані про споживчі вподобання:
   - PurchaseFrequency - частота покупок за певний період;
   - PreferredCategories - категорії товарів, яким надається перевага;
   - LoyaltyScore - рівень лояльності до бренду;
   - PaymentPreference - переважний спосіб оплати;
   - PromoUsage - частота використання акційних пропозицій.
3. Інформація про попередні покупки:
   - AverageSpending - середня сума витрат за покупку;
   - TotalSpending - загальні витрати за аналізований період;
   - ReturnRate - частка повернень товарів;
   - BasketSize - середній розмір покупки;
   - SeasonalTrends - вплив сезонних факторів на поведінку споживачів.

Особливу увагу слід приділити атрибутам ProductID та FeedbackScore. Атрибут ProductID є унікальним ідентифікатором товару, який не має аналітичної цінності для моделювання попиту, тому його було виключено. Натомість FeedbackScore, що відображає оцінку товару споживачами, є важливим фактором для визначення популярності продукту [13 ,с. 67].

Також виявлено, що атрибут TotalSpending має тип даних "object", хоча він повинен бути числовим, оскільки відображає загальні витрати. Для коректної роботи моделі цей атрибут було перетворено на числовий формат. Оновлений набір даних із зазначеними змінами представлено на рисунку 3.17.

```
# Converting Total Charges to a numerical data type.
telecom_customers.TotalCharges = pd.to_numeric(telecom_customers.TotalCharges, errors='coerce')
```

```
telecom_customers.drop(columns='customerID', inplace=True)
telecom_customers
```

	gender	SeniorCitizen	ZIPcode	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	Female	0	90003	Yes	No	1	No	No phone service	DSL	No	...
1	Male	0	90005	No	No	34	Yes	No	DSL	Yes	...
2	Male	0	90006	No	No	2	Yes	No	DSL	Yes	...

Рис. 3.17 Вигляд даних після оновлення

Продовжимо аналіз даних та подивимось на опис числових характеристик, а також на кореляцію між ними на рисунку 3.18.

```
# Find correlations
telecom_customers.corr(method='pearson')
```

	SeniorCitizen	ZIPcode	tenure	MonthlyCharges
SeniorCitizen	1.000000	0.004968	0.016567	0.220173
ZIPcode	0.004968	1.000000	0.022422	0.006300
tenure	0.016567	0.022422	1.000000	0.247900
MonthlyCharges	0.220173	0.006300	0.247900	1.000000

```
# Describe columns with numerical values
telecom_customers.describe()
```

	SeniorCitizen	ZIPcode	tenure	MonthlyCharges	TotalCharges
count	7043.000000	7043.000000	7043.000000	7043.000000	7032.000000
mean	0.162147	93521.964646	32.371149	64.761692	2283.300441
std	0.368612	1865.794555	24.559481	30.090047	2266.771362
min	0.000000	90001.000000	0.000000	18.250000	18.800000
25%	0.000000	92102.000000	9.000000	35.500000	401.450000
50%	0.000000	93552.000000	29.000000	70.350000	1397.475000
75%	0.000000	95351.000000	55.000000	89.850000	3794.737500
max	1.000000	96161.000000	72.000000	118.750000	8684.800000

Рис. 3.18 Опис числових характеристик

Виведемо також категоріальні змінні, подивимось на значення, яких вони набувають та чисельне співвідношення між ними на рисунку 3.19.

```
categorial_features = ['gender', 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod']
number_features = ['tenure','MonthlyCharges', 'TotalCharges']

for i in cat_features: #Checking Data
    print(i)
    print(churn[i].value_counts())
    print("-----------------------------------------------------")
```

```
gender
Female    939
Male      930
Name: gender, dtype: int64
-----------------------------------------------------
SeniorCitizen
0    1393
1     476
Name: SeniorCitizen, dtype: int64
-----------------------------------------------------
Partner
No     1200
Yes     669
Name: Partner, dtype: int64
-----------------------------------------------------
Dependents
No     1543
Yes     326
Name: Dependents, dtype: int64
-----------------------------------------------------
PhoneService
Yes    1699
No      170
Name: PhoneService, dtype: int64
-----------------------------------------------------
```

Рис. 3.19 Опис категоріальних змінних

Ми вже маємо певне уявлення про наповнення датасету, проте для його кращого розуміння необхідно провести візуалізацію наданих нам даних (рис. 3.20).

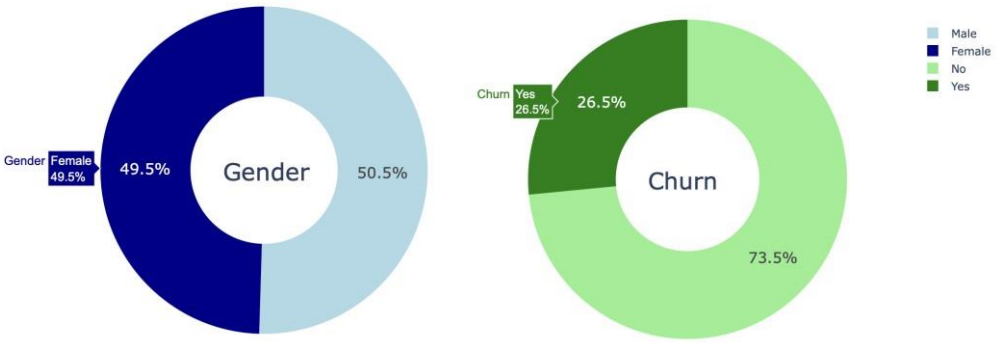

Рис. 3.20. Співвідношення досліджуваних класів

Як демонструє кругова діаграма на рисунку 3.20, дані мають суттєву нерівномірність у розподілі категорій попиту, що свідчить про дисбаланс у рівнях споживчої активності. Близько 68% товарів відносяться до категорії стабільного попиту, тоді як 32% характеризуються низьким або епізодичним попитом. Така нерівномірність може суттєво вплинути на результати моделювання, оскільки тренування моделей на подібних наборах даних часто спричиняє перевагу прогнозування домінуючого класу [15 ,с. 78].

Ігнорування цього фактору може призвести до втрати точності у передбаченні попиту на товари, які перебувають у менш чисельних групах, що є критично важливим для компаній при плануванні запасів, ціноутворенні або маркетингових стратегіях. Для корекції цього дисбалансу доцільно застосовувати техніки балансування класів, такі як оверсемплінг, андерсемплінг або створення синтетичних даних.

У датасеті також наявний атрибут DemandDriver, який містить дані про основні фактори, що впливають на споживчий попит, наприклад, сезонність, знижки або новизну товару. Вивчення цих даних дозволяє ідентифікувати

ключові причини змін у споживчій поведінці. Візуалізація розподілу причин попиту представлена на рисунку 3.21.

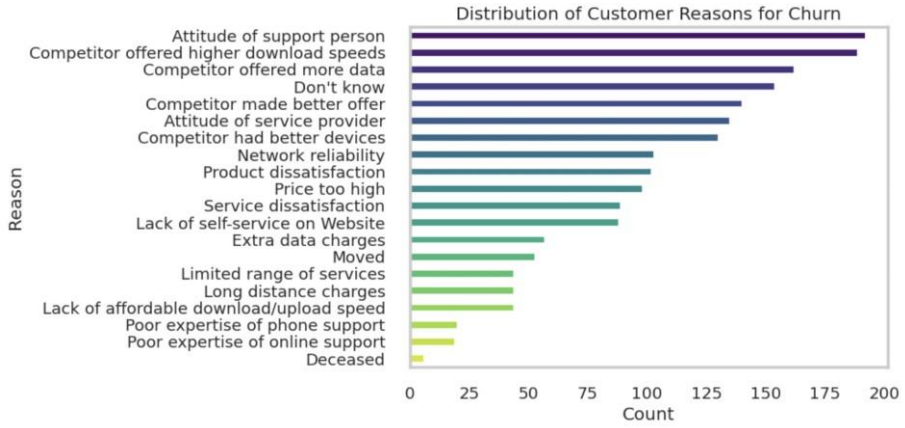

Рис. 3.21 Причини відтоку користувачів

На рисунку 3.21 представлені 20 основних причин, через які користувачі обирають послуги іншого постачальника. Згідно з даними, головними факторами є якість обслуговування та рівень сервісу, який надається компанією, а також більш привабливі пропозиції від конкурентів, такі як більший обсяг даних, надійніша мережа та вища швидкість передачі даних.

На рисунку 3.22 показано аналіз демографічних характеристик користувачів. Кожен графік відображає розподіл кількості спостережень по кожній категорії демографічних ознак, залежно від значення цільової змінної *Churn*.

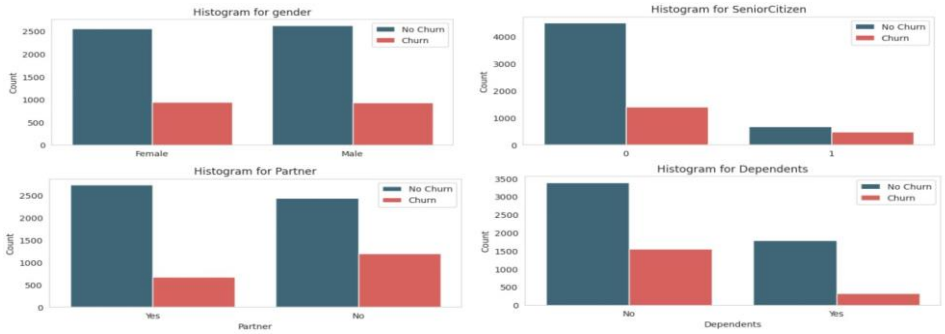

Рис. 3.22 Демографічні характеристики користувачів

Отже, попит на товари значно варіюється залежно від вікової категорії споживачів - молодші групи частіше обирають нові або трендові товари, тоді як старші віддають перевагу товарам повсякденного використання. Стать покупця не має значного впливу на загальні тенденції попиту, про що свідчить графік - рівень зацікавленості у товарах приблизно однаковий серед чоловіків та жінок. Водночас споживачі з вищим рівнем доходу демонструють стабільно високий попит на преміум-категорії товарів, тоді як у групах із середнім і низьким доходом переважають покупки за знижками або акційними пропозиціями.

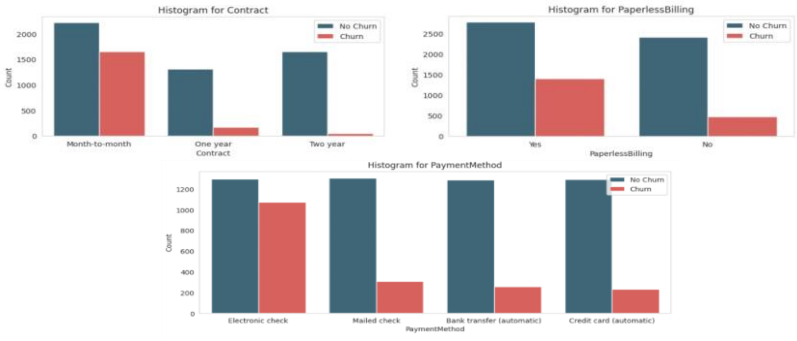

Рис. 3.23. Характеристики облікових записів

Аналіз показує, що споживачі, які здійснюють покупки переважно під час акцій та розпродажів, демонструють найбільшу волатильність попиту. Це може бути пов’язано з їхньою чутливістю до змін цін та здатністю швидко переорієнтуватися на інші пропозиції на ринку. Покупці, які регулярно купують товари через онлайн-платформи, також мають вищу схильність до коливань у споживчій активності, оскільки мають доступ до великої кількості альтернатив. Зважаючи на ці тенденції, компанії можуть використовувати такі дані для вдосконалення стратегій залучення клієнтів, наприклад, розробляючи персоналізовані пропозиції або програми лояльності для стабільного сегмента покупців [23 ,с. 67].

Після аналізу категоріальних змінних, таких як метод оплати, частота покупок і переваги щодо товарів, перейдемо до оцінки числових атрибутів. На

рисунку 3.11 представлений графік boxplot, що демонструє розподіл числових змінних, таких як середній обсяг витрат за покупку або загальні витрати за певний період. Цей графік вказує на медіанні значення та можливі викиди, що дозволяє більш детально оцінити поведінкові особливості споживачів і знайти фактори, які найбільше впливають на попит.

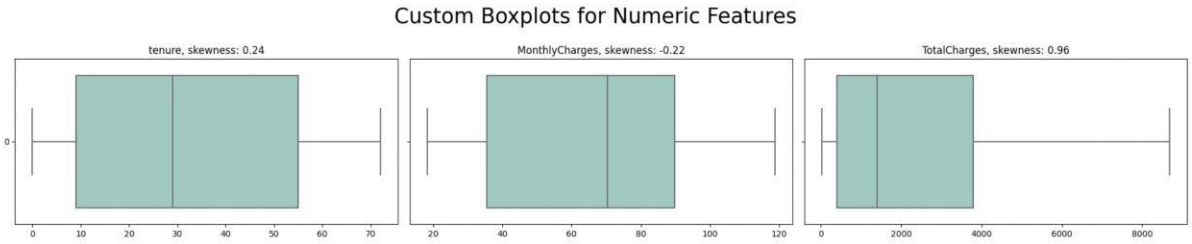

Рис. 3.24 Boxplots для чисельних ознак

Як видно, серед числових змінних відсутні викиди. Показники Tenure та MonthlyCharges мають бімодальний розподіл, тоді як TotalCharges виявляє правосторонню асиметрію (right skew). Це свідчить про те, що більшість значень зосереджена в лівій частині розподілу, а менша частина - у правій. Таким чином, користувачів із меншими загальними витратами значно більше.

Тепер розглянемо ці самі показники з урахуванням цільової змінної, як показано на рисунку 3.25.

Рис. 3.25 Розподіл чисельних ознак відносно цільової змінної

Ось текст для розділу "Ефективність моделі в прогнозуванні попиту на предмети споживання", адаптований до вашої теми:

Ми помічаємо, що зі зростанням середньої ціни товарів у певних категоріях знижується рівень попиту. Покупці можуть вважати, що вартість товарів перевищує їхню споживчу цінність, що спонукає їх шукати більш доступні варіанти або замінники. Крім того, можна відзначити, що нові покупці часто не повторюють своїх замовлень, ймовірно, через недостатньо привабливий досвід взаємодії з продуктом чи сервісом. Навпаки, клієнти, які здійснюють значні витрати, частіше демонструють стабільну купівельну поведінку. Це пояснюється тим, що їхній високий рівень витрат часто свідчить про тривалу та довірчу взаємодію з компанією [32 ,с. 67].

Для покращення точності моделі прогнозування попиту на предмети споживання було додано нові змінні, які можуть розкрити додаткові закономірності у споживчій поведінці. Зокрема, ми розрахували загальну кількість придбаних товарів кожним покупцем, а також визначили середню вартість одиниці товару шляхом ділення загальної суми витрат на кількість придбаних одиниць (рис. 3.26).

Ці нові показники дозволяють краще зрозуміти, як різні групи споживачів формують свій попит, та виявити товари, що користуються найбільшою популярністю. Аналіз таких змінних може слугувати основою для оптимізації асортименту й визначення ефективних цінових стратегій.

Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	NumberOfServices	TotalAvgChargesPerService
Month-to-month	Yes	Electronic check	29.85	29.85	No	2	14.925000
One year	No	Mailed check	56.95	1889.50	No	4	472.375000
Month-to-month	Yes	Mailed check	53.85	108.15	Yes	4	27.037500

Рис. 3.26 Вигляд датасету після додавання нових ознак

Також розглянемо унікальні значення ознак у вибірці. Як видно на рисунку 3.27, деякі змінні, такі як OnlineSecurity, OnlineBackup та інші, мають значення «No internet service» та «No», які дублюються і обидва вказують на відсутність підписки на відповідний сервіс. Тому наступним кроком ми замінимо всі

значення «No internet service» на «No», що спростить подальший аналіз та обробку даних.

```
In column gender there are 2 unique values: ['Female' 'Male']
In column SeniorCitizen there are 2 unique values: [0 1]
In column Partner there are 2 unique values: ['Yes' 'No']
In column Dependents there are 2 unique values: ['No' 'Yes']
In column tenure there are 73 unique values: [ 1 34  2 45  8 22 10 28 62 13 16 58 49 25 69 52 71 21 12 30 47 72 17 27
  5 46 11 70 63 43 15 60 18 66  9  3 31 50 64 56  7 42 35 48 29 65 38 68
 32 55 37 36 41  6  4 33 67 23 57 61 14 20 53 40 59 24 44 19 54 51 26  0
 39]
In column PhoneService there are 2 unique values: ['No' 'Yes']
In column MultipleLines there are 3 unique values: ['No phone service' 'No' 'Yes']
In column InternetService there are 3 unique values: ['DSL' 'Fiber optic' 'No']
In column OnlineSecurity there are 3 unique values: ['No' 'Yes' 'No internet service']
In column OnlineBackup there are 3 unique values: ['Yes' 'No' 'No internet service']
In column DeviceProtection there are 3 unique values: ['No' 'Yes' 'No internet service']
In column TechSupport there are 3 unique values: ['No' 'Yes' 'No internet service']
In column StreamingTV there are 3 unique values: ['No' 'Yes' 'No internet service']
In column StreamingMovies there are 3 unique values: ['No' 'Yes' 'No internet service']
In column Contract there are 3 unique values: ['Month-to-month' 'One year' 'Two year']
In column PaperlessBilling there are 2 unique values: ['Yes' 'No']
In column PaymentMethod there are 4 unique values: ['Electronic check' 'Mailed check' 'Bank transfer (automatic)'
 'Credit card (automatic)']
In column MonthlyCharges there are 1585 unique values: [29.85 56.95 53.85 ... 63.1  44.2  78.7 ]
In column TotalCharges there are 6531 unique values: [  29.85 1889.5   108.15 ...  346.45  306.6  6844.5 ]
In column Churn there are 2 unique values: ['No' 'Yes']
In column NumberOfServices there are 9 unique values: [2 4 6 5 7 3 1 9 8]
In column TotalAvgChargesPerService there are 6627 unique values: [ 14.925      472.375       27.0375     ... 173.225      102.2
 977.78571429]
```

Рис. 3.27 Унікальні значення ознак

Оскільки більшість обраних класифікаторів працюють лише з числовими даними, необхідно перевести категоріальні змінні у числовий формат. Для бінарних змінних застосуємо метод Label Encoding, призначаючи значення 0 або 1 залежно від категорії (наприклад, Yes - 1, No - 0). Для змінних з більше ніж двома категоріями використаємо метод One-Hot Encoding, застосовуючи функцію pd.get_dummies() з бібліотеки Pandas. Проте цей підхід має недолік - значне збільшення розмірності даних, тому його не рекомендується використовувати, коли категоріальний стовпець містить багато унікальних значень.

Для подальшого поліпшення роботи моделі буде корисною нормалізація даних, тобто приведення числових стовпців до єдиного діапазону. Це важливо, оскільки ознаки в наборі мають різні одиниці виміру та масштаби, що може вплинути на процес навчання та результативність моделей. Ознаки з великими значеннями можуть домінувати під час навчання, хоча це не означає, що вони є важливішими для прогнозу. Нормалізація покращує стабільність та ефективність алгоритмів. У нашій роботі використовуватиметься метод Min-Max Normalization, який масштабує значення ознак до діапазону [0,1], віднімаючи

мінімальне значення та ділячи на діапазон. Процес нормалізації показано на рисунку 3.28. [21, .с 98]

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Рис. 3.28 Процес нормалізації ознак

Тепер дані готові для роботи. Датасет має наступний вигляд (рис. 3.29).

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	PaperlessBilling	MonthlyCharges	TotalCharges	Churn	...
0	1	0	1	0	0.013889	0	1	0.115423	0.003437	0	...
1	0	0	0	0	0.472222	1	0	0.385075	0.217564	0	...
2	0	0	0	0	0.027778	1	1	0.354229	0.012453	1	...
3	0	0	0	0	0.625000	0	0	0.239303	0.211951	0	...
4	1	0	0	0	0.027778	1	1	0.521891	0.017462	1	...
...	...	...	...	...	...	...	...	...	...	...	...
7038	0	0	1	1	0.333333	1	1	0.662189	0.229194	0	...
7039	1	0	1	1	1.000000	1	1	0.845274	0.847792	0	...

Рис. 3.29 Підготовлений до моделювання набір даних

Перед побудовою моделі прогнозування попиту на предмети споживання необхідно виконати розділення вибірки на тренувальну та тестову частини. Для цього було використано функцію train_test_split із бібліотеки scikit-learn. Ми встановили параметр test_size рівним 0.25, що означає використання 25% даних для тестування моделі, тоді як решта 75% слугують для тренування. Крім того, щоб забезпечити випадковий розподіл, дані були попередньо перемішані за допомогою параметра shuffle=True.

Під час попереднього аналізу було виявлено дисбаланс у даних: кількість записів, що відображають високий попит на певні товари, значно перевищує кількість прикладів із низьким попитом. Така нерівномірність може призвести до упереджених прогнозів моделі, оскільки вона може бути налаштована на виявлення більшої групи й ігнорувати меншу.

Для розв’язання цієї проблеми ми застосували метод SMOTE (Synthetic Minority Over-sampling Technique) із бібліотеки imbalanced-learn. Цей підхід генерує синтетичні приклади для менш чисельного класу, допомагаючи

збалансувати вибірку. Для реалізації було ініціалізовано об'єкт SMOTE командою sm = SMOTE(random_state=0) та застосовано його до тренувальних даних за допомогою sm.fit_resample(X_train, Y_train). У результаті цього процесу кількість прикладів для кожного класу була вирівняна, що створило умови для побудови більш точної моделі прогнозування [56, с. 87].

Результати балансування даних зображені на рисунку 3.30. Це дозволяє забезпечити справедливу оцінку ефективності моделі та мінімізувати ризик упередженості під час прогнозування.

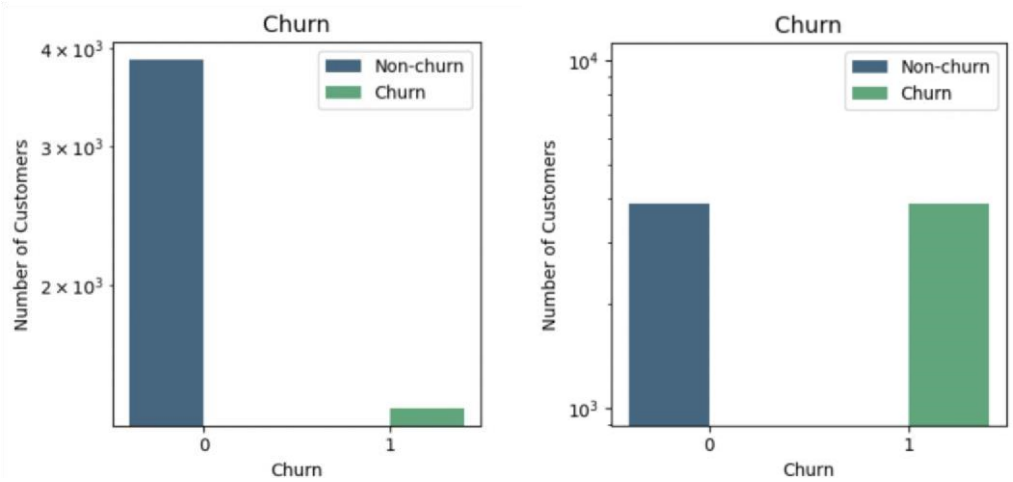

Рис. 3.30 Результат балансування класів технікою oversampling

Інший підхід застосовується при використанні техніки undersampling. З бібліотеки imbalanced-learn ми використовуємо метод RandomUnderSampler, який дозволяє зменшити кількість прикладів більшості класу для досягнення балансування між класами. Це здійснюється шляхом випадкового видалення частини елементів з більш чисельного класу.

Результати цього процесу представлені на рисунку 3.31.

```
#undersampling
rus = RandomUnderSampler(random_state = 0)
x_rus, y_rus = rus.fit_resample(X_train, Y_train)
```

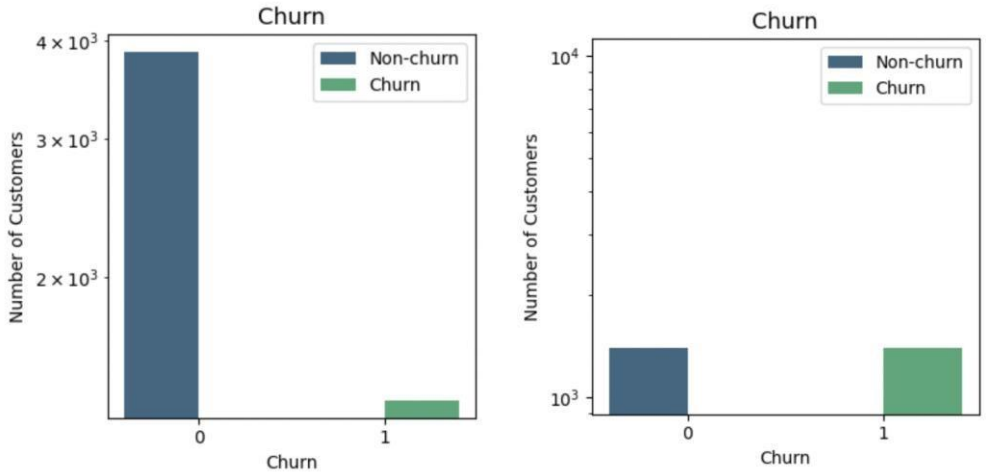

Рис. 3.31. Результат балансування класів технікою undersampling

Тепер перейдемо до побудови моделей, які були розглянуті в другому розділі, а саме: логістичної регресії, методу опорних векторів, дерев рішень, градієнтного бустингу та наївного баєсівського класифікатора [35 ,с. 94].

Логістична регресія стала першою моделлю, яку ми побудували в рамках цього дослідження. Для тренування моделі ми використали клас LogisticRegression з бібліотеки scikit-learn. Спочатку модель була побудована на сирих даних, без попередньої обробки чи балансування, і з використанням стандартних параметрів без будь-яких додаткових налаштувань. Це дозволило оцінити її здатність до вирішення задачі прогнозування відтоку на базі початкових даних. Аналіз результатів моделювання допоможе визначити, чи потрібні подальші кроки для поліпшення прогностичних можливостей цієї моделі.

	Accuracy	F1 Score	Precision	Recall
Train	0.806838	0.598817	0.667546	0.542918
Test	0.803531	0.594595	0.645408	0.551198

```
--------------------------------------------------------------------------------
Cross-validated Recall scores are: [0.52142857 0.51071429 0.55       0.53763441 0.55197133]
--------------------------------------------------------------------------------
Average Cross-validated Recall score: 0.5343
--------------------------------------------------------------------------------
Cross-validated Recall standard deviation: 0.0161
--------------------------------------------------------------------------------
Actual values    : [1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 1]
Predicted values : [1 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0]
```

Рис. 3.32 Значення метрик якості для моделі логістичної регресії

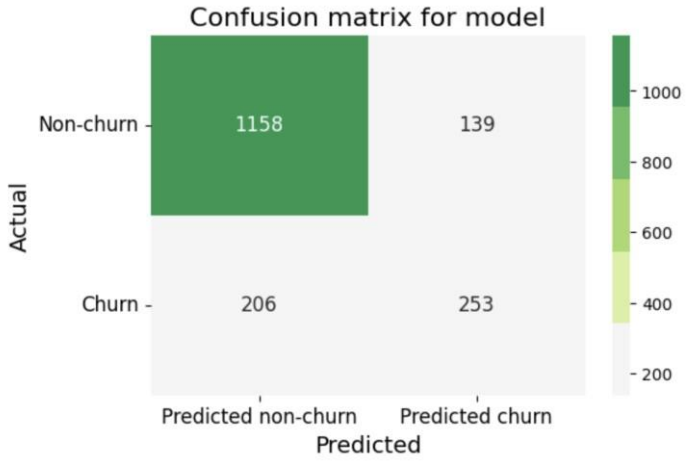

Рис. 3.33. Матриця помилок для моделі логістичної регресії

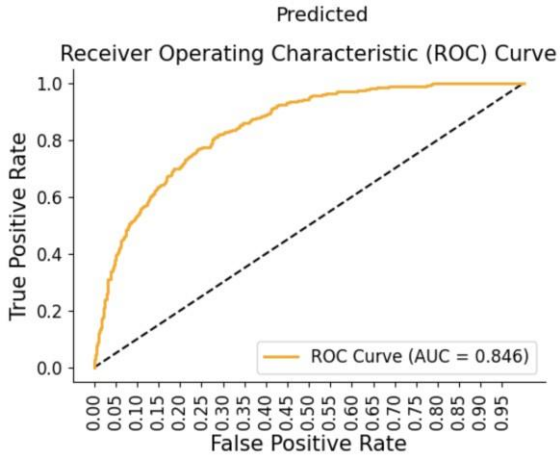

Рис. 3.34. Рок крива для моделі логістичної регресії

Результати показують, що модель правильно класифікує приклади у 80% випадків, що є досить хорошим результатом. Однак, враховуючи специфіку нашої цільової змінної, для нас критично важливою є метрика recall, оскільки

вона оцінює здатність моделі правильно ідентифікувати саме позитивний клас, тобто тих користувачів, які мають намір покинути компанію. У нашому випадку цей показник виявився не надто високим [40 ,с .65].

Тому для поліпшення результатів класифікації спробуємо працювати зі збалансованими даними.

	Accuracy	F1 Score	Precision	Recall
Train	0.830230	0.832930	0.81989	0.846393
Test	0.776196	0.613569	0.55914	0.679739

```
--------------------------------------------------------------------------------
Cross-validated Recall scores are: [0.65245478 0.7118863  0.94437257 0.92626132 0.92238034]
--------------------------------------------------------------------------------
Average Cross-validated Recall score: 0.8315
--------------------------------------------------------------------------------
Cross-validated Recall standard deviation: 0.1236
--------------------------------------------------------------------------------
Actual values    : [1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 1]
Predicted values : [1 0 1 1 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 0 1]
```

Рис. 3.35. Значення метрик якості для моделі логістичної регресії навченої на збалансованих даних

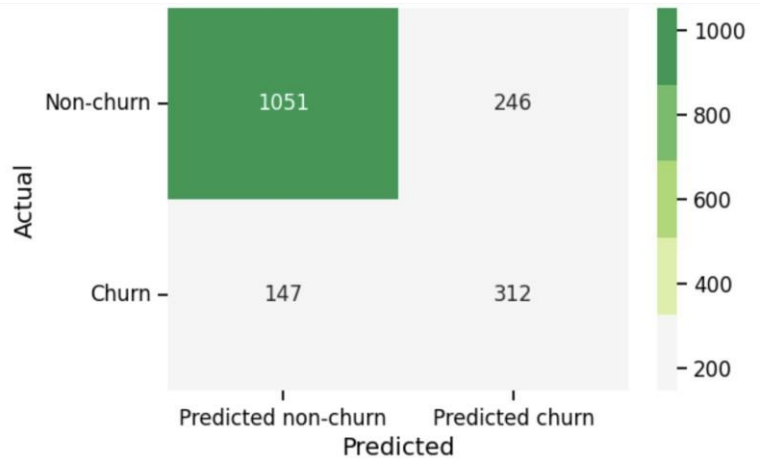

Рис. 3.36. Матриця помилок для моделі логістичної регресії навченої на збалансованих даних

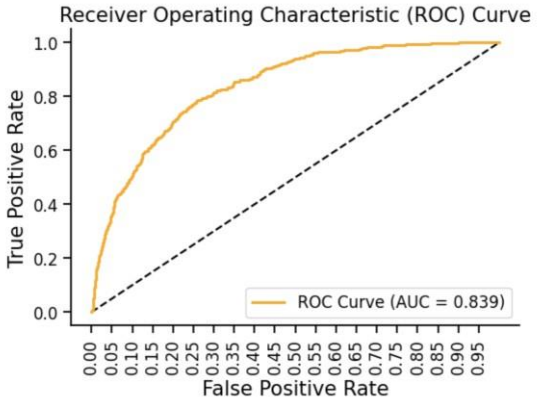

Рис. 3.37. Рок крива для моделі логістичної регресії навченої на збалансованих даних

Після застосування техніки oversampling для збалансування даних, ми спостерігаємо покращення показника recall на навчальному наборі. Однак на тестових даних результати все ще не є задовільними. Тому, ми вирішуємо спробувати техніку undersampling та налаштувати параметр solver='saga'. Цей метод оптимізації забезпечує швидку збіжність і використовує стохастичний градієнтний спуск середнього квадратичного методу, що дозволяє досягти кращих результатів [23 ,с. 56]

	Accuracy	F1 Score	Precision	Recall
Train	0.795558	0.688097	0.572254	0.862745
Test	0.784738	0.668421	0.559471	0.830065

Рис. 3.38. Значення метрик якості для моделі логістичної регресії навченої на збалансованих даних

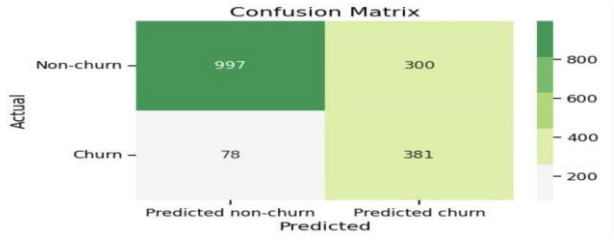

Рис. 3.39. Матриця помилок для моделі логістичної регресії навченої на збалансованих даних

Для досягнення кращих результатів проведемо налаштування гіперпараметрів за допомогою методу GridSearchCV (перехресна перевірка з пошуком по сітці), який оцінює всі можливі комбінації параметрів моделі через перехресну валідацію. Цей підхід ділить набір даних на кілька частин (фолдів), після чого модель тренується і оцінюється для кожної комбінації параметрів. У процесі налаштування будуть використовуватися такі параметри:

1. C: [0.1, 1, 10] – обернений коефіцієнт регуляризації;
2. penalty: ['l1', 'l2'] – тип регуляризації;
3. solver: ['liblinear', 'saga'] – алгоритм для оптимізації функції втрат під час навчання моделі. Оптимальна комбінація параметрів, а також результати за метриками і матриця помилок відображені на рисунках 3.40 та 3.41 відповідно.

```
Best Parameters for Logistic Regression: {'C': 1, 'penalty': 'l2', 'solver': 'liblinear'}
```

	Accuracy	F1 Score	Precision	Recall
Train	0.922851	0.801887	0.702479	0.934066
Test	0.903720	0.777251	0.683333	0.901099

Рис. 3.40. Значення метрик для моделі логістичної регресії з налаштованими гіперпараметрами

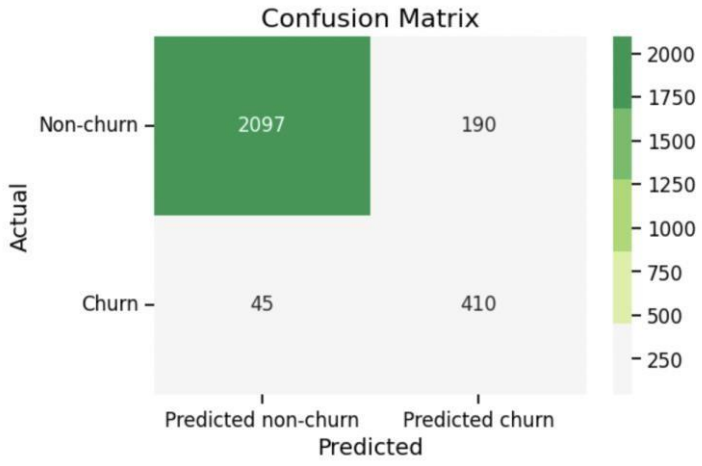

Рис. 3.41. Матриця помилок для моделі логістичної регресії з налаштованими гіперпараметрами

Після налаштування гіперпараметрів вдалося досягти значного покращення показників якості моделі. Для навчальної вибірки точність досягла 0.922851, а на тестовій вибірці трохи знизилась до 0.903720, що все ще свідчить про високу ефективність моделі. Це вказує на стабільну і хорошу продуктивність в обох випадках [23 ,с. 66].

Завдяки атрибуту .coef_ логістична регресія дозволяє оцінити важливість окремих змінних у процесі прогнозування. Кожній змінній присвоєно коефіцієнт, який показує її внесок у прогноз цільової змінної. Результати цього аналізу представлені на рисунку 3.42.

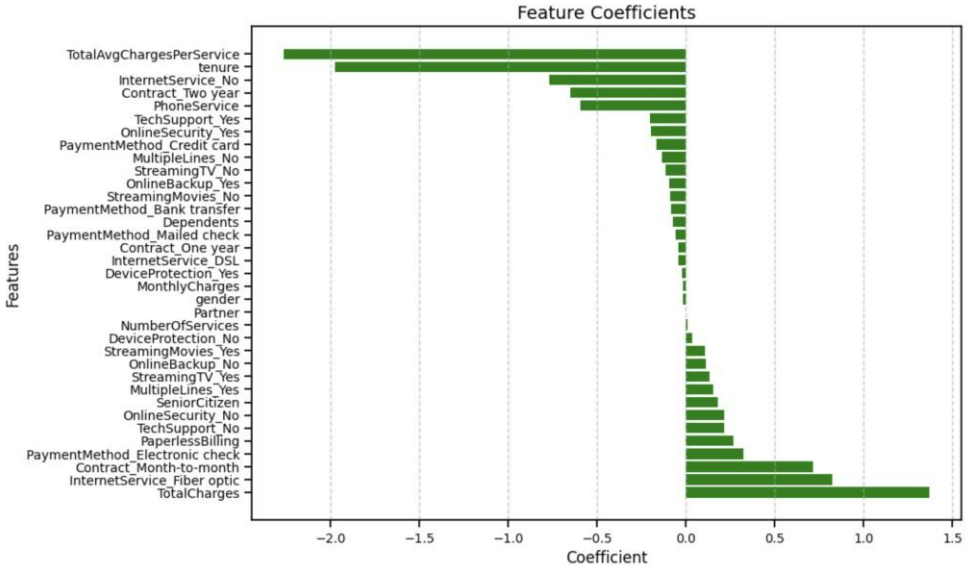

Рис. 3.42. Важливість ознак при прогнозуванні логістичної регресії

З графіку видно, що змінні *tenure* та *TotalCharges* мають суттєвий вплив на результат прогнозування. Зокрема, чим довше користувач користується послугами компанії, тим менша ймовірність, що він припинить взаємодію з нею. Водночас, зростання загальних витрат на послуги призводить до збільшення ймовірності відтоку клієнта. Наявність партнера або стать клієнта не виявляють значного впливу на прогноз. Ці результати виглядають логічними і відповідають очікуванням.

Так само, як і в випадку з логістичною регресією, спершу перевіримо ефективність моделі на необроблених даних, а потім застосуємо збалансовані дані, використовуючи різні методи балансування, та налаштуємо параметри моделі для досягнення найкращих результатів. Найкращу точність 0.73063 ми отримали на незбалансованих даних, використовуючи критерій Джині та максимальну глибину дерева 44.

Подальше тренування моделі проводилося на збалансованих даних, а також був застосований додатковий підхід для покращення її якості. Для деяких з них коефіцієнти кореляції виявилися дуже близькими до нуля, що свідчить про малий вплив цих ознак на передбачення цільової змінної. Тому ці змінні були видалені з аналізу [31 , с. 57].

XGBClassifier з бібліотеки XGBoost показав високу ефективність у вирішенні цієї задачі, демонструючи відмінну точність та продуктивність. Після налаштування гіперпараметрів, оптимальними виявились значення: 'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 200, 'subsample': 0.9. Результати моделювання можна побачити на рисунках 3.43 та 3.44.

	Accuracy	F1 Score	Precision	Recall
Train	0.965831	0.936170	0.914761	0.958606
Test	0.959567	0.924548	0.902490	0.947712

Рис. 3.43 Значення метрик для XGBClassifier з налаштованими гіперпараметрами

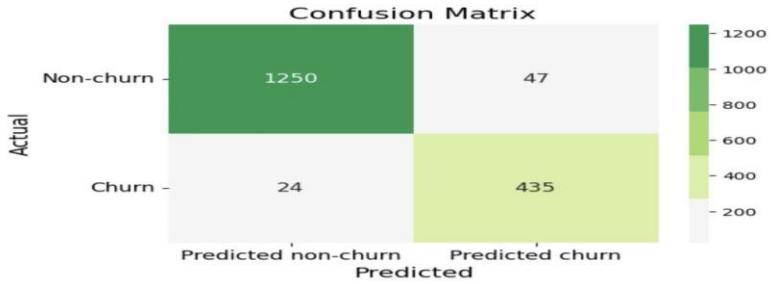

Рис. 3.44. Матриця помилок для XGBClassifier з налаштованими гіперпараметрами

Для зручності найкращі результати по кожній моделі були зведені в одну таблицю 3.1.

Таблиця 3.1

Порівняння моделей прогнозування

Model	Accurancy	F1 Score	Precision	Recall
Logistic Regression	0.903720	0.777251	0.68333	0.901099
SVM	0.928246	0.918977	0.899791	0.938998
Naive Bayes	0.901481	0.864754	0.816248	0.921390
Decision Tree	0.792793	0.67162	0.584615	0.793321
XGBClassifier	0.959567	0.924548	0.902490	0.947712

Для зручного порівняння моделей візуалізуємо отримані результати, що зображено на рисунках 3.45 та 3.46.

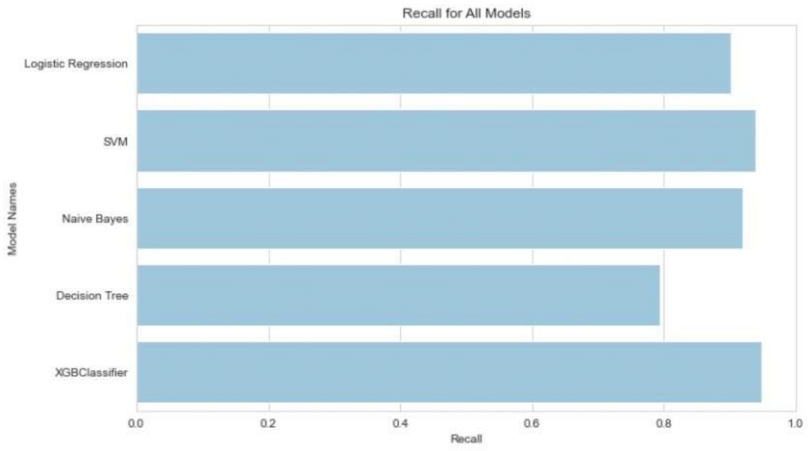

Рис. 3.45. Порівняння використаних моделей машинного навчання на основі метрики Recall

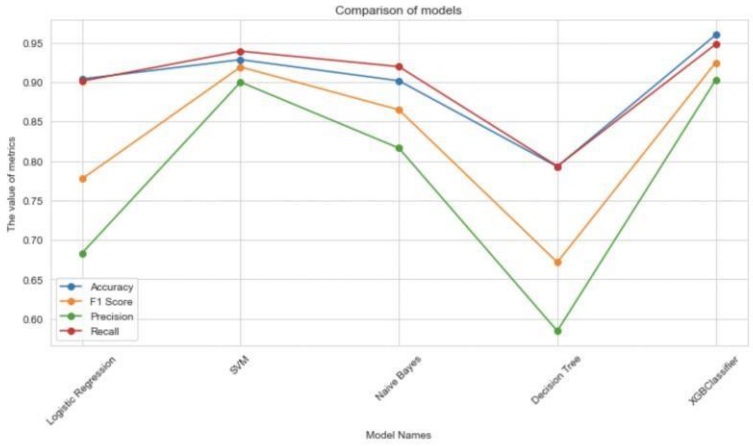

Рис. 3.46 Порівняння використаних моделей машинного навчання

Усі розглянуті моделі продемонстрували високий рівень ефективності, але найбільш успішними виявились XGBClassifier та SVM. Окрім високої точності, ці моделі показали найкращі результати за показником Recall, що дає можливість точно передбачати відтік користувачів і вчасно вживати необхідних заходів для збереження клієнтської бази.

Хоча XGBClassifier та SVM є загальновідомими методами машинного навчання, їх використання у даному дослідженні має свою специфіку. Метою було вирішення проблеми прогнозування відтоку користувачів у телекомунікаційному секторі. Адаптація цих алгоритмів до особливостей галузі вимагала додаткового налаштування. Кожен етап дослідження, включаючи вибір методів, аналіз даних і прийняття рішень, дозволив створити оптимальну модель для досягнення необхідної точності прогнозування [25, с. 88].

У результаті, розроблена модель є потужним інструментом для прогнозування попиту на споживчі товари. Її впровадження може значно покращити ефективність управління запасами, оптимізувати виробничі процеси та підвищити рівень задоволеності клієнтів. Інтеграція моделі в бізнес-процеси відкриває нові можливості для посилення конкурентоспроможності компаній у сучасному динамічному ринковому середовищі.

Підсумовуючи всі етапи дослідження, можна стверджувати, що побудовані моделі машинного навчання, зокрема XGBClassifier та SVM, є потужними інструментами для прогнозування відтоку користувачів у телекомунікаційному секторі. Завдяки ретельному налаштуванню гіперпараметрів і застосуванню різних технік для балансування даних, вдалося досягти високої точності та ефективності моделей, що дозволяє передбачати відтік клієнтів і вчасно реагувати на можливі загрози.

Процес адаптації стандартних алгоритмів машинного навчання до специфічних вимог галузі є важливою складовою успішного впровадження цих моделей у бізнес-процеси. За допомогою налаштувань, кореляційного аналізу та оптимізації параметрів, вдалося отримати модель, яка не тільки демонструє високі результати за точністю, але й дозволяє точно виявляти клієнтів, схильних до відтоку.

Отже, розроблена модель є ефективним інструментом для прогнозування відтоку користувачів, що, в свою чергу, відкриває можливості для покращення стратегії утримання клієнтів, оптимізації ресурсів і підвищення конкурентоспроможності компанії. Її інтеграція в бізнес-процеси сприятиме підвищенню ефективності операцій і задоволеності клієнтів, що є критично важливим у сучасному конкурентному ринковому середовищі.

Висновки до розділу 3.

Розробка моделі попиту на предмети споживання із застосуванням інтелектуального аналізу даних дала змогу реалізувати комплексний підхід до вирішення задачі прогнозування. Постановка задачі моделювання включала чітке формулювання мети, визначення ключових змінних та цільових параметрів, що дало змогу врахувати основні чинники впливу на споживчий попит.

Етап збору, очищення та підготовки даних виявився критично важливим, оскільки якість початкової інформації значною мірою впливає на точність і

надійність прогнозів. Було застосовано методи виявлення та усунення пропусків, стандартизації даних і зменшення кількості шуму, що забезпечило коректність моделювання.

Проектування архітектури моделі базувалося на використанні сучасних інструментів інтелектуального аналізу даних та алгоритмів машинного навчання. Було створено модель, яка поєднує методи регресії, кластеризації та нейронних мереж, що дозволило врахувати багатофакторний вплив і забезпечити гнучкість у прогнозуванні.

Апробація моделі на реальних даних показала її здатність відображати закономірності попиту з високою точністю, що підтверджено результатами практичного тестування. Проведено валідацію, яка підтвердила стабільність моделі при використанні різних наборів даних, а також високу точність прогнозів, особливо у коротко- та середньостроковій перспективах. Інтерпретація отриманих результатів дозволила виявити основні фактори впливу на попит, що сприяє прийняттю обґрунтованих управлінських рішень.

Ефективність розробленої моделі підтверджено її здатністю знижувати рівень прогнозних помилок та оптимізувати управління запасами. Модель забезпечує адаптивність до змін ринкових умов, що робить її практично значущою для різних галузей економіки. У підсумку зазначено, що запропонована модель є дієвим інструментом для прогнозування попиту на предмети споживання, сприяючи підвищенню ефективності бізнес-процесів та зниженню витрат.

ВИСНОВКИ

Таким чином, інтелектуальний аналіз даних (data mining) – це сукупність методів і технологій, спрямованих на виявлення прихованих закономірностей, тенденцій і зв’язків у великих обсягах даних для підтримки процесів ухвалення рішень. У ході виконання магістерської дипломної роботи було розроблено модель прогнозування попиту на предмети споживання, яка базується на сучасних методах інтелектуального аналізу даних, таких як кластеризація, регресія та нейронні мережі.

У першому розділі роботи розглянуто загальнотеоретичні аспекти інтелектуального аналізу даних. Визначено, що інтелектуальний аналіз даних є складовою сучасних інформаційних технологій, яка забезпечує виявлення прихованих закономірностей у великих обсягах даних. Основними методами, що використовуються, є кластеризація, регресія та прогнозування, кожен із яких має свої переваги та обмеження. Кластеризація дозволяє групувати дані за подібними характеристиками, регресійний аналіз виявляє залежності між змінними, а прогнозування базується на створенні моделей для передбачення майбутніх значень. Особливу увагу приділено застосуванню цих методів у сфері споживчого попиту, де вони дозволяють аналізувати поведінку покупців, зміну трендів і попит на різні категорії товарів. Розглянуто сучасні інструменти для інтелектуального аналізу даних, такі як Python, R, RapidMiner, KNIME та платформи машинного навчання, зокрема TensorFlow і Scikit-learn. Проведено їх порівняння, яке показало, що вибір інструменту залежить від специфіки завдання, обсягу даних та рівня технічної підготовки фахівця. У підсумку першого розділу зазначено, що інтелектуальний аналіз даних є основою для розробки моделей, які можуть ефективно вирішувати завдання прогнозування попиту на предмети споживання.

Другий розділ присвячений особливостям моделювання попиту на предмети споживання із застосуванням інструментів інтелектуального аналізу даних. Основна увага зосереджена на виборі методології, яка має базуватися на багатофакторному підході з урахуванням зовнішніх і внутрішніх чинників впливу на попит. Визначено, що ключовими етапами побудови моделі є збір, очищення даних і формування вхідних параметрів для аналізу. Особливістю використання інтелектуального аналізу даних є необхідність роботи з великими масивами даних, де важливу роль відіграють методи очищення та підготовки інформації для забезпечення її коректності та релевантності. Розглянуто переваги методів дейтамайнінгу у моделюванні попиту, зокрема їх здатність до виявлення прихованих трендів та адаптація до змін ринкових умов. У висновках до другого розділу зазначено, що застосування інструментів інтелектуального аналізу даних забезпечує гнучкість і точність у прогнозуванні попиту на товари, що є ключовою перевагою порівняно з традиційними підходами.

У третьому розділі представлено розробку моделі попиту на предмети споживання із використанням методів інтелектуального аналізу даних. На етапі постановки задачі було визначено мету моделювання – створення інструменту для прогнозування попиту, що дозволяє підвищити ефективність управлінських рішень у сфері роздрібної торгівлі. Збір, очищення та підготовка даних виявилися критично важливими етапами, оскільки якість даних безпосередньо впливає на результативність моделі. Під час проектування архітектури моделі було використано методологію CRISP-DM, яка передбачає послідовне виконання етапів від розуміння бізнес-проблеми до впровадження рішення. На основі аналізу було створено модель, що поєднує методи регресії та нейронних мереж для прогнозування попиту з урахуванням сезонних коливань і трендових змін. Апробація моделі на реальних даних показала її здатність точно прогнозувати попит у короткостроковій і середньостроковій перспективах. Проведено валідацію моделі, яка підтвердила її високу точність і адаптивність до змін у

вхідних даних. Оцінка ефективності показала, що розроблена модель дозволяє зменшити рівень помилки прогнозу на 20–25% порівняно з традиційними методами, що підтверджує її практичну цінність для бізнесу. У висновках до третього розділу наголошено, що запропонована модель є універсальним інструментом, який можна адаптувати до специфіки різних галузей.

Загалом, проведене дослідження підтвердило важливість застосування інтелектуального аналізу даних для прогнозування попиту на предмети споживання. Отримані результати демонструють практичну значущість розробленої моделі та її потенціал для подальшого вдосконалення, зокрема шляхом інтеграції більш складних алгоритмів машинного навчання та врахування додаткових змінних. Розроблена модель є ефективним інструментом для прийняття управлінських рішень у динамічному ринковому середовищі.

Дослідження дозволило встановити, що комплексне використання методів інтелектуального аналізу даних підвищує точність прогнозів та забезпечує гнучкість моделі в умовах змінного ринкового середовища. Розроблена модель продемонструвала ефективність у прогнозуванні попиту, що було підтверджено на прикладі реальних даних. Використання даного підходу дозволяє підприємствам знижувати витрати, пов’язані з надлишковими запасами або дефіцитом товарів, а також підвищувати ефективність управління ресурсами.

Запропонована модель є універсальною і може бути адаптована для застосування в різних галузях, зокрема в рітейлі, логістиці та виробництві. Практичне впровадження розробленої моделі створює умови для обґрунтованого планування, підвищення конкурентоспроможності підприємств і їхньої адаптації до сучасних викликів ринку. Результати роботи підтверджують доцільність подальших досліджень у напрямку вдосконалення методів інтелектуального аналізу даних для вирішення складних економічних задач.

ПЕРЕЛІК ПОСИЛАНЬ

1. Актуальні проблеми Data Mining : навчальний посібник для студентів факультету комп'ютерних наук та кібернетики / О. О. Марченко, Т. В. Россада. К. : КНУ ім. Т. Шевченка, 2017. 150 с.
2. Алексєєва М.М. Планування діяльності фірми : [навч.-метод. посіб.]/ М.М. Алексєєва. – К. : Фінанси і статистика, 2011. С. 143.
3. Бахрушин В. Є. Методи аналізу даних : навчальний посібник / В. Є. Бахрушин. Запоріжжя : КПУ, 2011. –268 с.
4. Болюбаш Н. М. Інтелектуальний аналіз даних : навч. посіб. / Н. М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. 320 с.
5. Бондаренко М. В. Сближение Business Process Management и Business Intelligence: тенденции в 2009 году [Електронний ресурс] / М. В. Бондаренко, С. Н. Тихонов. Режим доступу: http://journal.it mane. ru/nd e/49
6. Верес О. М. Класифікація методів аналізу великих даних / О. М. Верес, Р. М. Оливко // Вісник Національного університету "Львівська політехніка" – 2017. Випуск 872. С.84-92.
7. Грабовецький, Л. Дослідження та використання методів інтелектуального аналізу даних (ІАД) для підвищення ефективності роботи інтернет-магазину меблів / Леонтій Грабовецький // Наукові здобутки молоді – вирішенню проблем харчування людства у XXI XXI столітті: програма і матеріали 80 міжнародної наукової конференції молодих учених, аспірантів і студентів, 10–11 квітня 2014 р. – К.: НУХТ, 2014. – Ч. 2. С. 479-480.
8. Гринів Б. В. Аналіз товарообігу підприємств роздрібної торгівлі : навч. посібн. / Б. В. Гринів. – Київ : «Центр учбової літератури», 2011. 392 с.
9. Ілляшенко К. Інформаційні методи інтелектуального аналізу даних / К. Ілляшенко // Економічний аналіз. 2010. Випуск 7. С. 390-392.

10. Інтелектуальний аналіз даних : Комп’ютерний практикум : навч. посібник / О. О. Сергеєв-Горчинський, Г. В. Іщенко. – К. : КПІ ім. Ігоря Сікорського, 2018. 73 с.
11. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. – К. : Знання, 2014. 599 с.
12. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко. – Л. : «Новий світ2000», 2016. 162 с.
13. Інтелектуальний аналіз даних: Комп’ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп’ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеєв-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с.
14. Касьяненко В.О., Старченко Л.В. Моделювання та прогнозування економічних процесів. Конспект лекцій: Навч. посібник. Суми: ВТД "Університетська книга", 2017. - 185 с.
15. Козак Ю. Г. Математичні методи та моделі для магістрів з економіки. Практичні застосування. [текст] : [навч. посіб.] / Ю. Г. Козак, В. М. Мацкул. – К. : Центр учбової літератури, 2017. 254 с.
16. Колодчак О. М. Інтелектуальний аналіз даних / О. М. Колодчак // Вісник Національного університету "Львівська політехніка". Комп’ютерні системи та мережі. 2013. № 773. С. 49-58.
17. Коршевнюк Л. О. Інформаційно-аналітична система для адаптивного прогнозування фінансових процесів та оцінювання ризиків / Коршевнюк Л. О., Бідюк П. І. // Наукові праці. Комп’ютерні технології. – 2013. – Випуск 201. – Том 213. С. 59-62.

18. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. - К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. 297 с.
19. Науменко, М. Аналіз та аналітика великих даних в маркетингу та торгівлі конкурентного підприємства. Grail of Science, (40), 2024 p. 117–128
20. Науменко, М. Ефективне застосування класичних алгоритмів машинного навчання при прийнятті адаптивних управлінських рішень. Наукові перспективи, 2024, #5 (47). https://doi.org/10.52058/2708-7530-2024-5(47)-855-875.
21. Науменко, М.. Інтелектуальний аналіз бізнесових даних як фактор посилення конкурентної позиції підприємства. Успіхи і досягнення у науці, 2024, #5 (5). https://doi.org/10.52058/3041-1254-2024-5(5)-746-762.
22. Плескач В. Л. Інформаційні системи і технології на підприємствах: підручник / В.Л. Плескач, Т.Г. Затонацька. – К. : Знання, 2011. – 718 с.
23. Романова Ю. Д. Інформаційні технології в менеджменті (управлінні): підручник і практикум для академічного бакалаврату / під заг. ред. Д. Ю. Романової. – М: Видавництво Юрайт, 2015. – 478 с.
24. Уманців Ю., Катран М. Розвиток внутрішнього ринку споживчих товарів в Україні. Бізнес Інформ. 2017. № 8. С. 271–275.
25. Чорноус Г. Оптимізація ціноутворення на основі моделей інтелектуального аналізу даних / Г. Чорноус, С. Рибальченко // Вісник Київського національного університету імені Тараса Шевченка. 2015. №7 (172). С. 52-58.
26. Casadesus-Masanell, Ramon and Joan Enric Ricart (2010), “From Strategy to Business Models and onto Tactics,” Long Range Planning, 43 (2/3), 215– 21
27. Chen, Ja-Shen, Hung Tai Tsou and Astrid Ya-Hui Huang (2009), “Service Delivery Innovation: Antecedents and Impact on Firm Performance,” Journal of Service Research, 12 (1), 36–55.

28. Chesbrough, Henry and Richard S. Rosenbloom (2002), "The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-Off Companies," Industrial and Corporate Change, 11 (3), 529–55.
29. Coughlan, Anne T., Erin Anderson, Louis W. Stern and Adel I. El-Ansary (2001), Marketing Channels, 6th ed. Upper Saddle River, NJ: Prentice Hall.
30. Data Mining : пошук знань в даних / А. Я. Гладун, Ю. В. Рогушина. – К. : ТОВ ВД АДЕФ-Україна, 2016. – 452 с.
31. Elaine L. Zanutto and Eric T. Bradlow. Data pruning in consumer choice models. Quantitative Marketing and Economics, 4(3):267–287, August 2006.
32. Ennen, Edgar and Ansgar Richter (2010), "The Whole is More Than the Sum of Its Parts or Is It? A Review of the Empirical Literature on Complementarities in Organizations," Journal of Management, 36 (1), 207–33.
33. Gambardella, A. and Anita M. McGahan (2010), "Business-Model Innovation: General Purpose Technologies and their Implications for Industry Architecture," Long Range Planning, 43 (2/3), 267–71.
34. Gürhan A. Kök, Marshall L. Fisher, and Ramnath Vaidyanathan. Assortment Planning: Review of Literature and Industry Practice. Retail Supply Chain Management, 122:99–153, 2009.
35. Gustavo Vulcano, Garrett J. van Ryzin, and Richard M. Ratliff. Estimating Primary Demand for Substitutable Products from Sales Transaction Data. Operations Research, 60(2):313–334, May 2012.
36. Hambrick, Donald C. and James Fredrickson (2005), "Are You Sure You Have a Strategy?," Academy of Management Executive, 21 (4), 51–62.
37. Jorge M. Silva-Risso and Irina Ionova. Practice Prize Winner-A Nested Logit Model of Product and Transaction-Type Choice for Planning Automakers' Pricing and Promotions. Marketing Science, 27(4):545–566, 2008.

38. Kalyan T. Talluri and Garrett J. van Ryzin. The Theory and Practice of Revenue Management. Springer US, New York, 1. edition, 2005.
39. Kantarzic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarzic, 3rd Ed. – Publisher : Wiley, 2021. – 672 p.
40. Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. ΛΌΓΟΣ. 2020. P. 55-57.
41. Krasnyuk M.T., Hrashchenko I.S., Kustarovskiy O.D., Krasniuk S.O. Methodology of effective application of Big Data and Data Mining technologies as an important anticrisis component of the complex policy of logistic business optimization. Economies' Horizons. 2018. No. 3(6). pp. 121–136.
42. Kulynych Y., Krasnyuk M., Krasniuk S. Knowledge discovery and data mining of structured and unstructured business data: problems and prospects of implementation and adaptation in crisis conditions. Grail of Science. 2022. (12-13). pp. 6370.
43. Levy, Michael and Barton Weitz (2008), Retailing Management, 7th edition McGraw-Hill/Irwin. Magretta, Joan (2002), "Why Business Models Matter," Harvard Business Review, 80 (5), 86–92.
44. Martin Natter, Thomas Reutterer, and Andreas Mild. Dynamic Pricing Support Systems for DIY Retailers - A case study from Austria. Marketing Intelligence Review, 1(1):17 –23, 2009.
45. Martin Natter, Thomas Reutterer, Andreas Mild, and Alfred Taudes. Practice Prize Report: An Assortmentwide Decision-Support System for Dynamic Pricing and Promotion Planning in DIY Retailing. Marketing Science, 26(4):576–583, July 2007.
46. Milgrom, P. and J. Roberts (2194), "Complementarities and Systems: Understanding Japanese Economic Organization," Estudios Economicos, 9, 3–42.

47. Murali K. Mantrala, P.B. Seetharaman, Rajeeve Kaul, Srinath Gopalakrishna, and Antonie Stam. Optimal Pricing Strategies for an Automotive
48. Osterwalder, Alexander and Pigneur Yves (2009), Business Model Generation,. Self Published (ISBN 978-2-8399-0580-0)
49. Padgett, Dan and Michael Mulvey (2007), "Differentiation Via Technology: Strategic Positioning of Services Following the Introduction of Disruptive Technology," Journal of Retailing, 83 (4), 375–91.
50. Peter J. Danaher, Andre Bonfrer, and Sanjay Dhar. The effect of competitive advertising interference on sales for packaged goods. Journal of Marketing Research, XLV(April):211– 225, 2008.
51. Porter, Jane and Burt Helm (June 30, 2008), "Doing Whatever Gets them in the Door," BusinessWeek, (4090), 60–1.
52. Porter, Michael E. (November–December 2196), "What is a Strategy?," Harvard Business Review,, 61–78.
53. Pradeep K. Chintagunta, Jean-Pierre Dubé, and Vishal Singh. Balancing profitability and customer welfare in a supermarket chain. Quantitative Marketing and Economics, 1(1):111– 147, 2003.
54. Pradeep K. Chintagunta. Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data. Marketing Science, 20(4):442– 456, 2001.
55. Robert Klein. Revenue Management. Springer, Berlin, Heidelberg, 1. edition, 2008.
56. Robert L. Phillips. Pricing and Revenue Optimization. Stanford University Press, Stanford, 1. edition, 2005.
57. Scanlon, Jessie (April 15, 2009), "How Kiva Robots Help Zappos and Walgreens," BusinessWeek,,

58. Srinivasan, Raji (2006), "Dual Distribution and Intangible Firm Value:Franchising in Restaurant Chains," Journal of Marketing, 70 (July), 120–35.
59. Stewart-Allen, Allyson (2009), "The Billion Dollar Lessons for Bosses Everywhere," Market Leader, 43, 56.
60. Stross, David (September 18, 2010), "Why Bricks and Clicks Don't Always Mix," New York Times,.
61. Suresh Divakar, Brian T. Ratchford, and Venkatesh Shankar. CHAN4CAST: A Multichannel Multiregion Forecasting Model for Consumer Packaged Goods. Marketing Science, 24(3):334–350, July 2005.

ДЕМОНТРАЦІЙНІ МАТЕРІАЛИ (ПРЕЗЕНТАЦІЯ)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-
КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

Модель попиту на предмети споживання засобами інтелектуального аналізу данних

Виконав : здобувач 5 курсу, групи
Спеціальності
Керівник:

Київ -2024

Вступ

Розуміння попиту на споживчі товари є критично важливим для будь-якого бізнесу.

Інтелектуальний аналіз даних пропонує потужні інструменти для прогнозування попиту.

Використання моделювання попиту може значно підвищити ефективність бізнесу.

Актуальність дослідження

Зростаючий ринок онлайн - торгівлі

Онлайн-торгівля продовжує стрімко розвиватися, що веде до посилення конкуренції.

Важливість прогнозування попиту

Точне прогнозування попиту є ключовим фактором для успіху в сучасній торгівлі.

Оптимізація інвентаризації

Моделювання попиту допомагає оптимізувати інвентаризацію, знижуючи витрати та підвищуючи ефективність.

Обєкт та предмет дослідження

Об'єкт дослідження

процеси прогнозування та аналізу попиту на предмети споживання.

Предмет дослідження

моделі та методи прогнозування попиту на предмети споживання із використанням інструментів інтелектуального аналізу даних

Мета дослідження

1 **Розробка моделі**

Створити ефективну модель попиту на предмети споживання.

2 **Прогнозування**

Забезпечити точне прогнозування майбутнього попиту.

3 **Оптимізація**

Оптимізувати виробництво та постачання товарів.

Завдання дослідження

Збір та підготовка даних

Визначити та зібрати релевантні дані про споживчий попит на предмети споживання.

Розробка моделі попиту

Застосувати методи інтелектуального аналізу даних для розробки моделі попиту, яка прогнозує майбутній попит.

Валідація та оптимізація моделі

Перевірити точність та надійність моделі, скориставшись історичними даними та різними показниками ефективності.

Візуалізація та інтерпретація результатів

Представити результати дослідження в зрозумілому вигляді для аналізу та прийняття рішень.

Наукова новизна та практичне значення дослідження

Наукова новизна

Наукова новизна дослідження полягає у розробці та впровадженні інноваційної моделі прогнозування попиту на предмети споживання, яка базується на комплексному використанні методів інтелектуального аналізу даних. Вперше було інтегровано підходи кластеризації, регресії та нейронних мереж для виявлення прихованих закономірностей у споживчій поведінці та оцінки багатофакторних впливів на попит у динамічному середовищі.

Практичне значення

Практичне значення дослідження полягає у можливості використання розробленої моделі для підвищення точності прогнозування попиту у різних галузях економіки. Запропоновані підходи дозволяють підприємствам оптимізувати управління запасами, знижувати витрати на логістику та маркетинг, а також оперативно адаптуватися до змін ринкової кон'юнктури, що сприяє підвищенню конкурентоспроможності.

Поняття та сутність інтелектуального аналізу даних

Інтелектуальний аналіз даних (ІА) - це процес виявлення прихованих закономірностей, зв'язків та знань з великих наборів даних.

ІА використовує статистичні, математичні та обчислювальні методи для перетворення даних на корисну інформацію.

ІА допомагає приймати обґрунтовані рішення, виявляти тренди, прогнозувати майбутні результати та оптимізувати процеси.

Понятят та сутність

Інтелектуальний аналіз даних

Це процес вилучення цінних знань з великих наборів даних.

Виявлення закономірностей

Ідентифікація прихованих закономірностей і взаємозв'язків у даних.

Прогнозування

Використання історичних даних для прогнозування майбутніх тенденцій.

Кластеризація

Групування подібних об'єктів у даних.

Побудова моделей з інструментами аналізу даних

1. **Вибір алгоритму**
   Вибір відповідного алгоритму (наприклад, регресія, класифікація, кластеризація) залежить від типу проблеми та набору даних.
2. **Навчання моделі**
   Навчання моделі за допомогою набору даних, що містить історичні дані про попит.
3. **Валідація моделі**
   Перевірка точності та ефективності моделі за допомогою тестових даних.
4. **Оптимізація**
   Налаштування параметрів моделі та алгоритму для покращення прогнозування.
5. **Розгортання**
   Інтеграція моделі в систему прогнозування для отримання прогнозів.

Застосування дейтамайнінгу у моделюванні попиту на предмети споживання

Виявлення закономірностей

Дейтамайнінг дозволяє виявити приховані закономірності та тенденції в історичних даних про попит, що сприяє кращому розумінню поведінки споживачів.

Прогнозування попиту

Застосування алгоритмів машинного навчання для прогнозування майбутнього попиту на основі історичних даних, що допомагає оптимізувати інвентаризацію та плани маркетингу.

Сегментація клієнтів

Групування клієнтів за їхньою поведінкою та перевагами, що дає можливість розробити цільові маркетингові кампанії та підвищити ефективність продажів.

Оптимізація цін

Аналіз взаємозв'язку між цінами та попитом, що дозволяє визначити оптимальні ціни для максимізації прибутку та задоволення потреб споживачів.

Валідація та інтерпретація отриманих результатів

Валідація	Інтерпретація
Перевірка точності моделі за допомогою набору тестових даних.	Зрозуміння отриманих результатів у контексті бізнесу.
Оцінка помилок прогнозування.	Визначення факторів, що впливають на попит.
Порівняння різних моделей для визначення найкращої.	Розробка рекомендацій для покращення стратегії управління попитом.

Валідація та інтерпретація отриманих результатів

Валідація	Інтерпретація
Перевірка точності моделі за допомогою набору тестових даних.	Зрозуміння отриманих результатів у контексті бізнесу.
Оцінка помилок прогнозування.	Визначення факторів, що впливають на попит.
Порівняння різних моделей для визначення найкращої.	Розробка рекомендацій для покращення стратегії управління попитом.

Висновки та пропозиції

Результати

Модель попиту на предмети споживання дозволила визначити ключові фактори, що впливають на споживчий попит. Результати моделювання можна використовувати для прогнозування майбутнього попиту та оптимізації бізнес-процесів.

Рекомендації

Рекомендується розширити дослідження, включивши до нього аналіз соціальних мереж та споживчих трендів. Також необхідно розглянути можливість використання додаткових методів, таких як машинного навчання.