

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Системотехніки
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістрський)

Дослідження застосування методів інтелектуального аналізу
даних в інформаційній системі готелю

(тема)

Виконав:

студент 2 курсу, групи ІТПм-22-2

Ярмак О.М.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформаційні технології
проекування

(повна назва освітньої програми)

Керівник к.т.н., доцент Тітов С.В.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

Гребеннік І.В.

(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
Кафедра _____ Системотехніки _____
Рівень вищої освіти _____ другий (магістрський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-професійна _____
Освітня програма _____ Інформаційні технології проектування _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
« ____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Ярмаку Олександр Михайловичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження застосування методів інтелектуального аналізу даних в інформаційній системі готелю
затверджена наказом університету від «20» 11 2023 р. № 1373 Ст
2. Термін подання студентом роботи до екзаменаційної комісії 21.01.2024 р.
3. Вихідні дані до роботи Науково-технічні публікації та дані Інтернет-джерел щодо тематики кваліфікаційної роботи; аналіз застосування інтелектуального аналізу даних в бізнес-сфері; аналіз застосування Data Mining в бізнес-сфері, аналіз застосування інтелектуального аналізу даних в сфері готельного бізнесу; аналіз застосування Data Mining в сфері готельного бізнесу; застосування інструментів Azure Machine Learning для аналізу попиту готелю.
4. Перелік питань, що потрібно опрацювати в роботі Вступ, аналіз предметної області, аналіз сучасного процесу використання методів дослідження інтелектуального аналізу даних, аналіз застосування досліджуваних технологій в існуючих системах, постановка задачі на дослідження, аналіз застосування методів аналізу даних в бізнес-сфері, аналіз шляхів збільшення попиту готелю за допомогою використання інструментів інтелектуального аналізу даних, аналіз та опис технології Data Mining, дослідження методів вирішення задачі, загальні кроки використання Data Mining, шляхи використання Data Mining для аналізу попиту готелю, загальні кроки використання Azure Machine Learning, шляхи використання Azure Machine Learning для аналізу попиту, опис аналізу попиту готелю за допомогою Azure Machine Learning.
5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Традиційна система ВІ, набори даних, логічний процес побудови алгоритму машинного навчання, процес завантаження файлу з вибіркою, створена робоча область, процес візуалізації

компонента, процес видалення нормалізованих залишків, процес відкидання неповних рядків з вибірки, перевірка характеристик, навчання моделі лінійної регресії, візуалізація результатів Evaluate model, схема порівняння навчання обох моделей, результат порівняння навчання обох моделей, підрахована модель у вигляді графіку, технічні дані аналізу, результати проведеного аналізу.

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	<i>Отримання завдання на виконання роботи</i>	<i>21.09.2023</i>	<i>виконано</i>
2	<i>Дослідження існуючих методів аналізу даних</i>	<i>22.09-30.09.2023</i>	<i>виконано</i>
3	<i>Огляд літератури</i>	<i>30.09-15.10.2023</i>	<i>виконано</i>
4	<i>Розробка плану вирішення поставленої задачі</i>	<i>15.10-14.11.2023</i>	<i>виконано</i>
5	<i>Розробка прототипу інформаційної системи</i>	<i>14.11-30.11.2023</i>	<i>виконано</i>
6	<i>Розробка методів аналізу даних</i>	<i>01.12-20.12.2023</i>	<i>виконано</i>
7	<i>Тестування та налагодження розроблених методів</i>	<i>21.12-05.01.2024</i>	<i>виконано</i>
8	<i>Оформлення пояснювальної записки</i>	<i>06.01-21.01.2024</i>	<i>виконано</i>
9	<i>Представлення на рецензування</i>	<i>22.01.2024</i>	<i>виконано</i>

Дата видачі завдання 21 вересня 20 23 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

к.т.н. доцент Тітов С.В.
(посада, прізвище, ініціали)

РЕФЕРАТ

Кваліфікаційна робота: 91 стор., 17 рис., 14 джерел інформації.

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, БІЗНЕС-АНАЛІЗ, DATA MINING, MICROSOFT AZURE, AZURE MACHINE LEARNING, ДІЯЛЬНІСТЬ ГОТЕЛЮ, ІНФОРМАЦІЙНА СИСТЕМА, ПОПИТ ГОТЕЛЮ.

Об'єктом дослідження є аналіз і застосування методів інтелектуального аналізу даних в інформаційній системі готелю.

Предметом дослідження є аналіз існуючих варіантів покращення розробленої інформаційної системи для підтримки діяльності готелю на прикладі аналізу попиту серед клієнтів.

Мета роботи: використовуючи теоретичні та практичні навички, провести дослідження і аналіз попиту серед клієнтів інформаційної системи готелю.

Методи дослідження – класифікація та прогнозування, кластерний аналіз за допомогою Azure Machine Learning.

Під час виконання роботи був проведений аналіз існуючих варіантів покращення існуючої інформаційної системи, що належить до діяльності готелю на прикладі аналізу попиту готелю серед клієнтів.

ABSTRACT

Qualification work: 91 pages, 17 figures, 14 sources of information.

INTELLIGENT DATA ANALYSIS, BUSINESS ANALYSIS, DATA MINING, MICROSOFT AZURE, AZURE MACHINE LEARNING, HOTEL OPERATIONS, INFORMATION SYSTEM, HOTEL DEMAND

The object of research is the analysis and application of methods of intellectual data analysis in the hotel information system.

The subject of the study is the analysis of existing options for improving the developed information system to support the hotel's activities, based on the example of the analysis of demand among clients.

The purpose of the work: using theoretical and practical skills, conduct research and analysis of demand among clients of the hotel's information system.

Research methods – classification and prediction, cluster analysis using Azure Machine Learning.

During the execution of the work, an analysis of existing options for improving the existing information system related to the hotel's activities was carried out on the example of an analysis of the hotel's demand among customers.

ЗМІСТ

Скорочення та умовні позначки	7
Вступ.....	8
1 Аналіз предметної області.....	10
1.1 Аналіз сучасного процесу використання методів дослідження інтелектуального аналізу даних.....	10
1.2 Аналіз застосування досліджуваних технологій в існуючих системах	17
1.2.1 Опис існуючих напрямків використання ІАД.....	21
2 ПОСТАНОВКА ЗАДАЧІ НА ДОСЛІДЖЕННЯ.....	26
2.1 Аналіз застосування методів інтелектуального аналізу даних в бізнес-сфері	26
2.2 Аналіз шляхів збільшення попиту готелю за допомогою використання інструментів інтелектуального аналізу даних	32
2.3 Аналіз та опис технології Data mining.....	34
3 Дослідження методів вирішення задачі.....	44
3.1 Загальні кроки використання Data Mining.....	44
3.1.1 Шляхи використання Data Mining для аналізу попиту готелю	46
3.2 Загальні кроки використання Azure Machine Learning.....	50
3.2.1 Шляхи використання Azure Machine Learning для аналізу попиту	52
3.3 Опис аналізу попиту готелю за допомогою Azure Machine Learning	55
Висновок	70
Перелік джерел посилання	71
Додаток А Графічні матеріали кваліфікаційної роботи.....	73
Додаток Б «Відомість кваліфікаційної роботи».....	91

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ

СУБД – система управління базами даних.

ІАД – Інтелектуальний аналіз даних.

БД – База даних.

OLAP – Online analytical processing.

CBR – Case based reasoning.

DM – Data mining.

DAX – Data Analysis Expressions

SQL – Structured Query Language (мова структурованих запитів).

ML – Machine Learning

BI – Business Intelligence

ETL – Extract, Transform, Load

MSAS – Multi-functional Satellite Augmentation System

SSAS – SQL Server Analysis Services

ВСТУП

В сфері сучасного готельного бізнесу, в якому конкуренція стає усе більш жорсткою, використання інтелектуального аналізу даних стає одним з найбільш необхідних елементів поліпшення якості обслуговування гостей та оптимізації управління готелем. На сьогоднішній день, інформаційні системи в готельній галузі є об'єктом посиленого дослідження для впровадження передових технологій і аналітичних методів, що мають допомагати вирішенню важливих завдань в готельній сфері.

Головною метою готельної інформаційної системи є підтримка діяльності готелю, що має в собі не тільки систему для обліку бронювань і заселень, а й додаткових послуг, що надаються, для покращення комфорту і якості проживання клієнтів.

Проте, з розвитком діяльності готелю обсяг даних неухильно зростає. Саме тому існує попит на розробку і впровадження нових методів аналізу для винесення корисної інформації та виявлення прихованих патернів для подальшого розвитку.

Мета дослідження – спрямування на розвинення і впровадження інтелектуального аналізу даних у інформаційну систему готелю, метою якого є покращення його працездатності як інформаційної системи, так і як бізнес-проекту.

Одним з найважливіх факторів для існування готелю як прибуткового бізнес-проекту є попит серед клієнтів. Даний фактор буде розглядатися більш детально як приклад аналізу працездатності за допомогою дослідження методів інтелектуального аналізу даних в інформаційній системі готелю.

Попит – економічний чинник, що вказує на кількість товарів або послуг, які бажають придбати на ринку споживачі при певному рівні цін і в певний період часу. Попит є однією з основних детермінант економічного підприємництва та взаємодії між покупцями і продавцями.

Основними факторами попиту є умови ринку, ціна, очікування споживачів та доходи споживачів. Усі ці складові є важливими для формування високого рівня попиту.

Попит є ключовим фактором в рішеннях стосовно виробництва, розподілі ресурсів в економіці та формуванні цін. Існує можливість використання інтелектуального аналізу даних не тільки для вивчення, передбачення попиту, а й для оптимізації планування діяльності готелю як бізнес-проекту.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз сучасного процесу використання методів дослідження інтелектуального аналізу даних

В зв'язку з покращенням технологій зберігання і запису даних світ отримав величезний об'єм даних та інформації в різноманітних областях. На сьогоднішній день сфера роботи будь-якого підприємства супроводжується фіксуванням та записом всіх кроків під час його діяльності. Очевидним є факт, що за відсутністю продуктивного аналізу та переробки інформації, такі об'єми «сирих» даних не є необхідними. Сучасним підходом до подібного аналізу та переробки є наступні кроки:

- об'єм даних може бути необмеженим;
- дані можуть бути представлені в різних форматах, таких як якісний, текстовий, кількісний;
- результати аналізу мають бути зрозумілими і конкретними;
- інструментами обробки «сирих» даних мають бути простими в використанні.

Стандартні алгоритми математичної статистики довгий час були основними і підтимували концепцію усереднення із вибірки, яка зводиться до операцій з фіктивними величинами. Виявилось, що методи математичної статистики є корисними переважно для перевірки гіпотез, які були сформульовані заздалегідь і для більш абстрактного аналізу, який є базисом оперативної аналітичної обробки даних.

У основі інтелектуального аналізу даних закладений концепт паттернів, які відображають елементи багатоаспектних відносин в даних. Дані паттерни є закономірними. Вони ж мають властивість підвиборок даних, що мають можливість стислої вираженості в зрозумілій для людини формі. Пошук паттернів, що використовує методи, які не є заздалегідь обмеженими припущеннями щодо структури вибірки і типами розподілів значень показників,

щоб були проаналізовані. Одним з найважливіших положень інтелектуального аналізу даних є нетривіальність паттернів, що шукаються. З цього витікає те, що ці паттерни мають відображати несподівану або неочевидну постійність у даних, які становлять, так звані, приховані знання.

Таким чином, інтелектуальний аналіз даних є процесом знаходження у «сирих» даних невідомих раніше і доступних та корисних інтерпретацій інформації, яка необхідна для прийняття рішень в різноманітних сферах діяльності людини.

Розвинення методик запису та збереження даних призвело до нестримного росту об'ємів зібраної і проаналізованої інформації. Об'єми даних великі настільки, що людина не має можливості провести їх аналіз самотійно. В той же час, необхідність проведення подібного аналізу досить очевидна, оскільки такі «сирі» дані мають інформацію, яка може використовуватися для прийняття рішень в майбутньому [1]. Стандартна математична статистика, як було сказано раніше, довгий час була претендетом до ролі головного інструменту аналізу даних, не була спроможна вирішувати ті проблеми, що виникали в процесі аналізу. Саме тому з'явилася необхідність в розвиненні нових сучасних методологій обробки і аналізу інформації. Місце цієї методології став інтелектуальний аналіз даних.

Виділяють наступні причини популярності використання ІАД:

- стрімке накопичення даних у великих об'ємах;
- загальносвітовий тренд на комп'ютеризацію бізнес-процесів;
- проникнення мережі Інтернет у всі сфери діяльності;
- прогрес у сфері інформаційних технологій.

Використовувані алгоритми у ІАД вимагають досить великої кількості обчислень. Раніше це був найголовнішим фактором стримування широкого застосування ІАД на практиці, але сьогоднішній зріст потужностей сучасних процесорів зняв цю проблему з порядку денного майже повністю. Тепер існує можливість проведення якісного аналізу мільйонів записів за відносно невеликий проміжок часу.

Під час розвитку ІАД були використані різноманітні галузі, такі як розпізнавання образів, прикладна статистика, теорія баз даних, штучний інтелект, тощо [2].

Алгоритмами та методами ІАД є наступні складові:

- лінійна регресія;
- методи найближчого сусіда та к-найближчого сусіда;
- методи пошуку асоціативних правил;
- методи візуалізації даних;
- штучні нейронні мережі;
- ієрархічні методи кластерного аналізу;
- неієрархічні методи кластерного аналізу;
- метод опорних векторів;
- дерева рішень;
- генетичні алгоритми;
- кореляційно-регресійний аналіз;
- символні правила;
- байєсові мережі;
- метод обмеженого перебору;
- безліч інших методів.

Необхідно підкреслити, що більша частина методів ІАД розроблялася в межах теорії штучного інтелекту. Єдиної думки щодо виду задач, які можна зарахувати до ІАД не існує. Більша частина авторитетних джерел перелічує наступні:

- асоціація;
- класифікація;
- аналіз виявлення відхилень;
- кластеризація;
- оцінювання;
- підведення підсумків;
- прогнозування;

- аналіз зв'язків;
- візуалізація.

Розглянемо деякі з них [3].

Класифікація є однією з найпростіших і найпоширеніших задач ІАД. У результаті вирішення задачі класифікації виявляються класи – ознаки, характеризуючі групи об'єктів набору даних, що досліджується. За допомогою цих ознак існує можливість зарахувати новий об'єкт до того чи іншого класу. Для рішення задачі класифікації є можливість користуватися наступними методами:

- метод байєсової мережі;
- метод найближчого сусіда;
- метод к-ближнього сусіда;
- нейронних мереж;
- індукції дерев рішень.

Кластеризація є логічним продовженням ідеї класифікації. Її особливістю є те, що класи об'єктів не визначені з самого початку. В результаті кластеризації відбувається розбиття об'єктів на групи. Одним з прикладів методу задачі кластеризації може бути особливий тип нейронних мереж під назвою карта Кохонена, що самоорганізуються без вчителя.

Асоціація. Під час вирішення задачі пошуку асоціативних правил шукаються закономірності подіями в наборі даних, пов'язаних між собою. Відмінностями асоціації від класифікації та кластеризації є пошук закономірностей, що проходить не на основі властивостей об'єкта, який аналізується, а між кількома подіями, що відбуваються одночасно. Одним з найвідоміших алгоритмів розв'язання задачі пошуку асоціативних правил є алгоритм Apriori.

Послідовність створює можливість пошуку тимчасових закономірностей між транзакціями. Основне завдання послідовності схоже на завдання асоціації, але з різницею, що її мета – встановлення закономірностей не з одночасно наступаючими подіями, а між тими подіями, що пов'язані у часі, тобто, які проходять з певним інтервалом в часі. Дану задачу також називають задачею

знаходження послідовних шаблонів. Основним правилом послідовності є формула: після події X через деякий час відбудеться подія Y.

Прогнозування. Результатом вирішення задачі прогнозування на основі особливостей вже існуючих даних є пропущені або майбутні значення цільових числових показників. Для вирішення такого роду задач широко використовуються нейронні мережі, методи математичної статистики тощо.

Візуалізація. Результатом методу візуалізації є створення графічного образу аналізованих даних. Для вирішення задачі візуалізації використовуються графічні методи, які демонструють наявність закономірностей в даних. Прикладом методу візуалізації може бути представлення даних у 2D- і 3D-вимірваннях.

Підведення підсумків. Дана задача, метою якої є опис конкретних груп об'єктів із аналізованого набору даних [4].

Задачі ІАД, залежно від моделей, які використовуються, можуть бути або прогнозуючими, або дескриптивними. Результатом вирішення описових задач, аналітик отримує паттерни, які описують дані, що піддаються інтерпретації. Дані задачі описують загальну концепцію аналізованих даних, визначають їх підсумкові, відмінні, інформативні особливості.

ІАД може складатися з двох або трьох стадій [4]:

- виявлення закономірностей, або вільний пошук;
- використання виявлених закономірностей для прогнозу невідомих значень, або прогностичне моделювання; додатково до даних стадій інколи вводиться стадія валідації – наступна за стадією вільного пошуку [5]. Головною метою валідації є перевірка достовірності знайдених закономірностей. Але, вважається, що валідація частіше за все – це частина першої стадії, бо у реалізації багатьох методів, таких як дерев рішень і нейронних мереж, передбачений розподіл загальної множини даних на навчальні та перевіркові, і останні дозволяють перевіряти достовірність отриманих результатів.
- аналіз виключень; стадія, що призначена для виявлення і пояснення аномалій, що були знайдені в закономірностях.

Вільний пошук. На даній стадії відбувається дослідження набору даних, метою якого є пошук прихованих закономірностей. Попередні гіпотези про вид закономірностей на даній стадії не визначаються. Закономірністю є постійно повторюваний взаємозв'язок, який визначає етапи та форми процесу становлення і розвитку різних процесів або явищ. Система ІАД на даній стадії визначає шаблони, для отримання яких, наприклад, в системах OLAP, аналітику потрібно придумувати і створювати множину запитів. На даному етапі аналітик звільняється від подібної роботи, бо шаблони шукає за нього сама система. Особливо корисним застосування даного підходу в надвеликих базах даних, в яких виявити закономірність за допомогою створення запитів досить складно. Для такого типу задач вимагається випробувати безліч різних варіантів. Вільний пошук подано наступними діями [5]:

- виявлення закономірностей умовної логіки;
- виявлення закономірностей асоціативної логіки.

Описані дії в межах стадії вільного пошуку виконуються за допомогою наступних дій:

- індукції правил умовної логіки. До цього пункту входять задачі класифікації і кластеризації, опис в компактній формі близьких або подібних груп або об'єктів;
- індукції правил асоціативної логіки. До цього пункту входять задачі асоціації та послідовності, а також витягування інформації з їх допомогою;
- визначення трендів та коливань. Це є початковим етапом задачі прогнозування.

Під час проходження стадії вільного пошуку так само має здійснюватись валідація закономірностей, тобто їх перевірка достовірності на частини даних, що не брали участі в формуванні закономірностей.

Прогностичне моделювання є другою стадією ІАД. Дана стадія використовує результати роботи першої стадії. На даному етапі закономірності, що були знайдені, використовуються безпосередньо для прогнозування. Прогностичне моделювання охоплює наступні дії:

- прогнозування невідомих значень;
- прогнозування розвитку процесів.

Під час прогностичного моделювання вирішуються задачі класифікації і прогнозування.

В процесі вирішення задачі класифікації результати роботи першої стадії використовуються для зарахування нового об'єкту із певною впевненістю до одного із відомих визначених наперед класів, підставою яких є відомі значення.

В процесі вирішення задачі прогнозування результати першої стадії використовуються для прогнозу невідомих значень цільової змінної або декількох змінних.

Закономірності, що були отримані на даній стадії, формуються від часткового до загального. В результаті одержується деяка загальна інформація про деякий клас об'єктів на підставі дослідження окремих представників даного класу.

Прогностичне моделювання, на відміну від стадії вільного пошуку, є дедуктивним. Закономірності, що були одержані на даній стадії, формуються від загального до часткового. В результаті одержується нова інформація про деякий об'єкт або групу об'єктів на підставі:

- інформація класу, до якого належать досліджувані об'єкти;
- інформація загального правила, яка діє у межах даного класу об'єктів.

Аналіз виключень є третьою стадією ІАД, яка проводить аналіз аналіз виключень або аномалій, що могли бути виявлені в знайдених закономірностях. Дією, яка виконується на даній стадії є виявлення відхилень. Для виявлення відхилень потрібно визначити норму, що розраховується під час стадії вільного пошуку. Також існує можливість використання стадії аналізу виключень як очищення даних [4].

1.2 Аналіз застосування досліджуваних технологій в існуючих системах

На базі аналізу літературних джерел можна зробити висновки щодо особливостей, перспектив використання і можливостей подальшого розвитку інтелектуального аналізу даних в сучасних умовах розвитку комп'ютерних технологій.

Сфера застосування ІАД не обмежена. Вона є скрізь, де є дані. Проте насамперед методи ІАД на сьогоднішній день цікавлять комерційні підприємства, метою яких є розгортання власних проектів на базі інформаційних сховищ даних. ІАД є великою цінністю для керівників та аналітиків в їх повсякденній діяльності. Вони ж усвідомлюють, що з методами ІАД існує можливість одержати відчутні переваги в конкурентній боротьбі на ринку. Нижче розглянуті сучасні системи, у основі яких покладений ІАД.

Предметно-предметно-орієнтовані аналітичні системи

Предметно-предметно-орієнтовані аналітичні системи є дуже різноманітними. Один з найширших підкласів даних систем, який одержав поширення в галузі дослідження фінансових ринків, має назву «технічний аналіз». Даний підклас є сукупністю кількох десятків методів прогнозу динаміки цін та вибору оптимальної структури інвестиційного портфеля, які ґрунтуються на основі різних емпіричних моделей динаміки ринку. Описані методи часто використовують нескладний статистичний апарат, проте максимально враховують сформовану специфіку своєю областю, наприклад, системи різних індексів, професійна мова тощо. На ринку існує багато програм даного класу. Як правило, вони є досить дешевими.

Статистичні пакети

Останні версії майже всіх відомих статистичних пакетів ставлять елементи ІАД поряд з традиційними статистичними методами. Проте основна увага у них приділяється класичним методикам, таким як регресійномий, факторномий, кореляційномий аналіз тощо. Одним з недоліків систем даного класу є вимога до спеціальної підготовки користувача. Також відзначається, що сучасні потужні

статистичні пакети є доволі об'ємними для масового застосування в фінансовій і бізнес-сферах. До того ж досить часто дані системи є дорогими і витратними. Також існує ще більш серйозний принциповий недолік статистичних пакетів, який обмежує їх застосування в ІАД. Більша частина методів, які належать у склад пакетів, опираються на статистичну парадигму, де головними фігурантами є усереднені характеристики вибірки. Дані характеристики, як вказувалося вище, під час дослідження реально існуючих складних життєвих феноменів часто є фіктивними величинами.

Нейронні мережі

Нейронні мережі є великим класом систем, архітектура яких має аналогію з побудовою нервової тканини із нейронів. У одній із найпоширеніших архітектур – багатошаровому перцептроні із зворотним поширенням помилки, імітується робота нейронів в складі ієрархічної мережі, в якій кожен нейрон вищого рівня поєднаний своїми входами із виходами нейронів нижчого шару. На нейрони найнижчого шару подаються значення вхідних параметрів, на основі яких необхідно приймати рішення, прогнозувати розвиток ситуації тощо. Дані значення розглядаються як сигнали, які передаються в наступний шар, послаблюючись або підсилюючись, залежно від числових значень, які приписуються міжнейронним зв'язкам. В результаті роботи на виході нейрона найвищого верхнього шару виробляється деяке значення, що розглядається як відповідь – реакція всієї мережі на введені значення вхідних параметрів. Щоб мережу можна було застосовувати і далі, необхідно провести її навчання на основі отриманих раніше даних, для яких відомі і значення вхідних параметрів, і правильні відповіді на них. Навчання полягає в підборі значень міжнейронних зв'язків, відповідей, які забезпечують найбільшу близькість мережі до відомих правильних відповідей. Одним з основних недоліків нейромережевої парадигми – це необхідність в дуже великому обсязі навчальної вибірки. Іншим великим недоліком є те, що навіть натренована нейронна мережа є так званою «чорною скринькою». Знання, які зафіксовані як значення декількох сотень міжнейронних зв'язків, зовсім не піддаються аналізу та інтерпретації людиною.

Системи міркувань на основі аналогічних випадків

Системи міркувань на основі аналогічних випадків як ідея на перший погляд є доволі простою. Для прогнозування на майбутнє або вибору правильного рішення дані системи знаходять близькі аналоги наявної ситуації в минулому і обирають ту саму відповідь, що була правильною для них. Саме тому даний метод також має назву «метод найближчого сусіда». В останній період часу поширення також одержав термін «міркування на основі пам'яті», який акцентує увагу на прийнятті рішення, підставою якого є вся інформація, що накопичена у пам'яті.

Системи CBR

Системи CBR демонструють продуктивні результати в різноманітних задачах. Одним із головних мінусів є те, що в них відсутнє створення моделей чи правил, які б узагальнювали попередній досвід. Під час вибору рішення системи ґрунтуються на даних всього масиву доступних історичних даних. Саме тому неможливо сказати, на основі яких конкретних чинників системи CBR будують власні відповіді. Іншим мінусом є безлад, що допускають системи CBR в процесі вибору міри «близькості». Від цієї міри в найвирішальніший спосіб залежить обсяг множини прецедентів, які необхідно зберігати у пам'яті для досягнення задовільної класифікації чи прогнозу.

Дерева рішень

Дерева рішень є одним з найпопулярніших підходів до розв'язання задач ІАД. Даний підхід створює ієрархічну структуру правил типу «якщо... то...», яка має вид дерева. Задля ухвалення рішення, до якого класу необхідно зарахувати деяку ситуацію чи об'єкт, необхідно відповісти на запитання, які стоять в вузлах даного цього дерева, починаючи із його коріння. Питання мають вигляд «значення параметра А більше х?». В разі, якщо відповідь позитивна, відбувається перехід до правого вузла наступного рівня, в разі негативної відповіді – до лівого вузла; після цього необхідно знову поставити запитання, яке пов'язане із відповідним вузлом. Популярність даного підходу пов'язана із його зрозумілістю. Проте дерева рішень нездатні знаходити найповніші і найточніші правила у даних. Вони реалізують принцип послідовного перегляду ознак і звертаються до

фрагментів справжніх закономірностей, створюючи тільки ілюзію логічного висновку.

Генетичні алгоритми

ІАД не є основною областю використання генетичних алгоритмів. Їх потрібно розглядати швидше як потужний засіб вирішення різних задач оптимізації та комбінаторних задач. В той же час, на сьогоднішній день, генетичні алгоритми ввійшли до стандартного інструментарію методів ІАД. Саме тому вони і були включені для розгляду. Першим кроком в процесі побудови генетичних алгоритмів є кодування вихідних логічних закономірностей в базі даних, що мають назву хромосоми, а весь набір подібних закономірностей має назву популяція хромосом. Наступним кроком відбору вводиться спосіб зіставлення різноманітних хромосом для реалізації концепції. Популяція обробляється за допомогою процедур мінливості, генетичної композиції репродукції. Дані процедури імітують біологічні процеси. Найважливішими серед них є випадкові мутації даних у індивідуальних хромосомах, переходи, або кросинговер та рекомбінація генетичного матеріалу, що міститься у індивідуальних батьківських хромосомах, а також міграції генів. В процесі роботи процедур на кожній стадії еволюції виходять популяції із все більш досконалішими індивідуумами. Генетичні алгоритми зручні тим, що їх легко розпаралелювати. Прикладом може слугувати розбиття покоління на декілька груп та робота із кожною із них незалежно, обмінюючись час від часу кількома хромосомами в процесі. Також існують і інші методи розпаралелювання генетичних алгоритмів. З іншого боку, генетичні алгоритми мають багато недоліків. Критерієм відбору хромосом та використовуваних процедур є евристичним і далеко не завжди гарантує знаходження найкращого рішення. Як і у реальному житті, еволюцію може заклинити на будь-якій непродуктивній галузі. Або навпаки, існують приклади, коли два безперспективні батьки, що будуть виключені із еволюції генетичним алгоритмом, виявляються придатними зробити високоефективного нащадка. Це стає особливо помітним в процесі розв'язування високорозмірних задач із складними внутрішніми зв'язками.

Системи для візуалізації багатовимірних даних

В тій чи іншій мірі, засіб для графічного зображення даних підтримується всіма системами SFL. До того ж, доволі велику частину ринку займають системи, які спеціалізуються винятково на даній функції. В системах для візуалізації багатовимірних даних значна увага приділяється толерантності користувацького інтерфейсу, який дає можливість асоціювати із аналізованими показниками різні параметри діаграми розсіювання записів бази даних. До подібних параметрів належить форма, колір, розміри, орієнтація щодо власної осі і інші властивості графічних елементів зображення. Окрім цього, системи візуалізації даних позначені зручними засобами для масштабування і обертання зображень.

1.2.1 Опис існуючих напрямків використання ІАД

В минулому розділі були проаналізовані існуючі системи ІАД. Даний підрозділ опише можливі напрямки використання ІАД.

Роздрібна торгівля

На сьогоднішній день підприємства роздрібною торгівлі збирають докладну інформацію про кожну окрему закупівлю. В процесі використовуються кредитні картки з маркою магазину і автоматизована система контролю. Далі будуть перелічені типові задачі, що можна виконувати за допомогою ІАД в сфері роздрібною торгівлі:

- аналіз купівельного кошика, або аналіз подібності. Призначається для виявлення товарів, що покупці прагнуть купувати разом. Знання купівельного кошика потрібне для поліпшення реклами, вироблення стратегії створення запасів товарів та способів їх розкладки в торговельних залах;

- дослідження часових шаблонів допомагає торговельним підприємствам ухвалювати рішення щодо створення товарних запасів. Це дає відповіді на питання «Якщо сьогодні покупець придбав деякий товар, то через якісь час він, ймовірно, придбає нові додаткові складові, що належать до товару?»;

– створення прогнозних моделей створює можливість торговельним підприємствам дізнаватися про характер потреб різних категорій клієнтів із визначеною поведінкою. Для прикладу можна взяти купівлю товарів відомих брендів чи відвідування розпродажів. Такі знання необхідні для розроблення точно спрямованих економічних заходів для просування товарів.

Банківська справа

Досягнення технології ІАД використовуються також і в банківській справі для виконання наступних задач:

– сегментація клієнтів. Розбиваючи клієнтів по різних категоріях, банки роблять свою маркетингову політику цілеспрямованішою та результативнішою, пропонуючи різні види послуг різним групам клієнтів;

– прогнозування змін клієнтури. ІАД допомагає банкам будувати прогнозні моделі цінності власних клієнтів та в відповідний спосіб обслуговувати кожную категорію покупців;

– виявлення шахрайства з кредитними картками. Аналізуючи минулі транзакції, що в процесі виявилися шахрайськими, банк виявляє деякі закономірності подібного шахрайства.

Телекомунікації

У сфері телекомунікацій методи ІАД допомагають компаніям енергійніше просувати власні програми маркетингу та ціноутворення, з метою утримування вже існуючих клієнтів та залучання нових. Серед типових заходів виділяються наступні:

– виявлення лояльності клієнтів. ІАД можна використовувати для визначення характеристик клієнтів, що нечасто користуються послугами даної компанії, із великою ймовірністю залишаться їй вірними. В підсумку засоби, що виділяються на маркетинг, є можливість витратити там, де віддача буде найбільша.

– аналіз записів про докладні характеристики викликів. Метою призначення подібного аналізу є виявлення категорій клієнтів із подібними стереотипами користування їх послугами і розроблення привабливих наборів цін та послуг.

Страховання

Страхові компанії протягом багатьох років накопичують досить великі обсяги даних. Далі наведені методи ІАД для використання в даній сфері:

– аналіз ризику. Шляхом виявлення взаємозв'язку чинників, які пов'язані із оплаченими заявами, працівники страховлі компанії мають можливість зменшити власні втрати по зобов'язаннях.

– виявлення шахрайства. Страхові компанії можуть понизити рівень шахрайства, відшукуючи деякі стереотипи в заявах про виплату страхового відшкодування, які характеризують взаємодію між лікарями, юристами і заявниками;

Медицина

Існує багато експертних систем для постановки медичних діагнозів. Вони побудовані переважно на базі правил, які описують сполучення різноманітних симптомів різних захворювань. За допомогою цих правил розпізнаються не тільки хворобу пацієнта, а й методи його лікування. Правила допомагають визначати показання та протипоказання, творювати умови найефективнішого лікування, обирати засоби медикаментозного впливу, передбачати результати призначеного курсу лікування, мати орієнтири в лікувальних процедурах тощо. Технології ІАД створюють можливість виявляти дані шаблони зазначених правил.

Молекулярна генетика та генна інженерія

В сфері молекулярної генетики та генної інженерії формулюється як визначення маркерів, під якими маються на увазі генетичні коди, які контролюють ті чи інші фенотипічні ознаки живого організму. Подібні коди можуть містити сотні, тисячі і більше пов'язаних між собою елементів. Сфера генетичних досліджень має велике фінансування. Останнім часом в даній галузі створено особливу зацікавленість до застосування методів ІАД.

Прикладна хімія

Методи ІАД широко використовуються в прикладній хімії. В даній сфері доволі часто на порядку денному стоїть питання про з'ясування особливостей хімічної будови тих або інших з'єднань та їх визначальних властивостей. Особливо актуальним дане завдання виявляється під час аналізу складних хімічних сполук, опис яких містить в собі сотні й тисячі структурних елементів і їх зв'язків.

Інші додатки

ІАД може застосовуватися у багатьох інших сферах діяльності, наприклад:

- політика гарантій. Виробникам необхідно враховувати кількість клієнтів, що подадуть гарантійні заявки, а також середню вартість заявок;

- заохочення клієнтів, які часто літають літаками. Авіакомпанії можуть виявити групу клієнтів, яких можна спонукати літати більше за допомогою заохочувальних заходів. Прикладом може слугувати наступна ситуація: авіакомпанія виявила категорію клієнтів, що літали на досить короткі відстані, при цьому, не накопичуючи достатньої суми відстаней для вступу до їхніх клубів. Саме тому, у такий спосіб компанія змінила правила прийому в клуб, задля заохочення кількостей польотів, так само, як і сумму відстаней.

- розвинення автомобільної промисловості. При складанні автомобілів виробники мають враховувати вимоги і потреби кожного окремого клієнта. Саме тому їм необхідна можливість прогнозування популярності певних характеристик та дані того, які характеристики переважно замовляються разом.

На базі розглянутого вище матеріалу можна зробити висновок, що потенціал ІАД створює поштовх до розвитку і розширення меж застосування технології у сучасному світі комп'ютерних технологій. Стосовно перспектив ІАД можливі наступні напрямки розвитку:

- створення формальних мов та логічних засобів, з допомогою яких формалізуються міркування та автоматизація, які стануть інструментом вирішення задач ІАД в конкретних предметних галузях;

- подолання істотного відставання можливостей інструментальних засобів ІАД від теоретичних досягнень в даній області;
- виділення типів предметних сфер із відповідними їм евристичними, формалізація яких полегшить вирішення відповідних задач ІАД, які належать до цих сфер;
- створення методів ІАД, які будуть спроможні не тільки витягувати закономірності із даних, але і формувати деякі теорії, які будуть спиратись на емпіричні дані.

Основною особливістю ІАД є те, що зв'язок широкого математичного інструментарію та останніх досягнень в сфері інформаційних технологій гармонійно поєднуються між собою не тільки чітко формалізованими методами, але й методами неформального аналізу. Тобто, якісний та кількісний аналіз даних.

Більшістю аналітичних методів, які використовуються у ІАД є відомими математичними методами та алгоритмами. Новим у їх використанні є можливість їх застосування в процесі вирішення тих або інших конкретних проблем, що зумовлена новими програмними та технічними засобами, що з'явилися.

Системи інтелектуального аналізу даних застосовуються не тільки як масовий продукт для бізнес-додатків, але й як інструменти для проведення унікальних досліджень в різних областях людської діяльності. Лідери ІАД пов'язують майбутнє таких систем із їх використанням як інтелектуальних додатків, що вбудовані в корпоративні сховища даних. Незважаючи на досить велику кількість методів ІАД, пріоритет поступово зміщується в бік логічних алгоритмів пошуку причинно-наслідкових правил у даних. З їх допомогою вирішуються задачі розпізнавання образів, прогнозування, класифікації сегментації БД, інтерпретації даних, встановлення асоціацій в БД тощо. Результати таких алгоритмів ефективні і легко можуть бути інтерпретовані за необхідністю.

2 ПОСТАНОВКА ЗАДАЧІ НА ДОСЛІДЖЕННЯ

2.1 Аналіз застосування методів інтелектуального аналізу даних в бізнес-сфері

Загострення конкурентної боротьби у сфері бізнес-діяльності створює досить затребувані підходи до експертного використання наявних даних для підвищення обґрунтованості і оперативності прийняття управлінських рішень. Тому бізнес-аналітиків нещодавно почала привертати увагу область, що пов'язана із високопродуктивною інтелектуальним аналізом даних. Найбільший інтерес до ІАД проявляють компанії, які, насамперед, працюють у умовах доволі високої конкуренції і мають чітку групу споживачів. Вони використовують будь-яку можливість для підвищення ефективності власного бізнес-проекту через ухвалення найбільш ефективних управлінських рішень. На сьогоднішній день широко використовуються методи інтелектуального аналізу даних у бізнес-середовищі, що підтверджують на практиці зростаючу актуальність даної інтелектуальної сфери діяльності.

Не зважаючи на те, що інтелектуальні інформаційні технології використовуються вже більше, ніж півстоліття, питання їх впровадження у оптимізацію бізнес-процесів підприємств і компаній вимагають подальшого дослідження та є досить актуальними на сьогоднішній день.

Найбільший інтерес до технологій інтелектуального аналізу даних, в першу чергу, проявляють компанії, які працюють в умовах високої конкуренції на ринку і мають чітку групу споживачів. Вони використовують будь-яку можливість для підвищення ефективності власного бізнесу через ухвалення більш ефективних управлінських рішень. Такі компанії шукають зв'язок між «внутрішніми» факторами, такими як ціна, затребуваність продукту тощо та «зовнішніми» факторами, такими як економічні показники, конкуренція тощо. Це дає можливість оцінити та спрогнозувати рівень продажів та задоволеності серед клієнтів, розмір доходів, а також сформулювати на основі отриманої інформації

подальшу стратегію розвитку. Буває, що віддача від застосування даних інструментів може становити сотні відсотків при порівняно невисокій вартості впровадження. Також, результатом обробки даних має бути такий інформаційний продукт, що дозволить прийняти конкретне управлінське рішення без глибокої участі робітника, який приймає рішення, у деталі базових даних чи проміжної аналітики. До того ж, на практиці є можливість існування ситуації, коли будь-яке рішення у тій чи іншій частині необхідно приймати обов'язково. Але, постає питання яким чином воно буде прийнято: основі об'єктивної інформації чи інтуїтивно.

На сьогоднішній день, методи інтелектуального аналізу даних широко використовуються в бізнес-середовищі, що підтверджують на практиці зростаючу актуальність даної інтелектуальної сфери діяльності. Використовуючи методи інтелектуального аналізу даних, пункт роздрібної торгівлі має можливість фіксувати інформацію про усі покупки клієнта, яка створює можливість виділити цільову аудиторію для демонстрації їй реклами і розіслати рекламні пропозиції клієнтам на основі історії їх покупок і вподобань. Найбільш поширеними на сьогоднішній день прикладами використання інструментів ІАД є різні інтернет-магазини і пошукові системи. На основі ретроспективних запитів користувача вони визначають профіль його бажань та інтересів. На основі визначення профілю можливо запропонувати товари та послуги, що також можуть зацікавити користувача. Серед прикладів подібних сервісів, що використовують методи ІАД є eBay, Google, Amazon тощо.

Є очевидним той факт, що, інструменти ІАД могли б допомогти фахівцям кадрової служби будь-якої компанії. При цьому, прикладами питань, на які шукають відповіді ці фахівці, можуть бути які працівники та наскільки часто роблять помилки, яку сумму компанія втрачає через помилки співробітників, вимоги задоволення здобувачів, яким чином можна оптимізувати режими роботи устаткування, яка існує можливість оптимізації штату та навантаження, як можна скоротити число збоїв та простоїв із технологічних причин тощо.

Інтелектуальний аналіз даних у бізнесі - це процес використання технологій штучного інтелекту, машинного навчання та аналізу даних для отримання цінного інсайту та прийняття стратегічних рішень. Цей підхід дозволяє бізнесу використовувати великі обсяги даних для отримання конкурентних переваг та оптимізації різних аспектів діяльності.

Бізнес-аналітика, або BI — це комп'ютерні методи та інструменти, створені для організацій, які забезпечують переклад транзакційної ділової інформації у форму, що буде придатною для бізнес-аналізу, а також засоби для роботи із обробленою таким чином інформацією.

Термін «Бізнес-аналітика» описує процес і методи взаємодії із вже очищеними та підготовленими даними для аналізу та використовує статистичні і кількісні інструменти для оцінки поточної ситуації і прогнозування. Саме тому існує і використовується термін «поглиблена аналітика» стосовно даного процесу.

Бізнес-аналітика, насамперед, займається очищенням та консолідацією даних, перетворюючи їх в зручний формат для аналізу. Метою подібних задач є інтерпретування великої кількості даних, загострюючи увагу тільки на ключових факторах, які впливають на ефективність, моделювати результат різноманітних варіантів дій та відстежувати результати прийнятих рішень. Головним призначенням BI є це саме прийняття рішень для бізнесу.

BI підтримує прийняття безліч бізнес-рішень: від операційних до стратегічних. Основні операційні рішення включають у себе позиціонування продукції чи цін на цю ж продукцію. Стратегічні бізнес-рішення включають у себе пріоритети, напрямки і цілі.

BI найбільш ефективна як система під час її об'єднання з даними, що були отримані із ринку, де працює підприємство, так звані «зовнішні дані», з даними із джерел на підприємстві, такими як, наприклад, виробничі і фінансові, так звані «внутрішні дані». В поєднанні зовнішні та внутрішні дані дають більш повну картину бізнеса, іншими словами, аналітику, яку не можна отримати у процесі аналізу даних тільки від одного із цих джерел.

Трирівнева архітектура традиційної системи BI складається із трьох шарів:

- рівня презентації;
- рівня додатків;
- рівня бази даних.

Основним завданням даної архітектури є те, яким чином можна виконати такі цілі рівня обслуговування як мінімальна швидкість проходження і максимальний час відгуку. Це можна пояснити тим, що управління зберіганням даних на низькорівневих шарах приховано від рівня додатку, що ускладнює передбачення часу виконання. Трирівнева архітектура традиційної системи ВІ показана на рисунку 2.1.

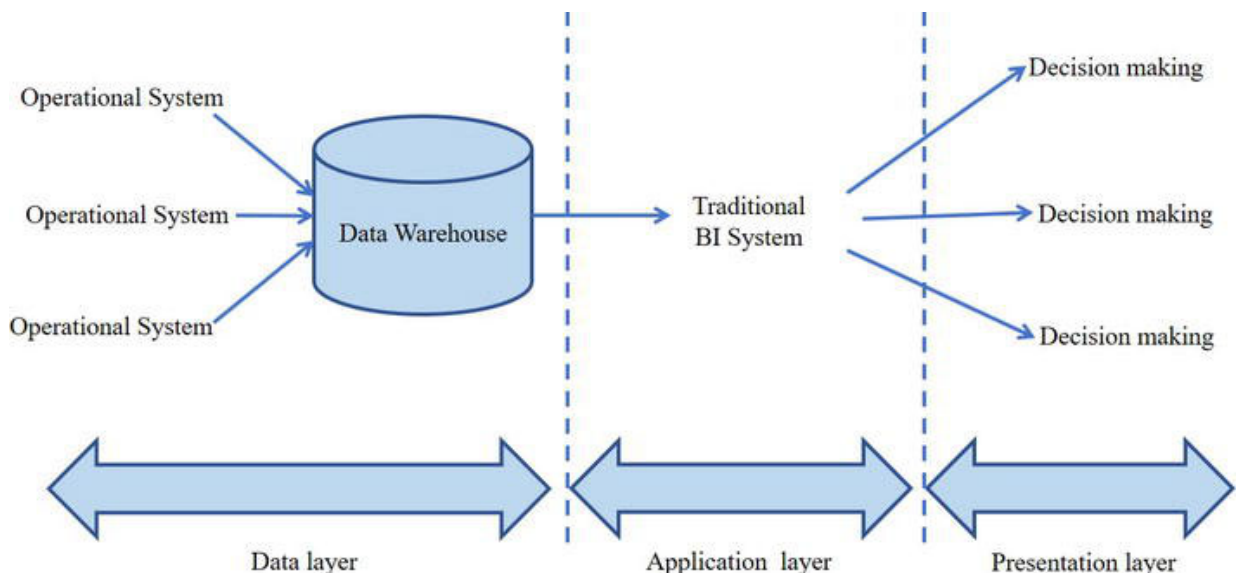


Рисунок 2.1 Традиційна система ВІ

На сьогоднішній день, ВІ-системи розвиваються за чотирма основними напрямками, а саме:

- аналіз даних. Для всебічного аналізу даних використовуються OLAP-інструменти. Вони дозволяють розглядати різні види даних, виявляти тренди та залежності від різних факторів;

- представлення даних. Для результативного представлення даних використовуються різноманітні графічні засоби, такі як графіки, діаграми, звіти тощо. Найпопулярнішим засобом візуалізації даних є інформаційні панелі, де

результати відображаються в вигляді індикаторів та шкал, які дозволяють контролювати поточні значення вибраних показників, порівнювати їх із мінімально або максимально допустимими значеннями і, за допомогою такого методу виявляти потенційні загрози для бізнесу;

– збереження даних. Дані у сховищі ВІ-системи мають спеціальну структуру для більш ефективного аналізу та обробки запитів. Така структура є більш ефективною, в порівнянні із звичайними базами даних, в яких інформація організована таким чином, щоб оптимізувати час обробки поточних транзакцій;

– інтеграція даних. Для формування та підтримки сховищ даних використовуються ETL-засоби. Це інструменти, які забезпечують отримання даних, їх перетворення, тобто приведення до потрібного формату і завантаження цих даних у сховище чи у іншу базу даних.

Загалом, коренем бізнес-вигоди інтелектуального аналізу даних є збільшення здатності виявляти сховані закономірності, кореляції, тенденції і аномалії в наборах даних. Усю цю інформацію є можливість використовувати для покращення процесу прийняття бізнес-рішень і стратегічного планування за допомогою поєднання звичайного аналізу даних і прогностичної аналітики.

Більш конкретні переваги аналізу даних включають в себе:

– краще обслуговування клієнтів. Завдяки аналізу даних компанії мають можливість виявляти потенційні проблеми з обслуговуванням клієнтів більш оперативно і надавати агентам контакт-центру актуальну інформацію для використання в процесі дзвінків і онлайн-чатів з клієнтами;

– збільшення часу безвідмовної роботи. Отримання оперативних даних з датчиків на виробничих машинах і іншому промисловому обладнанні підтримує програми прогнозованого технічного обслуговування для виявлення потенційних проблем до їх виникнення, тим самим, допомагаючи уникнути незапланованих пауз та простоїв, що є шкідливим для отримання прибутку;

– ефективний маркетинг і продаж. Інтелектуальний аналіз даних допомагає маркетингологам краще зрозуміти поведінку, потреби і вподобання клієнтів, що дозволяє на базі отриманої інформації створювати цільові маркетингові і рекламні

кампанії. Подібним чином відділи продажів мають можливість використовувати результати інтелектуального аналізу даних для підвищення коефіцієнту конверсії потенційних клієнтів і продавати додаткові продукти і послуги вже існуючим клієнтам;

- економія бюджету. Інтелектуальний аналіз даних допомагає заощаджувати кошти за рахунок підвищення операційної ефективності бізнес-процесів і зменшення надмірності і марнотратства корпоративних витрат;

- покращене управління постачанням. Організації можуть помічати ринкові тенденції і більш точно створювати прогнозування попиту на продукцію. Це дає змогу краще керувати запасами товарів. Менеджери ланцюгів постачання також можуть використовувати інформацію із аналізу даних для оптимізації складування, розподілу і інших логістичних операцій;

- посилення управління ризиками. Менеджери ризиків і керівники підприємств можуть більш якісно оцінювати правові, фінансові, кібербезпекові і інші ризики і загрози для компанії і створювати плани управління ними;

- прогнозування та оптимізація. Використання аналітики та моделей прогнозування для передбачення майбутніх тенденцій та оптимізації операцій. Наприклад, прогнозування споживання товарів або оптимізація ланцюга постачання;

- персоналізований маркетинг. Використання аналітики та машинного навчання для створення персоналізованих стратегій маркетингу. Аналіз великих обсягів даних про клієнтів дозволяє створити індивідуальні пропозиції та знижки;

- аналіз соціальних мереж та взаємодії. Використання аналізу соціальних мереж для збору даних про взаємодію з брендом, відгуки клієнтів та використання цих даних для удосконалення продуктів та послуг;

- аналіз клієнтської інформації. Використання даних про клієнтів для розуміння їх потреб та уподобань. Інтелектуальний аналіз може допомагати створювати більш ефективні стратегії обслуговування та збільшувати рівень задоволення клієнтів;

– процеси прийняття рішень. Використання аналітики та машинного навчання для полегшення процесів прийняття рішень у бізнесі. Моделі можуть надавати конкретні поради на основі аналізу даних;

– ефективність операцій. Використання аналітики для виявлення та усунення неефективності операцій. Це може включати оптимізацію роботи виробничих процесів, управління запасами та розподіл ресурсів;

– бізнес-прогнозування. Використання аналізу даних для прогнозування ринкових тенденцій, кон'юнктури та інших факторів, що можуть впливати на стратегію розвитку бізнесу.

Інструменти бізнес-аналізу допомагають компаніям залишатись конкурентоспроможними на ринку і максимізувати потоки доходу. Організації будь-якого розміру і етапу використовують програмне забезпечення ВІ для аналізу, управління і візуалізації бізнес-даних.

Таким чином, використання інтелектуального аналізу даних можуть призвести до підвищення і прибутків компаній в різних галузях, а також конкурентних переваг, що відрізняють компанії від їх бізнес-конкурентів.

2.2 Аналіз шляхів збільшення попиту готелю за допомогою використання інструментів інтелектуального аналізу даних

В сучасних умовах конкуренції в готельній галузі, де важливим фактором є не лише якість обслуговування, але й ефективна маркетингова стратегія, інтелектуальний аналіз даних виступає ключовим інструментом для розуміння та прогнозування попиту.

Збільшення попиту на готель за допомогою інтелектуального аналізу даних може бути виконане за допомогою різних стратегій. Далі будуть представлені конкретні методи:

Сегментація клієнтів. Використання аналітики для ідентифікації різних сегментів клієнтів дозволяє готелям краще розуміти свою цільову аудиторію. Наприклад, можливі сегменти включають бізнес-подорожуючих, туристів,

молодіжні групи, сімейні подорожі тощо. Для кожного сегменту можна створити унікальні пакети, пропозиції та послуги, що відповідають їхнім потребам і вподобанням.

Прогнозування попиту. Використання аналіз даних для прогнозування попиту в різні періоди року або для конкретних подій та свят, минулих тенденцій і прогнозу попиту на номери готелю. Адаптування цін та пропозиції відповідно до прогнозу попиту.

Персоналізовані рекомендації. Використання інтелектуальний аналіз для рекомендацій гостям щодо екскурсій, ресторанів та інших послуг на основі їхніх попередніх виборів та вподобань.

Маркетингові кампанії та реклама. Використання даних для персоналізації маркетингових кампаній та реклами. Спрямування реклами на основі демографічних, географічних та поведінкових даних клієнтів.

Залучення через соцмережі. Використання дані з соцмереж для розуміння популярних тенденцій та обговорень у сфері подорожей. Залучання аудиторії через соцмережі, пропонуючи унікальні акції та конкурси.

Вдосконалення сервісу. Використання аналітичні інструменти для аналізу відгуків та оцінок гостей. Впровадження зміни в обслуговуванні та інфраструктурі готелю на основі отриманих даних.

Прогнозування змін клієнтури. будівництва прогнозних моделей цінності власних клієнтів та в відповідний спосіб обслуговувати кожну категорію покупців.

Виявлення лояльності клієнтів. Визначення характеристик клієнтів, що нечасто користуються послугами даної компанії, із великою ймовірністю залишаться їй вірними в майбутньому.

Ці стратегії дозволяють готелям більш ефективно відповідати на потреби та очікування своїх клієнтів, що сприяє збільшенню попиту.

2.3 Аналіз та опис технології Data mining

В результаті розвитку інформаційних технологій кількість даних, які накопичуються у електронному вигляді, зростає досить швидкими темпами. Такі дані існують у різних форматах. В той же час, більшість цієї інформації не приносить ніякої користі, саме тому що її дуже важко обробити для знаходження даних, які зможуть покращити той чи інший чинник. Виникає проблема пошуку корисної інформації із великого масиву даних. Даний розділ розглядає технології Data Mining, які допомагають знайти корисні знання із великих масивів даних з різних сфер.

Data mining є процесом виявлення та сортування великих об'ємів даних для виявлення зв'язків та закономірностей, що мають можливість допомогти вирішити бізнес-проблеми за допомогою інтелектуального аналізу даних. Інструменти і методи data mining дозволяють бізнесам підприємствам передбачати майбутні тенденції, проводячи аналіз і приймати обґрунтованіші бізнес-рішення на основі отриманої інформації.

Для деяких підприємств цілі видобутку даних визнаються для виявлення різноманітних особливостей розвинення маркетингових здібностей і прогнозу перспектив на базі попередніх спостережень і сучасних тенденцій. Існує вимога щодо перевірки даних для підтримки розпродажів і додаткових цілей підприємця або компанії. Окрім цього, існує можливість продовжувати практику видобутку даних для розпізнавання незвичайної продуктивності і виявлення дивної поведінки представників, що практикують деякі технології.

Сфера Data Mining допомає формуванню вмінь і знань у сфері інтелектуального аналізу даних із метою розробки повної моделі функціонування і розвитку реального бізнесу.

Технології сховищ даних і бізнес-аналізу почали з'являтися наприкінці 1980-х, початку 1990-х років. Вони почали забезпечувати більшу здатність аналізувати зростаючі обсяги даних, що займались створенням і збором організації. Термін «Data mining» використовувався ще до 1995 року, в той час

коли у Монреалі відбулася перша міжнародна конференція із виявлення інтелектуального аналізу даних.

Data mining є головною частиною загальної аналітики даних та однією з основних дисциплін в сфері даних, що використовує передові аналітичні методи для пошуку корисної інформації у наборах даних.

На сьогоднішній день data mining є важливим компонентом успішних аналітичних ініціатив у різноманітних організаціях. Інформацію, що генерується, є можливість використовувати у програмах бізнес-аналітики та розширених аналітичних програмах, що включають аналіз історичних даних, а також в аналітичних програмах в реальному часі, що роблять перевірку потокових даних в процесі їх створення чи збору.

Ефективний data mining допомагає у різних чинниках планування бізнес-стратегії і управління операціями. В нього входять функції, що стосуються клієнтів, такі як продажі і підтримка клієнтів, маркетинг, виробництво, управління ланцюгом постачання тощо. Data mining визначається як ефективний інструмент для виявлення шахрайства, управління ризиками, планування кібербезпеки та вирішення ряду критичних бізнес-випадків. Цей аналітичний підхід забезпечує можливість виявлення несанкціонованої або аномальної активності, виявлення потенційних ризиків та розробку стратегій забезпечення кібербезпеки для ефективного функціонування бізнес-систем. Також він відіграє важливу роль у охороні здоров'я, уряді, наукових дослідженнях, математиці, спорті тощо.

Data mining, насамперед, є відповідальністю фахівців із обробки даних та інших кваліфікованих спеціалістів в сфері бізнес-інтелекту. Але цей процес також може бути виконаний бізнес-аналітиками, керівниками і працівниками, які виступають у ролі експертів з обробки даних у внутрішніх структурах організації. Основні компоненти включають у себе застосування машинного навчання і статистичного аналізу, а також вирішення завдань управління даними, що необхідні для підготовки інформації для подальшого аналізу. Завдяки використанню алгоритмів машинного навчання і інструментів штучного інтелекту

була автоматизована значна частина процесу, що значно спростило видобуток обширних даних, таких як бази даних клієнтів, транзакційні записи і журнали із серверів, а також інформація із мобільних додатків та сенсорів.

Процес data mining зазвичай розбивають на чотири основні етапи:

Збір даних починається з визначення та накопичення релевантних даних для подальшої аналітичної програми. Ці дані можуть знаходитися в різноманітних вихідних системах, сховищах даних або озерах даних, особливо поширених у великих даних, що включають комбінацію структурованих і неструктурованих даних. Важливо також враховувати можливість використання зовнішніх джерел даних. Незалежно від походження даних, фахівець із обробки даних часто переміщує їх до озера даних для проведення подальших етапів аналітичного процесу.

Етап підготовки даних включає в себе серію кроків, спрямованих на підготовку даних для подальшого аналізу. Процес розпочинається з вивчення даних, їх профілювання та попередньої обробки, після чого виконується очищення даних для виправлення помилок та вирішення інших проблем, пов'язаних із якістю даних.

Після завершення процесу підготовки даних, фахівець з даних обирає відповідну техніку data mining та реалізує один або кілька алгоритмів для проведення інтелектуального аналізу. У сфері програм машинного навчання зазвичай необхідно протестувати алгоритми на вибіркових наборах даних, щоб виявити необхідну інформацію перед їхнім застосуванням на повному обсязі даних.

Аналіз та інтерпретація даних завершуються створенням аналітичних моделей, які можуть служити основою для прийняття рішень та виконання інших стратегічних дій в бізнесі. Спеціаліст із обробки даних або інший учасник групи з аналізу даних також відповідає за передачу отриманих результатів керівникам компаній та користувачам, часто використовуючи візуалізацію даних та різні методи представлення інформації.

Для аналізу даних у різних програмах обробки інформації можна використовувати різноманітні методи. Один із широко використовуваних підходів інтелектуального аналізу даних (data mining) - це розпізнавання шаблонів, що включає декілька методів, включаючи виявлення аномалій. Основною метою виявлення аномалій є визначення викидних значень або невідповідностей у наборах даних. Основні методи аналізу даних будуть перелічені нижче:

У сфері data mining правила асоціацій представляють собою оператори «якщо-тоді», які визначають зв'язки між елементами даних. Для оцінки цих зв'язків використовуються критерії підтримки та достовірності. Підтримка вимірює, наскільки часто пов'язані елементи з'являються в наборі даних. Достовірність відображає кількість разів, коли вислів «якщо-тоді» виявився точним.

У класифікаційному підході елементи в наборах даних призначаються різним категоріям. До прикладів методів класифікації входять дерева рішень, класифікатори Байєса, k-найближчий сусід та логістична регресія.

У задачі кластеризації елементи даних, що мають спільні характеристики, групуються в кластери. Методи, що використовуються для цього завдання, включають кластеризацію k-середніх, ієрархічну кластеризацію та моделі Гауса.

Регресія представляє собою метод виявлення зв'язків у наборах даних, визначаючи прогнозовані значення на основі набору змінних. Прикладами регресії є лінійна регресія та багатовимірна регресія. Крім того, для виконання регресії можна використовувати дерева рішень та деякі інші методи, які використовуються у класифікації.

В аналізі послідовності та шляху дані використовуються для виявлення шаблонів, де певна послідовність подій або значень призводить до наступних подій чи станів.

Аналіз асоціацій виявляє закономірності між взаємопов'язаними подіями в наборі даних. Найвідоміший алгоритм для вирішення задачі пошуку асоціативних правил - це алгоритм Apriori. Завданням аналізу зв'язків є виявлення взаємозв'язків або залежностей у наборі даних

Прогнозування включає оцінку майбутніх значень показників на основі характеристик історичних даних. Для цього використовуються методи математичної статистики, нейронні мережі та інші аналітичні підходи.

Виявлення відхилень, аналіз відхилень або викидів - це процес виявлення та аналізу даних, які значно відрізняються від загальної маси даних, включаючи виявлення нехарактерних шаблонів.

Оцінювання включає прогнозування неперервних значень ознак.

Візуалізацією є процес створення графічного відображення аналізованих даних. Для вирішення завдань візуалізації використовують графічні методи, що ілюструють наявні закономірності в даних.

Підбивання підсумків є процесом створення опису конкретних груп об'єктів за допомогою аналізу вхідного набору даних

Описані вище завдання поділяються на описові та предиктивні. Описові задачі спрямовані на поліпшення розуміння аналізованих даних, зокрема на кластеризацію та пошук асоціативних правил. Ключовим аспектом у цих моделях є простота та зрозумілість результатів для сприйняття людиною.

Рішення предиктивних завдань розглядається на двох етапах. На першому етапі будується модель на основі набору даних з відомими результатами. На другому етапі ця модель використовується для прогнозу результатів на основі нових даних. У цьому випадку вимагається висока точність роботи побудованих моделей. До цього типу задач відносяться класифікація та регресія, а також деякі аспекти пошуку асоціативних правил, якщо їхні результати можуть бути використані для прогнозування певних подій.

За способами рішення задачі поділяються на ті, які вирішуються за допомогою вчителя і без його допомоги. Категорія навчання з учителем представлена такими задачами, як класифікація, оцінка, прогнозування, в той час як без учителя вирішуються завдання кластеризації.

У випадку рішення за допомогою вчителя аналіз даних проводиться у кілька етапів. Спочатку застосовується конкретний алгоритм Data Mining для побудови моделі аналізованих даних - класифікатора. Потім проводиться навчання

класифікатора, перевіряється якість його роботи, і, у разі необхідності, відбувається додаткове навчання класифікатора. Цей процес повторюється до досягнення необхідного рівня якості або виявлення некоректної роботи алгоритму з даними або відсутності структури даних, здатної для аналізу. До цього типу задач відносяться класифікація та регресія.

Рішення без допомоги вчителя об'єднує завдання, які розкривають описові моделі, такі як виявлення закономірностей в часових рядах макроіндикаторів. Ці завдання можуть бути вирішені без попередніх знань про дані для аналізу. До них належать кластеризація та пошук асоціативних правил.

Під час аналізу часто необхідно визначити, до якого з відомих класів відносяться досліджувані об'єкти, тобто як їх класифікувати. Задачу класифікації розглядають як завдання визначення значення одного з параметрів аналізованого об'єкта на підставі значень інших параметрів. Досліджуваний параметр часто називають залежною змінною, а параметри, що беруть участь у його визначенні, - незалежними змінними. Задача класифікації і регресії розв'язується у два етапи. На першому виділяється навчальна вибірка, до якої входять об'єкти, для яких відомі значення як незалежних, так і залежних змінних. На підставі навчальної вибірки будується модель визначення значення залежної змінної, яку часто називають функцією класифікації або регресії. Для отримання максимально точної функції до навчальної вибірки пред'являються такі основні вимоги: кількість об'єктів, що входять до вибірки, має бути достатньою; до вибірки мають входити об'єкти, що представляють усі можливі класи у задачі класифікації або всю область значень у задачі регресії; для кожного класу в задачі класифікації або кожного інтервалу області значень у задачі регресії вибірка має містити достатню кількість об'єктів. На другому етапі побудовану модель застосовують до об'єктів аналізу. Задача класифікації і регресії має геометричну інтерпретацію. Пошук асоціативних правил є поширеним застосуванням Data Mining. Суть задачі полягає у визначенні наборів об'єктів, які часто зустрічаються, в інформаційних масивах. Ця задача є окремим випадком задачі класифікації. При аналізі потрібною є інформація про послідовність подій, що відбуваються. При виявленні

закономірностей у таких послідовностях можна з певною ймовірністю передбачати появу подій у майбутньому, що дає змогу приймати правильніші рішення. Така задача є різновидом задачі пошуку асоціативних правил і називається секвенційним аналізом. Він широко використовується, наприклад, в телекомунікаційних компаніях для аналізу даних про аварії на різних вузлах мережі.

Задача кластеризації полягає у розділі об'єктів на групи схожих об'єктів, відомих як кластери, що представляють собою сукупності осіб або предметів. Задачі поділу множини елементів на кластери відомі як кластерний аналіз.

Кластеризація може бути застосована практично в будь-якій області, де потрібно дослідження експериментальних чи статистичних даних. Для задачі кластеризації характерно відсутність різниць між змінними та об'єктами. Кластерний аналіз дозволяє розглядати значний обсяг інформації і ефективно стискає великі масиви інформації, роблячи їх компактними. Є деякі особливості, притаманні задачі кластеризації.

Рішення залежить від природи об'єктів даних та їх атрибутів, а також від представлення кластерів і передбачуваних відносин об'єктів даних та кластерів. Потрібно враховувати можливість/неможливість приналежності об'єктів кільком кластерам та визначення поняття приналежності кластеру.

Дані деталізуються для подальшої обробки, що передбачає виявлення і використання формалізованих закономірностей або дистиляцію шаблонів.

Технологія дистиляції шаблонів включає витягання зразка інформації з початкових даних та перетворення його у формальні конструкції. Цей процес відбувається на стадії вільного пошуку, яка є важливою для групи методів Data Mining. Методи цієї групи включають логічні методи, методи візуалізації, методи крос-табуляції та методи, що базуються на рівняннях.

Логічні методи, такі як нечіткі запити, символічні правила, дерева рішень і генетичні алгоритми, підтримують інтерпретацію та представлення знайдених закономірностей у вигляді зрозумілого для користувача формату. Методи крос-

табуляції дозволяють знаходити шаблони, а методи на основі рівнянь виражають закономірності у вигляді математичних виразів.

Інша класифікація відділяє методи Data Mining на статистичні та кібернетичні методи, засновані на підходах до навчання математичних моделей.

Статистичні методи Data Mining включають в себе ряд дій для попереднього аналізу характеру статистичних даних, такі як перевірка гіпотез стаціонарності, нормальності, незалежності, однорідності, оцінка функції розподілу та її параметрів. Вони також охоплюють виявлення зв'язків і закономірностей через лінійний і нелінійний регресійний аналіз та кореляційний аналіз. Багатовимірний статистичний аналіз, такий як лінійний і нелінійний дискримінантний аналіз, кластерний аналіз, компонентний аналіз та факторний аналіз, також входить у цей спектр методів. Крім того, статистичні методи Data Mining враховують динамічні моделі та прогнози на основі часових рядів. Загалом, ці методи можна розділити на чотири групи: описовий аналіз та опис початкових даних; аналіз зв'язків, включаючи кореляційний і регресійний аналіз, факторний аналіз і дисперсійний аналіз; багатовимірний статистичний аналіз, що охоплює компонентний аналіз, дискримінантний аналіз, багатовимірний регресійний аналіз і канонічні кореляції; аналіз часових рядів, який включає динамічні моделі і прогнозування.

Кібернетичні методи Data Mining включають в себе різноманітні підходи, такі як еволюційне програмування, асоціативна пам'ять, нечітка логіка, дерева рішень, системи обробки експертних знань, штучні нейронні мережі, генетичні алгоритми.

Нейронні мережі представляють собою клас моделей, які взяли натхнення з роботи мозку людини і призначені для вирішення різноманітних задач аналізу даних після проходження етапу навчання на даних. Ці моделі базуються на імітації біологічних нейронних мереж мозку, де нейрони моделюються однотипними штучними елементами. Нейронна мережа може бути представлена напрямним графом із зваженими зв'язками, де штучні нейрони є вершинами, а синаптичні зв'язки - дугами.

Застосування нейронних мереж включає автоматизацію процесів розпізнавання образів, прогнозування показників діяльності підприємства, медичну діагностику, прогнозування, адаптивне управління, створення експертних систем, організацію асоціативної пам'яті, оброблення аналогових і цифрових сигналів, синтез та ідентифікацію електронних систем. Нейронні мережі можуть бути використані, наприклад, для передбачення обсягів продажу виробів, показників фінансового ринку, розпізнавання сигналів і конструювання самонавчальних систем. Сама нейронна мережа є комплексом нейронів, які утворюють шари, де кожен нейрон пов'язаний з іншими нейронами в тих чи інших шарах.

Інструменти Data Mining доступні від різних постачальників і, зазвичай, входять до складу програмних платформ, які також включають інші типи аналізу даних та розширені засоби аналітики. Основні функції програмного забезпечення для інтелектуального аналізу даних включають опції підготовки даних, вбудовані алгоритми, підтримку прогнозного моделювання, середовище розробки з графічним інтерфейсом користувача та інструменти для розгортання та оцінки ефективності моделей.

Серед постачальників інструментів для аналізу даних можна відзначити такі компанії, як Alteryx, AWS, Databricks, Dataiku, DataRobot, Google, H2O.ai, IBM, Knime, Microsoft, Oracle, RapidMiner, SAP, SAS Institute, Tibco Software та інші. Зокрема, існують і різноманітні безкоштовні технології з відкритим кодом, такі як DataMelt, Elki, Orange, Rattle, scikit-learn і Weka. Деякі постачальники програмного забезпечення також пропонують варіанти з відкритим кодом; наприклад, Knime об'єднує аналітичну платформу з відкритим вихідним кодом і комерційне програмне забезпечення для управління додатками для обробки даних, а такі компанії, як Dataiku та H2O.ai, мають безкоштовні версії своїх інструментів.

Data Mining порівнюється з аналітикою даних і зберіганням даних. Інтелектуальний аналіз даних часто вважається синонімом аналізу даних, але його можна розглядати як конкретний аспект аналізу, який автоматизує виявлення

інформації великими наборами даних. Отриману інформацію можна використовувати для подальшої обробки даних та інших аналітичних застосувань.

Сховища даних, зазвичай, спрямовані на забезпечення можливості пошуку даних, створюючи репозитарії для наборів даних. Традиційно історичні дані зберігаються в корпоративних сховищах даних або менших базах даних, створених для конкретних бізнес-підрозділів чи для зберігання певних підмножин даних. Однак зараз програми інтелектуального аналізу даних часто використовують озера даних, що об'єднують історичні та потокові дані, та базуються на платформах великих даних, таких як Hadoop і Spark, базах даних NoSQL або хмарних об'єктах зберігання.

3 ДОСЛІДЖЕННЯ МЕТОДІВ ВИРШЕННЯ ЗАДАЧІ

3.1 Загальні кроки використання Data Mining

Сегментація клієнтів за допомогою Data Mining дозволяє готелям краще розуміти свою аудиторію та надавати більш персоналізовані послуги. Існує кілька кроків, які можна виконати з використанням Data Mining у Microsoft Analysis Services.

Збір та підготовка даних. Зібрати та об'єднати дані про клієнтів, такі як історія бронювань, дані про використані послуги, інформація про відгуки та інше.

Використання ETL-процеси для підготовки даних для аналізу.

Визначення цільових змінних, які можуть передбачити чи аналізувати, такі як вартість бронювання, тривалість перебування, імовірність повторного бронювання тощо.

Вибір алгоритму Data Mining, який потрібен. Data Mining в MSAS підтримує різні алгоритми, такі як класифікація, кластеризація, прогнозування тощо. Наприклад, для сегментації клієнтів можна використовувати алгоритм кластеризації для групування клієнтів зі схожими характеристиками.

Тренування Моделі. Використання інструментів SSAS DM для тренування моделі на основі існуючих даних.

Встановлення параметрів моделі та оцініть її ефективність на основі навчальних даних.

Валідація та тестування моделі. Валідація моделі на тестових даних для перевірки її точності та адаптабельності. Модель має правильно сегментувати клієнтів і може виявляти корисні патерни.

Сегментація та профілювання. Використання отриманих кластерів для сегментації клієнтів за різними параметрами, такими як споживання послуг, вартість проживання, частота відвідувань тощо. Створювання профілів кожного сегменту клієнтів.

Впровадження результатів. Застосування отриманих результатів для персоналізації маркетингових кампаній, пропозицій та обслуговування. Спрямування реклами на основі інтересів кожного сегменту.

Моніторинг та оновлення. Моніторинг ефективності сегментації та оновлення моделі при необхідності. Регулярний аналіз нових даних та враховування їх у процесі сегментації.

Сегментація клієнтів за допомогою Data Mining дозволяє готелям краще розуміти потреби своїх клієнтів та пристосовувати свою стратегію обслуговування та маркетингу для кожного сегменту.

Галузеві приклади використання data mining в деяких галузях як частина аналітичних програм.

В галузі роздрібної торгівлі, інтернет-магазини використовують Data Mining для збору даних про клієнтів та реєстрації їхньої активності в Інтернеті з метою цільового налаштування маркетингових кампаній, рекламних заходів та пропозицій для окремих покупців. Прогностичне моделювання та інтелектуальний аналіз даних дозволяють також реалізовувати рекомендаційні системи, що пропонують відвідувачам веб-сайту можливі покупки, а також управляти запасами та ланцюгами постачань.

У фінансовій сфері банки та компанії, що надають кредитні картки, використовують інструменти Data Mining для створення моделей фінансового ризику, виявлення шахрайських транзакцій та аналізу заявок на позики та кредити. Інтелектуальний аналіз даних також важливий для маркетингових стратегій та виявлення можливостей збільшення продажів серед існуючих клієнтів.

У галузі страхування Data Mining використовується для визначення цін на страхові поліси, рішення про схвалення заявок на страхування та моделювання ризиків для потенційних клієнтів.

Виробничий сектор використовує програми Data Mining для оптимізації роботи на виробничих підприємствах, управління ланцюгами постачань та підвищення безпеки продукції.

У розважальній галузі сервіси потокового відтворення використовують Data Mining для аналізу переглядів та прослуховувань користувачів, створення персоналізованих рекомендацій на основі їхніх звичок.

Сфера охорони здоров'я використовує інтелектуальний аналіз даних для діагностики захворювань, лікування пацієнтів та аналізу медичних зображень, таких як рентгенівські знімки. Медичні дослідження також великою мірою залежать від аналізу даних, машинного навчання та інших аналітичних методів.

3.1.1 Шляхи використання Data Mining для аналізу попиту готелю

Аналіз попиту готелю за допомогою Data Mining може включати в себе ряд методів та технік для виявлення корисної інформації та закономірностей у великих обсягах даних. Нижче описані можливі варіанти, як можна використати Data Mining для аналізу попиту готелю:

Визначення цілей. Визначення конкретних питань, на які необхідно отримати відповіді. Це може стосуватися аналізу факторів, які найбільше впливають на попит, або визначення типу клієнта, які частіше залишаються.

Збір даних. Зібрати дані, пов'язані з готельним бізнесом, такі як бронювання, дані клієнтів, інформація про номери, час перебування тощо, а також даних, які можуть впливати на попит: історія бронювань, додаткові послуги, інформація про події в регіоні, погодні умови тощо.

Визначення цільової змінної для прогнозування, наприклад, кількість бронювань на певний період.

Очищення та підготовка даних. Виконання операції по очищенню та підготовці даних, враховуючи видалення дублікатів, заповнення пропущених значень, агрегацію та інші операції.

Експлораторний аналіз даних. Використання статистичні методи та візуалізацію даних для виявлення патернів, кореляцій та особливостей, які можуть бути важливими для розуміння попиту.

Видобування знань. Застосування алгоритмів Data Mining для видобутку прихованої інформації. Наприклад, асоціативні правила для виявлення взаємозв'язків між різними атрибутами для виявлення комбінації послуг, які обирають клієнти.

Прогнозування попиту. Використання методів прогнозування для передбачення майбутнього попиту. Можливе використання методів часових рядів або регресії для прогнозування кількості бронювань.

Кластеризація клієнтів. Використання алгоритмів кластеризації для групування клієнтів за схожими характеристиками, щоб легше розробити таргетовані маркетингові стратегії. Використання алгоритмів класифікації Data Mining для розподілу клієнтів на різні групи в залежності від їхніх характеристик та попередніх виборів. Аналіз маркетингових стратегій на характеристиках кожного сегменту клієнтів для більш точного таргетування.

Персоналізовані пропозиції. Використання Data Mining для аналізу історії бронювань та виборів клієнтів. Генерація персоналізованих пропозицій та рекламні промокоди на основі індивідуальних потреб та виборів кожного клієнта.

Асоціативний аналіз. Використання алгоритмів асоціативного аналізу для виявлення зв'язків між різними продуктами або послугами, що може допомогти у формуванні пакетів та пропозицій.

Оцінка та впровадження знань. Оцінка та валідація знань, отриманих з аналізу, та впровадження отриманих висновків у стратегії бізнесу.

Моніторинг та оновлення. Встановлення системи моніторингу, яка дозволяє вам слідкувати за змінами в попиті, оновлювати моделі та аналізувати їхню продуктивність в реальному часі. Впровадження прогнозних результатів у планування ресурсів готелю, включаючи персонал, інвентар, та ціноутворення. Моніторинг реального попиту та порівняння його з прогнозами, адаптуючи модель при необхідності. Імплементация системи персоналізованих рекомендацій може значно полегшити процес вибору для клієнтів та зробити їхній візит більш комфортним та приємним.

Тренування моделі. Використання тренувальних даних для навчання моделі прогнозування. Встановлення параметрів моделі, такі як тривалість прогнозу, інтервали часу тощо.

Валідація та тестування моделі. Валідація моделі на тестових даних, щоб перевірити її точність та ефективність.

Тестування моделі на різних періодах для перевірки її здатності адаптуватися до змін.

Прогнозування попиту. Використання підготовленої моделі для прогнозування кількості бронювань на майбутній період. Оцінка не визначених факторів, такі як свята, події, конференції, які можуть впливати на попит.

Генерація рекомендацій. Використання моделі для генерації персоналізованих рекомендацій для кожного клієнта на основі їхньої історії та попередніх виборів. Рекомендації можуть стосуватися додаткових послуг, ресторанів, екскурсій чи інших зручностей.

Інтеграція з платформами. Інтеграція системи рекомендацій з веб-сайтом готелю, мобільним додатком чи іншими каналами взаємодії з клієнтом.

Аналіз та візуалізація результатів. Вивчення результатів прогнозів та аналіз відхилення між прогнозованими та фактичними значеннями. Використання візуалізації для представлення прогнозів та їхньої ефективності.

Прогнозування попиту готелю з використанням Data Mining може допомогти готелям ефективніше планувати та керувати своїми ресурсами, що в кінці кінців може призвести до оптимізації прибутку та покращення задоволення клієнтів.

Аналіз зручностей та послуг. Використання Data Mining для аналізу використання зручностей та послуг готелю через створення рекламних кампаній, що висвітлюють популярні або недооцінені послуги, щоб привертати більше клієнтів.

Аналіз відгуків та рейтингів. Використання Data Mining для аналізу відгуків та рейтингів клієнтів. Створення реклами, яка відзначає позитивні аспекти та реагує на негативні аспекти для покращення іміджу готелю.

Оптимізація ціноутворення. Використання Data Mining для аналізу цін на номери та конкурентоспроможності готелю. Оптимізація стратегії ціноутворення та створюйте рекламні кампанії, що відзначають конкурентні переваги. Data Mining може стати потужним інструментом для створення спрямованих та ефективних маркетингових кампаній готелю, покращуючи таргетинг та задоволення клієнтів. Використання Data Mining для прогнозування витрат на ресурси та оптимізації ціноутворення. Забезпечення рівноваги між якістю обслуговування та ефективністю витрат.

Аналіз поведінки в соцмережах. Використання Data Mining для аналізу поведінки користувачів у соцмережах, таких як Facebook, Twitter, Instagram. Визначення, які типи вмісту найбільше привертають увагу та взаємодію користувачів. Використання даних для створення таргетованих рекламних кампаній на платформах соцмереж. Налаштування орієнтації реклами за віком, місцем, інтересами та іншими параметрами вашої цільової аудиторії. Реагування на відгуки та коментарі. Використання Data Mining для моніторингу відгуків та коментарів у соцмережах. Застосування Data Mining у соцмережах дозволяє готелям не лише впливати на свою аудиторію більш ефективно, але й покращувати стратегії маркетингу через аналіз та використання даних для здійснення інформованих рішень.

Прогнозування тенденцій та інтересів. Використання алгоритмів прогнозування для визначення майбутніх тенденцій та інтересів у соцмережах. Передбачання, які аспекти готелю можуть взаємодіяти з користувачами у майбутньому.

Персоналізовані рекламні кампанії. Використання отриманих даних для створення персоналізованих рекламних кампаній. Спрямування реклами на основі інтересів, попередніх взаємодій та характеристик користувачів у соцмережах.

Прогнозування ефективності кампаній. Використання аналізу ефективності Data Mining для оцінки результатів маркетингових кампаній. Визначення, які кампанії приводять до більшого взаємодії та конверсії.

Персоналізоване обслуговування. Використання Data Mining для аналізу історії бронювань та перебувань клієнтів. Створення персоналізованих профілів клієнтів для забезпечення індивідуального підходу до кожного гостя.

Адаптація пропозицій та зручностей. Аналіз використання різних зручностей та послуг готелю за допомогою Data Mining. Надання персоналізованих рекомендацій щодо додаткових послуг та пропозицій.

Прогнозування обсягу роботи. Використання алгоритмів прогнозування Data Mining для прогнозування обсягу роботи в готелі (наприклад, кількість гостей, бронювань). Оптимізація розкладу персоналу та ресурсів відповідно до прогнозованих потреб.

Аналіз відгуків та зауважень. Використання Data Mining для аналізу відгуків та зауважень гостей. Виявлення патернів та тенденцій, що можуть бути використані для вдосконалення обслуговування.

Автоматизація обслуговування. Використання Data Mining для ідентифікації рутинних завдань, які можна автоматизувати (наприклад, автоматичне замовлення таксі, прибирання). Зменшення часу очікування та підвищуйте зручність для гостей.

Моніторинг та оцінка ефективності. Спостереження за результатами вдосконалення сервісу та моніторинг зміни в задоволенні клієнтів. Використання аналітики для оцінки ефективності вдосконалення та для виявлення нових можливостей для оптимізації.

Таким чином, використання Data Mining у готельному бізнесі може допомогти вдосконалити різні аспекти обслуговування, що призводить до покращення задоволення гостей та оптимізації операцій готелю.

3.2 Загальні кроки використання Azure Machine Learning

Машинне навчання — це галузь штучного інтелекту, що фокусується на розробці алгоритмів та моделей, які дозволяють комп'ютерам вчитися з даних і виконувати завдання без явного програмування. Основна ідея машинного

навчання полягає в тому, щоб система могла вдосконалювати свою продуктивність на основі досвіду, а не потребувати жорсткого програмного коду для кожної конкретної задачі.

Azure надає широкий набір сервісів та інструментів для роботи з наукою про дані та машинним навчанням. Ці сервіси призначені для допомоги даним вченим та розробникам у побудові, розгортанні та управлінні повноцінними рішеннями в області науки про дані. Нижче наведено ключові компоненти пропозицій Azure в галузі науки про дані:

Сервіс Azure Machine Learning:

Робочий простір: Робочий простір Azure Machine Learning - це хмарна платформа для управління та оркестрації повними робочими процесами в галузі машинного навчання. Вона надає централізоване місце для роботи з даними, кодом, моделями та ресурсами розгортання.

Тетраді: Azure Notebooks дозволяє створювати та запускати тетраді Jupyter в хмарі. Це корисно для колективної роботи в галузі науки про дані та експериментування.

Експерименти: Ви можете використовувати Azure Machine Learning для створення, відстеження та управління експериментами. Це включає в себе запуск та порівняння кількох моделей машинного навчання для знаходження найкращої.

Розгортання моделей: Azure Machine Learning дозволяє розгортати моделі машинного навчання у вигляді веб-сервісів для інтеграції в додатки.

Azure Databricks:

Azure Databricks - це швидка, проста та спільна аналітична платформа на основі Apache Spark. Вона надає середовище для обробки великих обсягів даних та виконання складних завдань аналітики.

Вона інтегрується з сервісами Azure, зокрема з Azure Machine Learning, що дозволяє використовувати потужність розподіленого обчислення для обробки даних великого обсягу та завдань машинного навчання.

Azure Data Factory:

Azure Data Factory - це хмарна служба інтеграції даних, яка дозволяє створювати робочі процеси на основі даних для автоматизації та оркестрації переміщення даних та їх перетворення.

Azure Cognitive Services:

Azure Cognitive Services надають заздалегідь побудовані можливості штучного інтелекту, які можна легко інтегрувати в ваші додатки. Це включає сервіси для комп'ютерного зору, розпізнавання мови, розуміння мови та інше.

Azure SQL Database:

Azure SQL Database - це повністю управляється реляційна база даних. Надає масштабовану, безпечну та інтелектуальну платформу баз даних для ваших даних-орієнтованих додатків.

Azure Synapse Analytics:

Раніше відомий як Azure SQL Data Warehouse, Azure Synapse Analytics - це аналітична служба, яка об'єднує в собі обробку великих даних та склад даних. Вона дозволяє аналізувати великі обсяги даних за допомогою як запитів, так і виділених ресурсів.

Azure DevOps:

Azure DevOps надає набір інструментів та сервісів для розробки, тестування та розгортання додатків. Включає функції керування вихідним кодом, безперервної інтеграції та безперервної поставки (CI/CD), що є важливим для розгортання моделей науки про дані в продакшені.

Ці служби можна комбінувати для створення повноцінних рішень в галузі науки про дані на хмарній платформі Azure. Залежно від конкретного використання та вимог існує вибрати ті сервіси, які найкраще підходять для вирішення поставлених задач.

3.2.1 Шляхи використання Azure Machine Learning для аналізу попиту

Аналіз попиту готелю за допомогою Azure Machine Learning може бути виконаний через різні етапи використання цього сервісу. Давайте розглянемо

конкретний сценарій аналізу попиту та як Azure Machine Learning може бути використаний для оптимізації цього процесу:

Збір та підготовка даних. Збір історичних даних про бронювання готелю, включаючи дати, кількість номерів, тривалість перебування та інші важливі параметри.

Підготовка даних: Виконання очищення та підготовки даних, враховуючи пропущені значення, шкалювання та інші операції.

Створення моделі для прогнозування попиту. Вибір алгоритму. Вибрати алгоритм машинного навчання для створення моделі прогнозування попиту. Наприклад, можна використовувати алгоритми регресії або часових рядів, залежно від характеру даних.

Навчання моделі: Використання Azure Machine Learning для навчання моделі на історичних даних та налаштування параметрів для отримання оптимального прогнозу.

Тестування та валідація моделі. Тестування моделі: Використовуйте тестові дані для перевірки точності та ефективності моделі.

Валідація. Використання валідаційні дані для підтвердження, що модель може адекватно прогнозувати попит на нових даних.

Розгортання моделі в хмару. Створення веб-сервісу. Використання Azure Machine Learning для розгортання навченої моделі як веб-сервісу, який можна викликати для отримання прогнозів попиту.

Моніторинг та оптимізація моделі:

Моніторинг в реальному часі. Використання інструментів Azure для моніторингу продуктивності веб-сервісу та збору метрик.

Оптимізація: Виявляйте аномалії в попиті та оптимізуйте модель за допомогою навчання на нових даних.

Аналіз результатів та прийняття рішень. Візуалізація результатів. Використання інструменти візуалізації даних для вивчення результатів прогнозування та виявлення закономірностей.

Прийняття рішень. На основі аналізу результатів розробити стратегії для оптимізації використання ресурсів готелю та впровадження маркетингових заходів для збільшення попиту.

Використання Azure Machine Learning в такому сценарії дозволяє автоматизувати та полегшити процес аналізу попиту, надаючи потужні інструменти для роботи з даними, навчанням моделей та їхнього ефективного впровадження в продукцію.

Використання Azure Machine Learning для підвищення попиту готелю може включати в себе різні аспекти машинного навчання та аналітики для оптимізації маркетингу, обслуговування клієнтів та управління ресурсами готелю. Ось деякі можливі використання:

Прогнозування попиту. Використання моделей машинного навчання для прогнозування попиту на номери в готелі. Можна враховувати різні фактори, такі як місцезнаходження, погода, святкові дні, події у місцевості та інші параметри.

Персоналізований маркетинг. Використання алгоритмів рекомендацій для створення персоналізованих пропозицій та знижок для окремих клієнтів на основі їхньої історії бронювань та вподобань.

Аналіз задоволення клієнтів. Використання аналізу настроїв та відгуків клієнтів, щоб виокремити та вирішити проблеми та покращити якість обслуговування. Можна використовувати моделі для передбачення задоволення клієнтів та вчасної реакції на проблеми.

Динамічне ціноутворення. Використання моделей для оптимізації ціноутворення на основі попиту, конкурентної ситуації, а також інших економічних та соціальних факторів.

Управління запитами та ресурсами. Використання аналізу для оптимізації розміщення запитів та використання ресурсів готелю відповідно до очікуваного попиту.

Прогнозування сезонності та пікових навантажень. Використання моделей для прогнозування сезонності та пікових навантажень, щоб ефективно планувати ресурси та оптимізувати обслуговування під час пікових періодів.

Створення системи рекомендацій. Використання моделей для створення системи рекомендацій, що допомагатиме клієнтам обирати оптимальні опції та додаткові послуги під час бронювання.

Для реалізації цих використань, необхідно використовувати інструменти Azure Machine Learning для розробки, тренування та впровадження моделей, а також для моніторингу та оптимізації їхньої продуктивності. За допомогою Azure Machine Learning існує можливість створити цілісний підхід до оптимізації діяльності готелю та привертання більшого числа клієнтів.

Azure Machine Learning може бути використаний для реалізації багатьох з завдань описаних в розділі 3.1.1, забезпечуючи зручний інтерфейс для розробки, навчання та впровадження моделей Data Mining.

3.3 Опис аналізу попиту готелю за допомогою Azure Machine Learning

Для аналізу попиту готелю була обрана Azure Machine Learning як один з найпростіших інструментів для використання машинного навчання. Azure Machine Learning є хмарним рішенням, яке дозволяє побудувати і використовувати складні моделі машинного навчання в простій і наочній формі, а також як екосистема, що призначена для поширення і монетизації готових алгоритмів.

Основними складовими Azure Machine Learning є:

Приклади датасетів, за допомогою яких існує можливість тестування інструментів, що мають багато вбудованих наборів даних. Це полегшує процес тестування потужності інструменту. Azure ML поставляється з довгим списком вбудованих наборів даних.

Azure ML має майже всі популярні засоби і алгоритми машинного навчання, а також формули оціночних показників.

За допомогою Azure ML існує можливість перетворення даних за допомогою параметрів фільтрації, параметрів перетворення, параметрів підсумовування і варіантів розрахунку матриці.

Також, існує можливість додавання власного набору даних. Azure ML містить декілька опцій для додавання наборів даних з локальної системи користувача. Можливі варіанти подібних наборів даних зображені на рисунку 3.1.

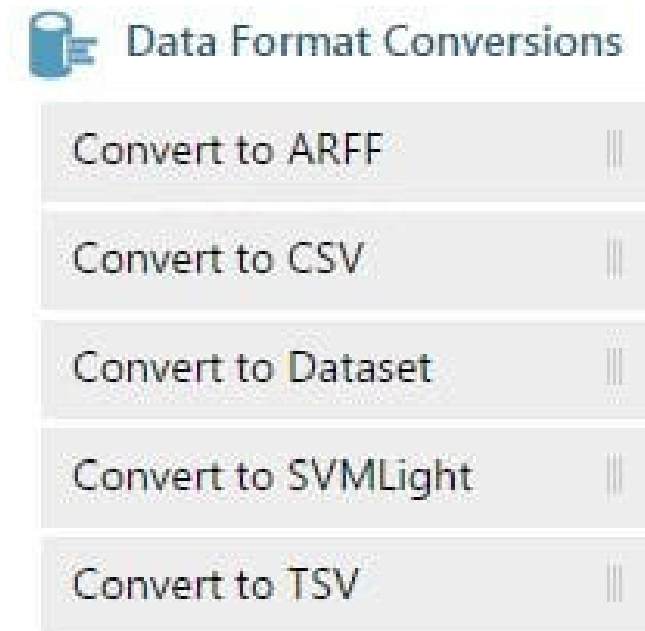


Рисунок 3.1 Набори даних

Нижче описується логічний процес побудови алгоритму машинного навчання.

Визначення мети. Усі алгоритми машинного навчання не мають сенсу без чітко визначеної мети проведення експерименту. У даному аналізі головною метою є прогнозування попиту на основі набору характеристик, які надає кінцевий користувач.

Збір даних. В процесі даного етапу формується вибірка даних, що є необхідною для подальшого навчання моделі. В даному проекті аналізу попиту серед клієнтів входять такі дані як розташування, номери, кухня, сервіс, парковка, персонал, бронювання, технічне обслуговування, безпека, додаткові послуги.

На етапі підготовки даних виконується їх обробка, включаючи формування характеристик, вилучення аномальних значень і розділ вибірки на навчальну та тестову.

Під час розробки моделі визначається вибір однієї чи кількох моделей даних та відповідних алгоритмів навчання, які, на думку розробника, мають надати необхідний результат. Цей процес часто супроводжується паралельним оцінюванням ефективності кількох моделей та візуальним аналізом даних для виявлення можливих закономірностей.

Під час етапу вивчення моделі, алгоритм навчання аналізує вибірку даних для виявлення потенційних прихованих закономірностей з метою розробки методу передбачення. Сам процес пошуку залежить від обраної моделі та алгоритму навчання.

Після завершення процесу навчання модель потребує оцінки її прогностичних можливостей. Це часто досягається, проганяючи модель на тестовому наборі даних та визначаючи рівень відхилення її прогнозів. Залежно від отриманих результатів і вимог до точності, модель може бути прийнята як завершена, або ж може виникнути необхідність в додатковому навчанні, включаючи введення нових вхідних параметрів або навіть зміну алгоритму навчання.

Після успішного проходження тестування навченої моделі вступає в силу етап її використання. У цьому контексті Azure ML виявляється невід'ємним, надаючи всі необхідні засоби для публікації, моніторингу та комерціалізації алгоритмів.

Логічний процес побудови алгоритму машинного навчання і схематично зображений на рисунку 3.2.

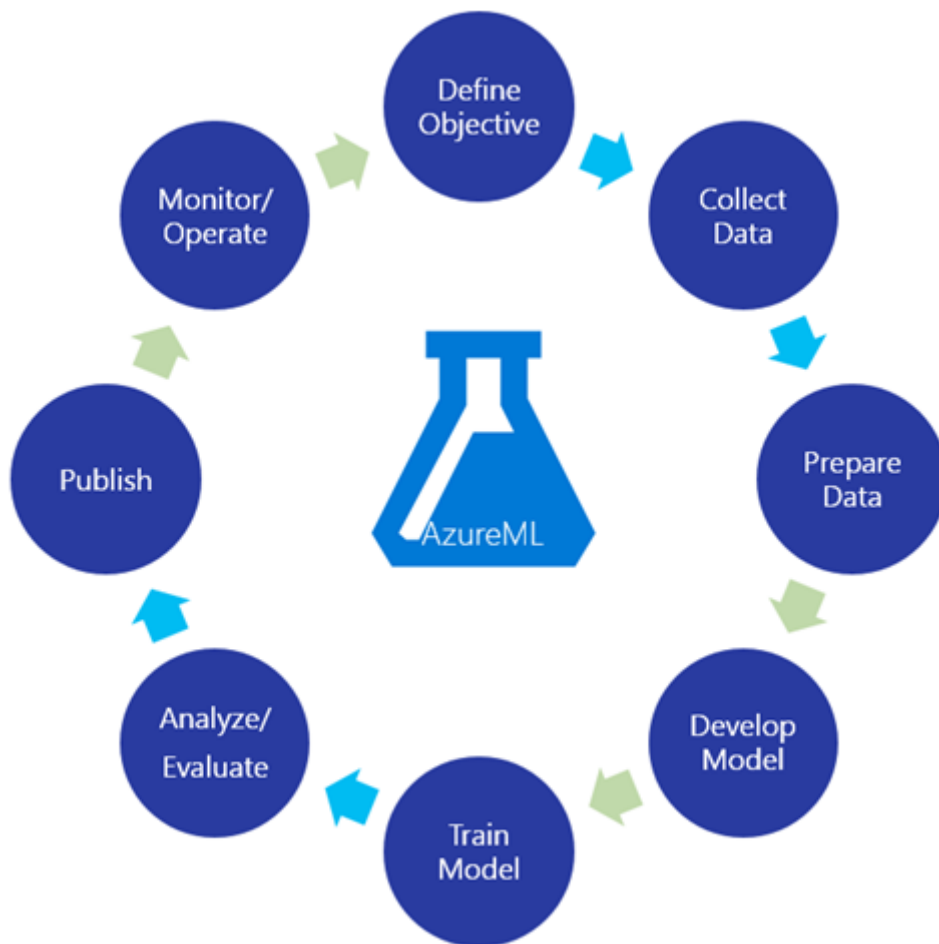


Рисунок 3.2 Логічний процес побудови алгоритму машинного навчання

Наступним кроком є завантаження вибірки даних, що будуть аналізуватися.

Для цього завантажується файл `imports-85.data`, що містить вибірку даних по готелю.

Для завантаження файлу із вибіркою в Azure ML Studio необхідно створити нову модель в нижній частині сторінки і в панелі послідовно обрати Dataset і From Local File. В меню завантаження вказується шлях до завантаженого файлу. Його назва і тип Generic CSV File with no header (.hn.csv). Процес завантаження файлу з вибіркою зображений на рисунку 3.3.

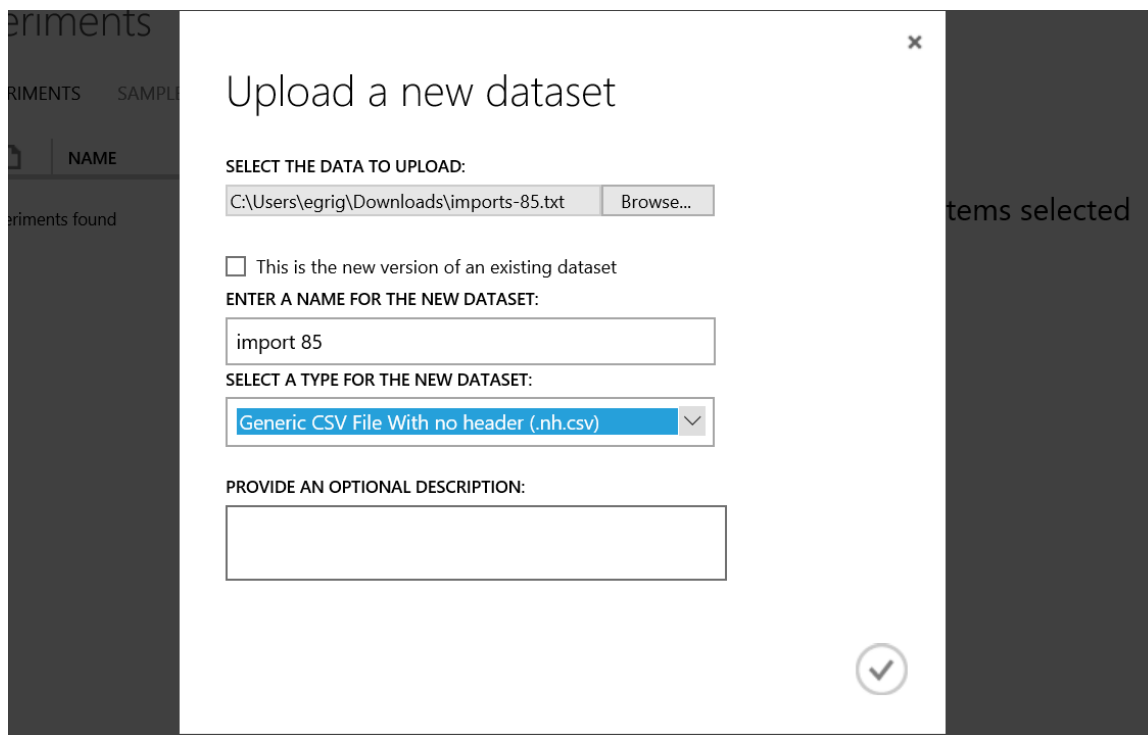


Рисунок 3.3 Процес завантаження файлу з вибіркою

Наступним кроком є створення нового експерименту. Створення нового експерименту відбувається наступним чином: New > Experiment > Blank Experiment. Результатом є створена нова робоча область експеримента із панеллю інструментів, але ще не заповнена даними. Дана робоча область зображена на рисунку 3.4.

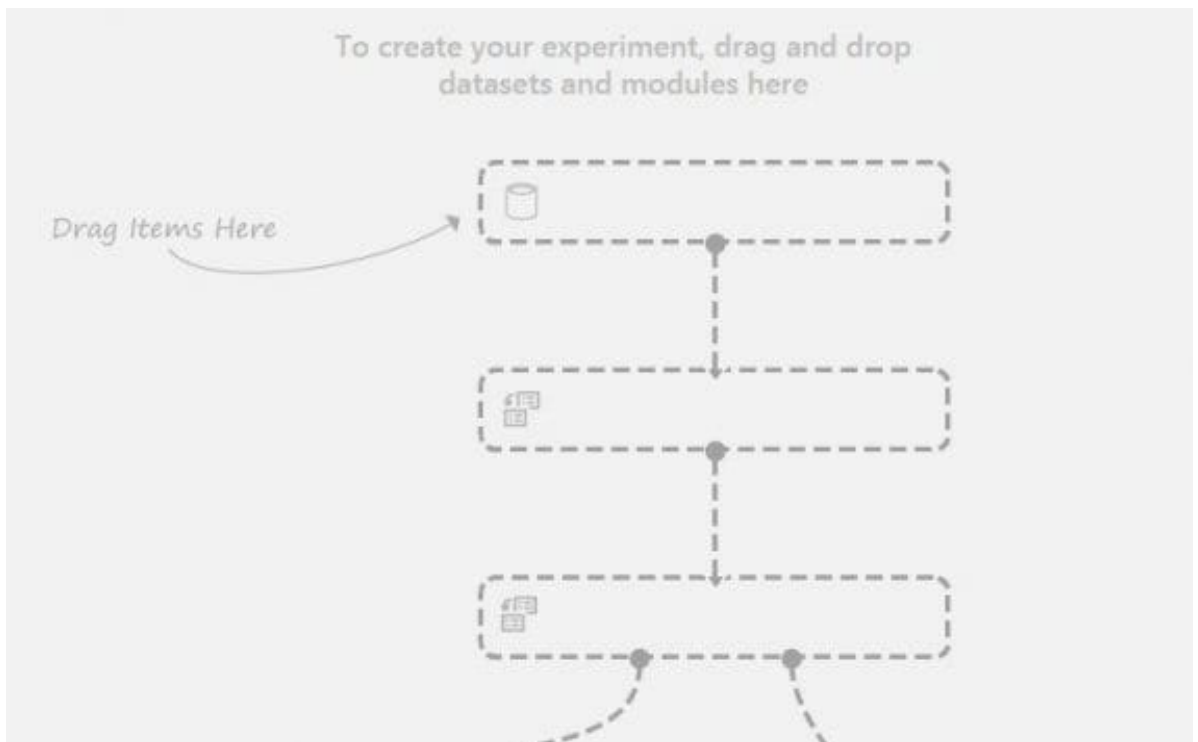


Рисунок 3.4 Створена робоча область

Наступним кроком є підготовка даних з вибірки.

При розробці моделей машинного навчання важливо систематично перевіряти результати попередніх експериментів після кожної зміни. Тому є сенс використовувати опцію «Visualize». Це призведе до відкриття вікна, яке надасть загальне уявлення про дані і їх розподіл.

В полі New column names панелі налаштувань вводяться імена стовпців, що описують дані з вибірки. Вони знаходяться в тому ж файлі, який раніше. Значення полів мають бути наступними: location, rooms, kitchen, service, parking, staff, reservation, maintenance, safety, additional services.

Для демонстрації роботи компонента, необхідно запустити модель за допомогою Run знизу. Після цього відбудеться вихід Metadata Editor. Даний процес зображено на рисунку 3.5.

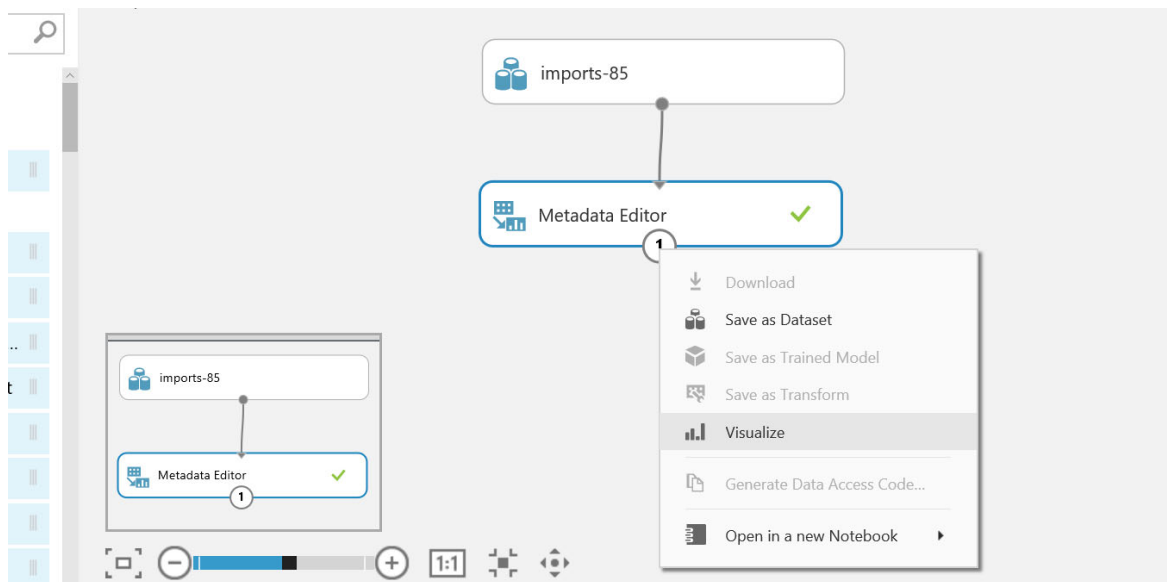


Рисунок 3.5 Процес візуалізації компонента

Наступним кроком є видалення нормалізованих залишків. Для цього відбувається перевід в робочу область Project Columns із тієї ж групи та відбувається з'єднання його із Metadata Editor. Далі обираються усі рядки, окрім normalized-losses. Дані налаштування приведені на рисунку 3.6.

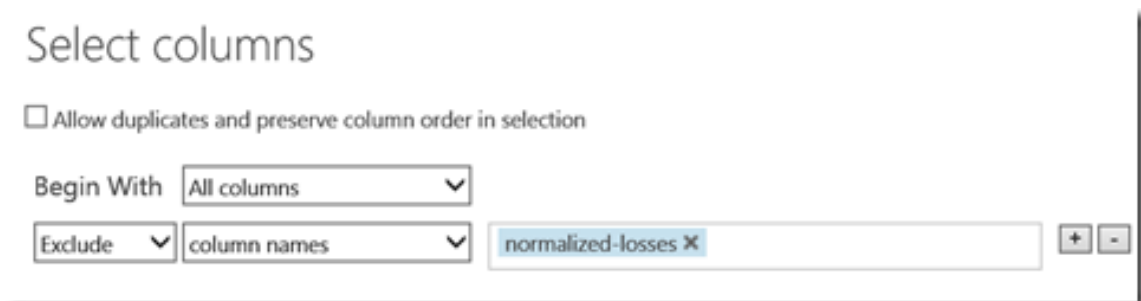


Рисунок 3.6 Процес видалення нормалізованих залишків

Якщо в вибірці є стовпці, у яких немає значення, можна обмежитись відкиданням неповних рядків. Для цього необхідно обрати функцію Missing Value Scrubber і з'єднати його з Project Columns. У полі For missing values змінюється значення на Remove entire row. Після цього необхідно запусити, візуалізуйте і переконайтеся, що рядки із порожніми значеннями зникли з вибірки. Даний процес зображено на рисунку 3.7.

Properties

▲ Missing Values Scrubber

For missing values

Remove entire row



Cols with all MV

KeepColumns



MV indicator column

DoNotGenerate



Рисунок 3.7 Процес відкидання неповних рядків з вибірки

Останнім питанням на етапі підготовки є чи всі характеристики впливають на попит готелю. В цьому етапі можна обмежитись наступним невеликим числом показників, список яких наведено нижче. Надалі завжди буде зберігатись можливість додати нові показники і перевірити гіпотезу про їхню достатність, порівнявши точність отриманих моделей. Нижче вказані перелічені конкретні складові з вибірки даних із відповідними характеристиками кожного з них.

- розташування – близькість транспорту;
- номери – чистота;
- кухня – включення сніданку в рахунок;
- сервіс – обслуговування номеру;
- парковка – автомийка;
- персонал – багатомовність;
- бронювання – зручність;

- технічне обслуговування – наявність інтернету;
- безпека – охорона;
- додаткові послуги – можливість замовити таксі.

Під час аналізу дані характеристики будуть вказуватися як proximity of transportation, cleanliness, breakfast included in the bill, room service, car wash, multilingualism, convenience, Internet availability, security, possibility to order a taxi.

Цей процес показаний на рисунку 3.8.

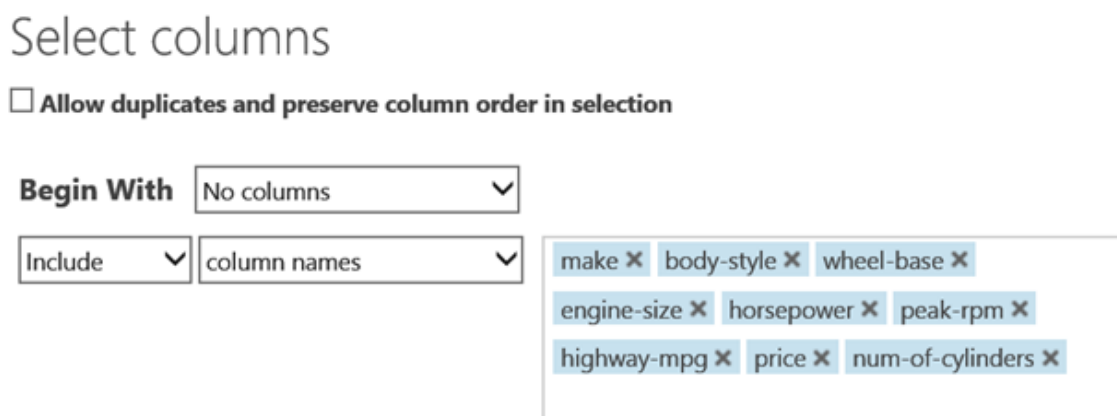


Рисунок 3.8 Перевірка характеристик

Наступним етапом є розбивка вибірки. Тепер дані готові для використання в процесі навчання. Однак у машинному навчанні може виникнути явище, відоме як "перенавчання" - це коли модель навчається даним без їх узагальнення. Це може призвести до недостатньої здатності моделі адекватно передбачати дані, які суттєво відрізняються від тих, на яких вона була навчена. Для управління цією ситуацією вибірку зазвичай розділяють на навчальну та тестову у відношенні приблизно 3:1. Тестова вибірка не бере участь у процесі навчання і використовується після завершення для оцінки точності прогнозування. Якщо помилка на тестовій вибірці суттєво відрізняється від помилки на навчальній, то може виникнути описаний вище ефект перенавчання. Для створення тестової вибірки необхідно перейти на робочу область експерименту і з'єднати із останнім Project Columns компонент Split Data з групи Data Transformation/Sample and Split.

Далі встановити частку рядків на першому виході 0.75 і переконатися, що встановлено позначку Randomize Split.

Наступним етапом є навчання моделі лінійної регресії. По-перше, необхідно розташувати компоненти Linear Regression, Train Model, Score Model та Evaluate Model на панелі інструментів. Train Model є універсальним інструментом, який навчає будь-яку модель на навчальних даних. В випадку даного проекту для конфігурації випадку необхідно підключити перший або лівий вихід Split Data та вихід Linear Regression до відповідних входів Train Model. В налаштуваннях Train Model вказати цільове значення outcome column як price. Тепер модель готова до навчання.

Окрім навчання, важливо оцінити результати. Score Model дозволяє обчислити вихідні значення навченої моделі на будь-якій вибірці та розрахувати прогнозовані результати. Для цього необхідно підключити вихід Train Model, що містить навчену модель, до відповідного входу Score Model, а тестову вибірку з другого виходу Split Data як вибірку даних. Далі з'єднати вихід Score Model з будь-яким із входів Evaluate Model для обчислення числових характеристик якості навчання. Результат даного процесу зображений на рисунку 3.9.

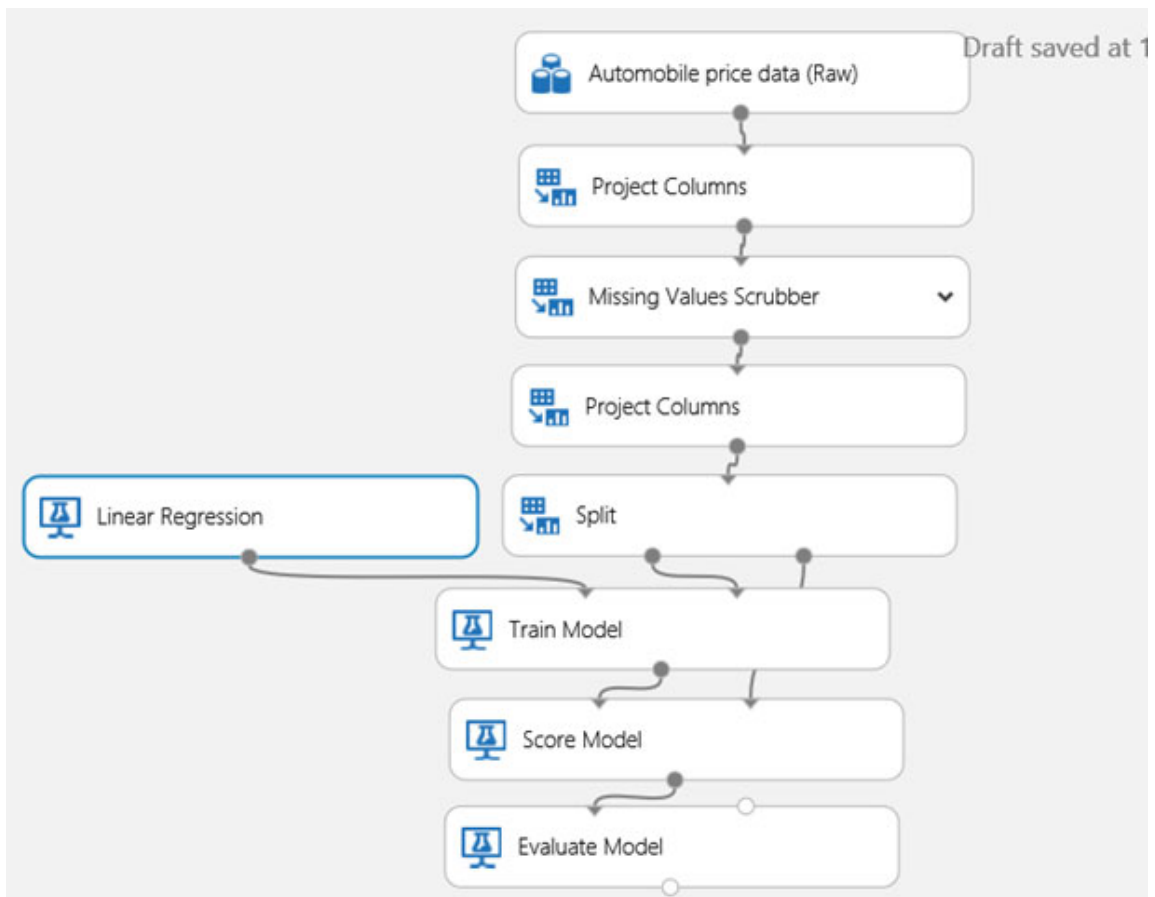


Рисунок 3.9 Навчання моделі лінійної регресії

Наступним кроком є запуск моделі і візуалізація результатів виконання Evaluate Model. Результат запуску зображений на рисунку 3.10.

rows	columns				
1	5	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
view as					Coefficient of Determination
		2462.272003	3516.579251	0.411243	0.183563
					0.816437

Рисунок 3.10 Візуалізація результатів виконання Evaluate Model

Коефіцієнт детермінації вказує на те, наскільки точно лінія регресії відтворює вихідні дані. Його значення коливається від 0 до 1, де 1 відповідає абсолютній точності. У випадку даного проекту коефіцієнт складає 82%. Цей результат може вважатися досить хорошим або не дуже, і це залежить від

постановки завдання та прийнятної толерантності до помилок. У випадку передбачення попиту 82% вважається відмінним результатом. Якщо існує намір покращення даного показнику, існує можливість додати інші стовпці за допомогою Project Columns або випробувати зовсім інший алгоритм, наприклад, Poisson Regression. Це відбувається наступним чином. Відбувається заміна лінійної регресії на Poisson Regression, здійснивши відповідну заміну компонента. Метод паралельного навчання проходить через підключення результату до другого виходу Evaluate Model. Це дозволить зручно порівняти результати навчання обох моделей. Схема порівняння зображена на рисунку 3.11.

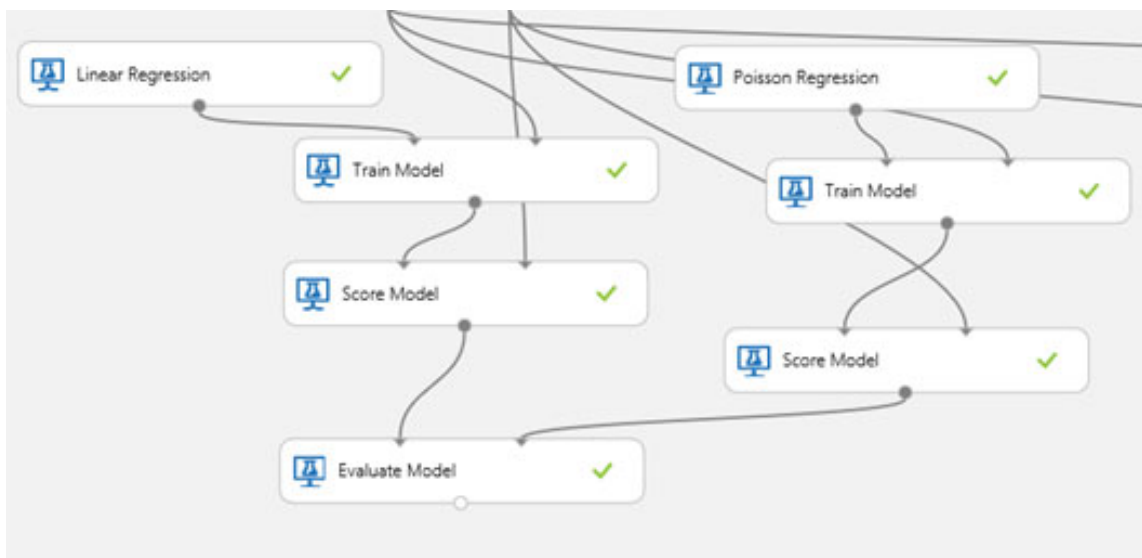


Рисунок 3.11 Схема порівняння навчання обох моделей

Судячи з результатів порівняння, дані значно краще описуються моделлю лінійної регресії. Саме тому існують підстави вибрати саме її як підсумкову.

Metrics		Metrics	
Mean Absolute Error	1762.779258	Mean Absolute Error	2984.241394
Root Mean Squared Error	2792.029358	Root Mean Squared Error	4311.095814
Relative Absolute Error	0.296941	Relative Absolute Error	0.502697
Relative Squared Error	0.137762	Relative Squared Error	0.328446
Coefficient of Determination	0.862238	Coefficient of Determination	0.671554

Рисунок 3.12 Результат порівняння навчання обох моделей

Для візуалізації будь-якого вузлу необхідно перейти до потрібного вузла і обрати його візуалізацію. Результат підрахованої моделі у вигляді графіків та результати проаналізованих даних зображені на рисунках 3.13, 3.14, 3.15 і 3.16 відповідно.

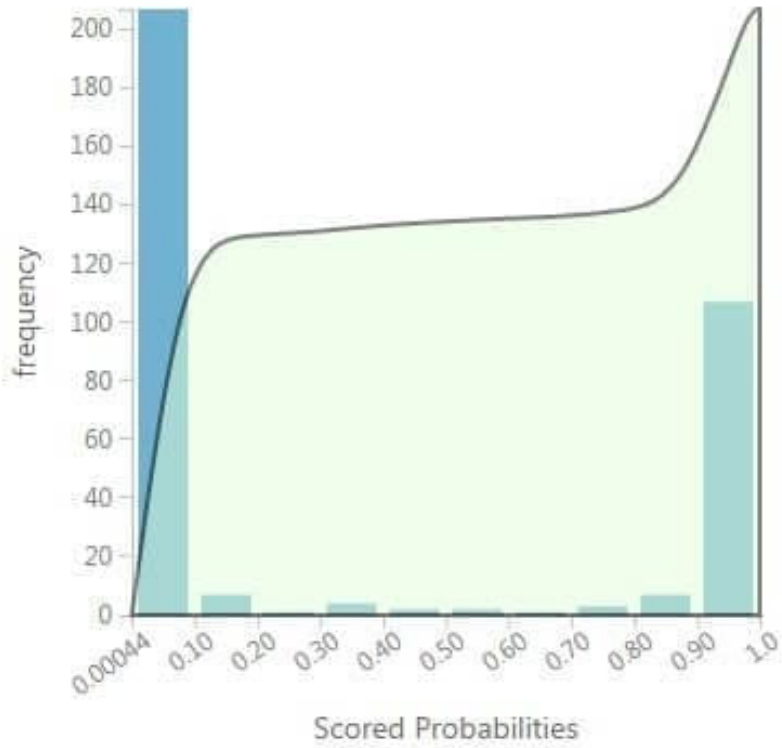


Рисунок 3.13 Підрахована модель у вигляді графіку

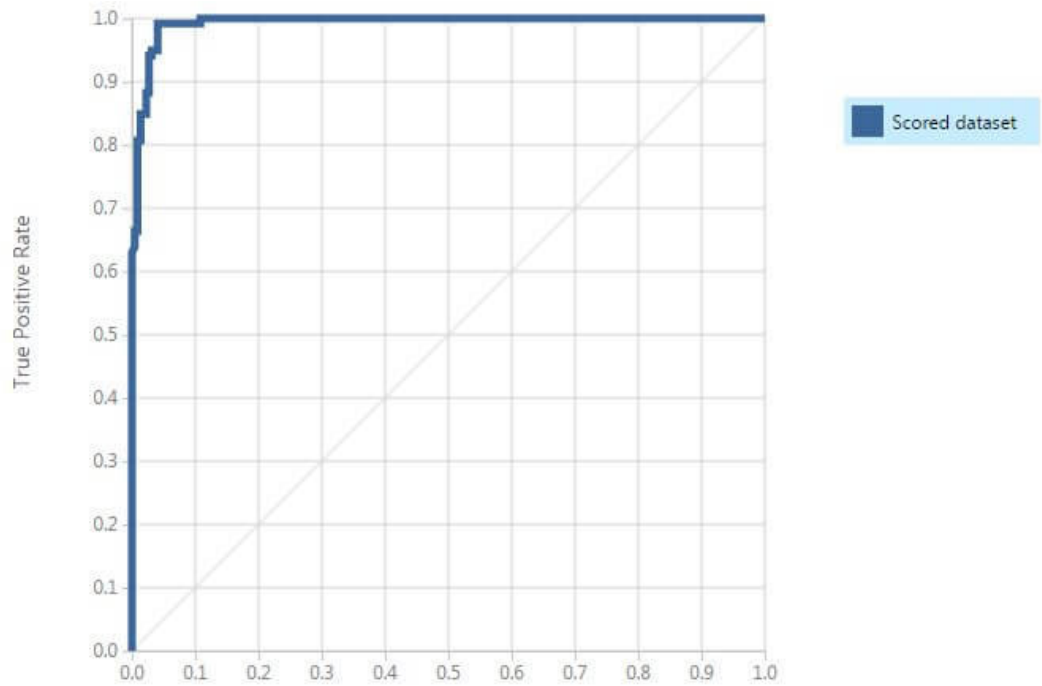


Рисунок 3.14 Підрахована модель у вигляді графіку

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
113	6	0.962	0.942	0.5	0.993
False Positive	True Negative	Recall	F1 Score		
7	215	0.950	0.946		
Positive Label	Negative Label				
1	0				

Рисунок 3.15 Технічні дані аналізу

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	102	5	0.314	0.935	0.903	0.953	0.857	0.927	0.977	0.017
(0.800,0.900]	6	1	0.334	0.950	0.927	0.947	0.908	0.952	0.973	0.021
(0.700,0.800]	3	0	0.343	0.959	0.941	0.949	0.933	0.964	0.973	0.021
(0.600,0.700]	1	0	0.346	0.962	0.945	0.949	0.941	0.969	0.973	0.021
(0.500,0.600]	1	1	0.352	0.962	0.946	0.942	0.950	0.973	0.968	0.025
(0.400,0.500]	0	2	0.358	0.956	0.938	0.926	0.950	0.973	0.959	0.034
(0.300,0.400]	4	0	0.370	0.968	0.955	0.929	0.983	0.991	0.959	0.034
(0.200,0.300]	1	0	0.372	0.971	0.959	0.929	0.992	0.995	0.959	0.034
(0.100,0.200]	0	7	0.393	0.950	0.933	0.881	0.992	0.995	0.928	0.065
(0.000,0.100]	1	206	1.000	0.349	0.517	0.349	1.000	1.000	0.000	0.993

Рисунок 3.16 Результати проведеного аналізу

В ході дослідження було виявлено більшу ефективність моделі лінійної регресії ніж Poisson Regression. Результатом аналізу попиту було виявлена популярність готелю серед клієнтів у 82%. Це доволі високий показник, який створює гарну мотивацію продовжувати розвивати даний бізнес-проект і далі, а також займатись покращенням послуг для подальшого росту.

Azure Machine Learning є високопродуктивний інструмент для роботи з алгоритмами машинного навчання. Можливо, це єдине середовище, яке дозволяє досить легко опублікувати свої алгоритми у вигляді окремого сервісу та подальше їх використання в власних додатках. Однією з головних переваг також є необов'язкові знання вищої математики. У цьому дослідженні використовувався один із найпростіших алгоритмів машинного навчання - алгоритм лінійної регресії. У Azure ML доступні десятки інших алгоритмів, розроблених вченими для різних завдань.

ВИСНОВОК

В ході кваліфікаційної роботи магістра був проведений аналіз сучасного положення інтелектуального аналізу даних як в загальному вигляді, так і при використанні новітніх технологій в різних галузях людської діяльності.

Був проведений аналіз застосування досліджуваних технологій в існуючих системах.

Особлива увага була приділена бізнес-сфері. Був проведений аналіз застосування елементів інтелектуального аналізу даних і бізнес-аналізу.

Також була досліджена область data mining, як в загальному вигляді, так і в сфері її використання в різноманітних бізнес-проектах.

Головною темою дослідження був аналіз попиту в готелі. Для проведення даного дослідження був використаний один з найпопулярніших і найдоступніших інструментів сфери Machine Learning, а саме Azure Machine Learning. Була проаналізована як загальна характеристика додатку, так і конкретні шляхи використання в області бізнесу.

Основна увага приділялася аналізу попиту готелю серед клієнтів. За допомогою Azure Machine Learning було створено дві моделі і порівняні між собою для виявлення більш якісного варіанту аналізу прогнозування попиту на основі вибірки даних.

В результаті порівняння, було виявлено, що дані значно краще описуються моделлю лінійної регресії. Саме тому існують підстави вибрати саме її як підсумкову.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Hotelogix PMS Hotel Software. [Електронний ресурс]. – Режим доступу: <https://www.hotelogix.com/> – Загол. з екрану.
2. Ліщинська Л. Б. Впровадження Business Intelligence для ефективного управління підприємством [Електронний ресурс]. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2021/paper/view/12330/> – Загол. з екрану.
3. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних. [Текст]. – 1-ше видання. – Київ, 2014. 599 с.
4. Чубукова І. А. Data Mining: навчальний посібник [Текст]. – 1-ше видання. – Інтернет-університет інформаційних технологій: БІНОМ: лабораторія знань, 2006. 382 с.
5. Бойко Н. І. Перспективні технології дослідження великих даних у розподілених інформаційних системах. Радіоелектроніка, інформатика, управління [Текст]. – 1-ше видання. – 2017. №4 с. 66-76.
6. Беккауер А. Використання технологій Data Mining для автоматизації бізнес-процесів на виробництві. Системи обробки інформації [Текст]. – 1-ше видання. – 2016. №1 с. 161-164.
7. Taylor D. Almost intelligent systems. How to gain competitive advantages by automating decision making [Текст]. – Cengage Learning, 2009. – 448 с.
8. Ситник В.Ф. Інтелектуальний аналіз даних (дейтамайнінг). [Текст]. – 1-ше видання. – КНЕУ, 2007. 376 с.
9. Microsoft Azure Software. [Електронний ресурс]. – Режим доступу: <https://www.azure.microsoft.com/en-us/> – Загол. з екрану.
10. Clean Architecture: A Craftsman's Guide to Software Structure and Design [Текст] – 1-ше видання. – Pearson, 2017. 430 с.
11. Martin Kleppmann. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems [Текст] – 1-ше видання. – O'Reilly Media, 2017. – 616 с.

12. Chris Fehily. SQL Database Programming [Текст] – 4-те видання. – Questing Vole Press, 2014 – 604 с.
13. Bert Scalzo. Database Benchmarking and Stress Testing: An Evidence-Based Approach to Decisions on Architecture and Technology [Текст]. – Apress, 2018. – 262 с.
14. Бізнес-аналітик. [Електронний ресурс]. – Режим доступу: <https://brainbasket.org/who-are-it-specialists-2/business-analyst/> – Загол. з екрану.