НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
Інститут телекомунікаційних систем
(повна назва інституту/факультету)

Кафедра телекомунікацій
(повна назва кафедри)

«До захисту допущено»

Завідувач кафедри
_____ Явіся В.С._____
(підпис) (ініціали, прізвище)

“ 04 ” ____червня____2020 р.

Дипломна робота

на здобуття ступеня бакалавра

зі спеціальності (спеціалізації) 172 Телекомунікації та радіотехніка
(код та назва спеціальності)

на тему: Використання алгоритмів машинного навчання в телекомунікаційних системах 5G

Виконав: студент IV курсу, групи ТЗ-62

Раченчук Іван Геннадійович_____
(прізвище, ім’я, по батькові) (підпис)

Керівник д.т.н., професор Ільченко М.Ю.
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант ___1,2,3___ __к.т.н, доцент Міночкін Д.А._____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент ___к.т.н, доцент Скулиш М.А._____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Засвідчую, що в цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Київ – 2020 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Інститут телекомунікаційних систем
(повна назва)

Кафедра Телекомунікацій
(повна назва)

Рівень вищої освіти – перший (бакалаврський)

Спеціальність (спеціалізація) 172 Телекомунікації та радіотехніка
(код і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Явіся В.С.___
(підпис) (ініціали, прізвище)
«_22_»____січня___2020 р.

ЗАВДАННЯ
на дипломну роботу студенту
Раченчуку Івану Геннадійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Використання алгоритмів машинного навчання в телекомунікаційних системах 5G

керівник роботи Ільченко Михайло Юхимович, д.т.н., проф.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «_30_»_березня_ 2020 р. №924-с

2. Строк подання студентом роботи 04.06.2020

3. Вихідні дані до роботи: сервіс запуску Jupyter Notebook-ів Google Colab, сторонні бібліотеки Python, засоби SciKit-Learn, моделі класифікації.

4. Зміст дипломної роботи (перелік завдань, які потрібно розробити):
1. Загальний огляд 5G, білінгу та управління абонентами;
2. Аналіз основних методів та алгоритмів машинного навчання;
3. Аналіз моделей прогнозування вибору тарифного плану двох різних класів абонентів

5. Орієнтовний перелік ілюстративного матеріалу: презентація по темі «Використання алгоритмів машинного навчання в телекомунікаційних системах 5G»

Слайд №1 Назва;

Слайд №2 Вступ. Актуальність та мета;

Слайд №3-4 Загальні відомості про 5G, білінг та управління абонентами;

Слайд №5 Загальні відомості про аналіз даних та машинне навчання;

Слайд №6-7 Огляд задач машинного навчання, методів та моделей;

Слайд №8-9 Проведення аналізу моделей;

Слайд №10 Результати аналізу;

Слайд №11 Висновки;

6. Консультанти розділів роботи[1*]

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Доцент Міночкін Д.А.	27.12.2019	26.02.2020
2	Доцент Міночкін Д.А.	26.02.2020	30.04.2020
3	Доцент Міночкін Д.А.	30.04.2020	12.05.2020

7. Дата видачі завдання 1.11.2019

Календарний план

№ з/п	Назва етапів виконання дипломного роботи	Строк виконання етапів роботи	Примітка
1	Опрацювання літературних джерел з теми дипломної роботи	27.12.2019	
2	Аналіз вимог завдання, вибір методів та засобів рішення поставлених завдань	20.01.2020	
3	Дослідження відомостей про системи 5G, білінгу та управління абонентами	26.02.2020	
4	Систематизація інформації про машинне навчання, його методи та моделі	27.03.2020	
5	Отримання вихідних даних для розробки моделі	30.04.2020	
6	Розробка моделі та аналіз результатів	12.05.2020	
7	Оформлення пояснювальної записки	29.05.2020	

Студент _____ І.Г. Раченчук
(підпис) (ініціали, прізвище)

Керівник роботи _____ Ільченко М.Ю.
(підпис) (ініціали, прізвище)

1

РЕФЕРАТ

Робота виконана на 58 сторінках, містить 28 ілюстрацій, 6 таблиць. При підготовці використовувалася література з 9 різних джерел.

З впровадженням телекомунікаційних систем 5G відбудеться початок нової промислової революції та змінить економіку завдяки новим можливостям, що змінять традиційні поняття про телекомунікації. Значно підвищиться продуктивність за рахунок зростання швидкості та зменшення затримки. Оскільки даних буде ще більше, то необхідно якомога швидше їх аналізувати та обробляти. Зокрема для цього використовується машинне навчання, основними перевагами якого є висока швидкість, самостійність та вміння пристосовуватись до змін без значного втручання оператора.

Метою роботи є використання алгоритмів машинного навчання в телекомунікаційних системах 5G.

В роботі розглянуті системи 5G, білінгу та управління абонентами. Також було проаналізовано задачі, методи, алгоритми та моделі машинного навчання. В результаті роботи було проведено аналіз моделей прогнозування вибору тарифного плану для двох класів користувачів з використанням машинного навчання.

Ключові слова: 5G, машинне навчання, Python, білінг, класифікація, SKLearn.

ABSTRACT

Work carried out on 58 pages containing 28 figures, 2 tables. The paper was written with references to 9 different sources.

5G will start the Fourth Industrial Revolution and change the economy because of new capabilities. It will change traditional concept of telecommunications. An efficiency will be also greatly increased by higher speeds and lower latency. Data will be greatly increased too and that's why machine learning exists. The data should be analyzed and processed in a short time period. The main advantages of machine learning are high speed, self-dependence and ability to adapt to changes without significant operator intervention.

The purpose of the bachelor's thesis is to describe using of machine learning algorithms for the 5G telecommunication systems.

The work discusses 5G systems, billing and subscribers management. Machine learning tasks, methods, algorithms and models were analyzed too. As a result, the prediction models analysis of subscribers chosen plan was carried out for two classes with machine learning.

Keywords: 5G, machine learning, Python, billing, classification, SKLearn.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ І ТЕРМІНІВ

API – Application programming interfaces, прикладний програмний інтерфейс.

IoT – Internet of Things, Інтернет речей.

ML – Machine Learning, машинне навчання.

GPRS – General Packet Radio Service, загальний сервіс пакетної радіопередачі.

EDGE – Enhanced Data Rates for GSM Evolution, технологія передачі даних, що забезпечує передачу інформації в мережі мобільного зв’язку.

HSPA – High Speed Packet Access, протокол високошвидкісного пакетного доступу.

VoLTE – Voice over LTE, передача голосу по мережі LTE.

EE – британський оператор мобільного зв’язку та інтернет-провайдер

BSS - Business support system, система підтримки бізнесу, компоненти для проведення операцій з клієнтами.

AI – Artificial Intelligence, штучний інтелект.

DL – Deep Learning, глибинне навчання.

CLIQUE – Clustering In QUEst, модель кластеризації на основі сітки.

ВСТУП

Легко зрозуміти, чому технології бездротового зв'язку стають все більш важливими: за рахунок полегшення розширеного доступу до інформації в дорозі, досягнення в області бездротового зв'язку сприяли поліпшенню бізнесу, освіти і технологій в усьому світі. У міру поліпшення мобільного зв'язку покращуються і зв'язки між людьми.

5G викличе четверту промислову революцію і змінить економіку. Через швидкостей і потужностей, які обіцяє принести 5G мережу, у неї є потенціал бути незамінною технологією. 5G важливе не тільки тому, що воно здатне підтримувати мільйони пристроїв на надшвидких швидкостях, але й тому, що воно здатне трансформувати життя людей у всьому світі.

Машинне навчання також стане ще більш вживаним при розвертанні мереж 5G, оскільки воно вже використовується в багатьох галузях, а швидкості нового покоління дадуть змогу розширити межі уяви про його використання.

Ці технології мають величезний потенціал, оскільки вирішують проблему обробки та аналізу великих об’ємів даних за короткий час та без значного втручання людини. Актуальною проблемою при впровадженні машинного навчання є вибір методу та алгоритму машинного навчання, що може вплинути на отримання тих чи інших результатів. Тому необхідно проводити налаштування відповідно до поставлених вимог.

Метою дипломної роботи є проведення аналізу використання алгоритмів машинного навчання в телекомунікаційних системах 5G. Для досягнення поставленої мети необхідно вирішити наступні задачі:

1. Провести аналіз моделей машинного навчання для бінарної класифікації.
2. На основі отриманих даних сформувати рекомендації щодо вибору моделі класифікації.

Об’єктом дослідження є надання вірних прогнозів щодо класифікування, а предметом – методи, алгоритми машинного навчання та моделі класифікації.

РОЗДІЛ 1. 5G, БІЛІНГ ТА УПРАВЛІННЯ АБОНЕНТАМИ

1.1 5G

5G – наступне покоління технології мобільної мережі. Це поліпшить роботу з мобільними пристроями і допоможе насолоджуватися більш високою швидкістю роботи в Інтернет, більш надійним підключенням до даних в місцях з високим навантаженням і практично миттєвим підключенням при відкритті додатків і веб-сайтів, грі в ігри або підключенні до ваших розумних домашніх пристроїв.

Але 5G не є заміною 4G. Це просто додає ще один рівень в мережу, щоб надати вам більш швидкий, плавний і найкращий мобільний досвід. Еволюція систем мобільного зв'язку до 5G вимагає впровадження нових технологій, що підвищують швидкість передачі і абонентську ємність, а також значно зменшують затримку в каналі.

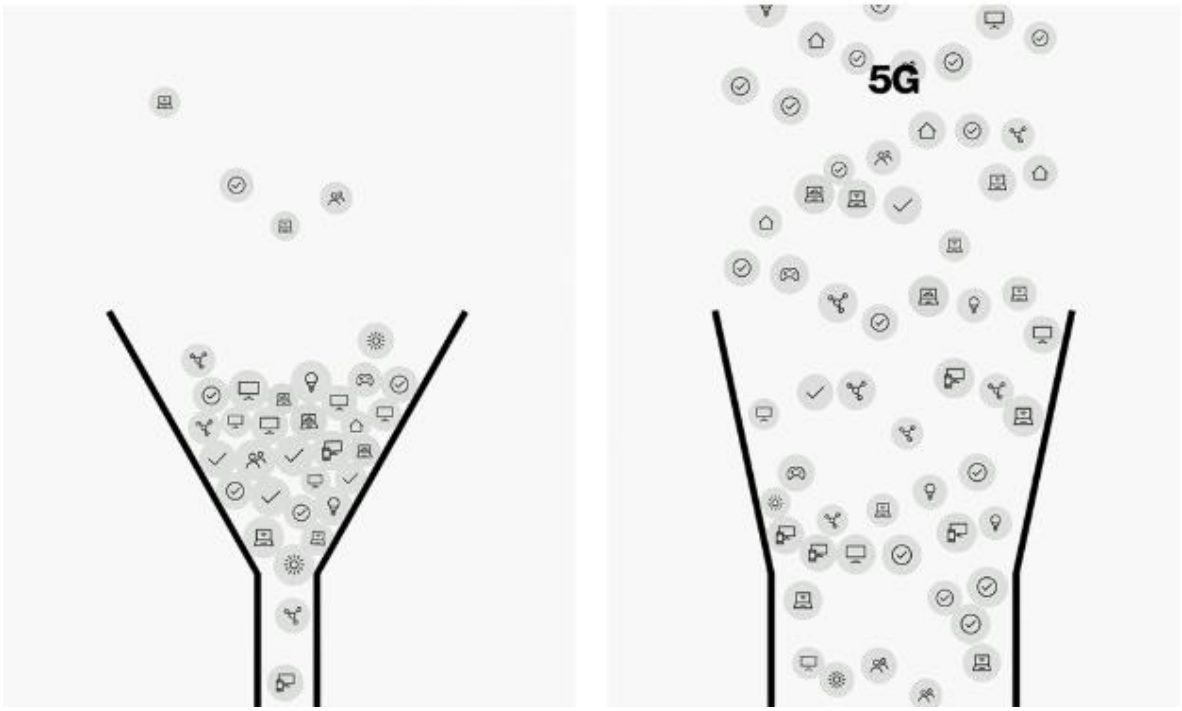


Рис. 1.1 Наглядне зображення, що дає уявлення про суть 5G

Необхідність переходу до 5G полягає в тому, що станом на 2019 рік витрати на трафік по мережам операторів зв’язку не покривається доходами від традиційних послуг. Пошук нових послуг не надає очікуваних результатів.

Також основний зріст трафіку та доходів відбувається саме в секторі пристроїв IoT (Internet of Things), що являє собою одну з основних цілей функціоналу 5G.

Мета створення та призначення мереж 5G. Мережі мобільного зв'язку минулих поколінь мали наступні призначення та функціонал:

- 1G: Послуги передачі мовлення по аналоговій мережі
- 2G: Послуги передачі мовлення по цифровій мережі, низькошвидкісні послуги передачі даних (GPRS, EDGE)
- 3G: Високошвидкісні послуги передачі даних (HSPA) з можливістю передачі голосу по мережі IP, мобільний доступ до Інтернет MBB (Mobile Broadband)
- 4G: Мобільний широкосмуговий доступ MBB на базі LTE, LTE-A, передача голосу (VoLTE)

Мережі 5G значно розширять обмежений функціонал мобільних мереж попередніх поколінь. Основними функціональними особливостями мереж 5G наступні:

- удосконалений мобільний широкосмуговий доступ eMBB enhanced MBB
- Найнадійніші комунікації з низькою затримкою ULLRC (Ultra Low Latency Reliable Communication)
- Масивні міжмашинні комунікації Massive IoT. IIoT, mMTC (massive machine Type Communication)

На основі цих трьох видів функціоналу будується вся різноманітність послуг та можливостей мереж 5G.

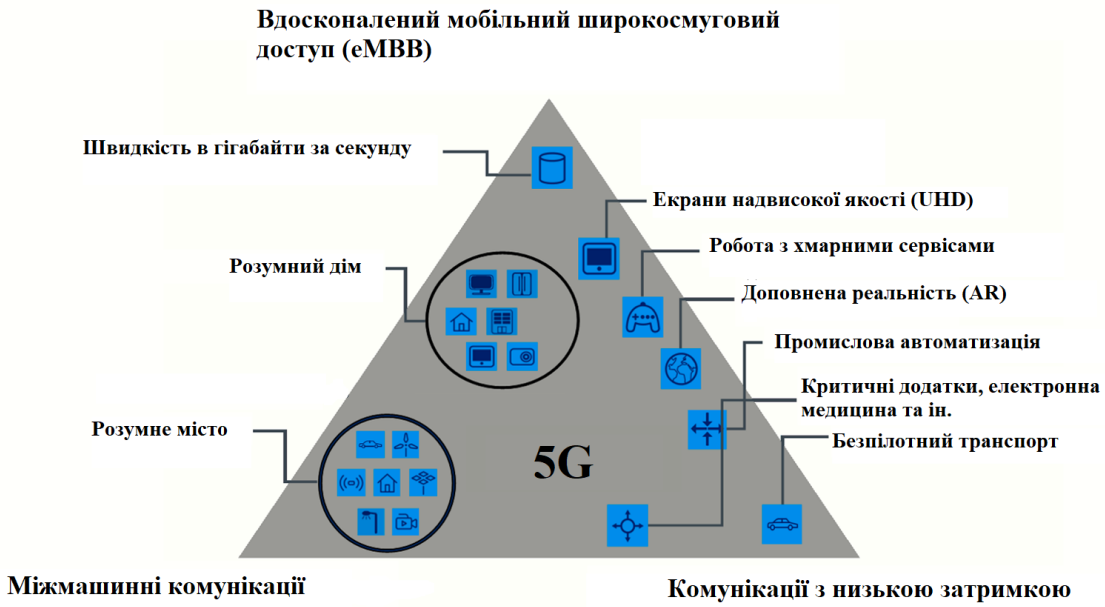


Рис.1.2 Основні види функціоналу, можливості та послуги мереж 5G

Стандарт 5G консорціуму 3GPP визначає наступні характеристики систем нового покоління:

- швидкість передачі не менше 10 Гбіт / с;
- одночасне підключення до 100 млн пристроїв / км2;
- затримка в радіоканалі не більше 1 мс.

Для забезпечення заданих характеристик застосовується смуга частот, що повинна мати безперервну ширину спектра 500-1000 МГц і займати сантиметровий і чи міліметровий діапазон електромагнітного спектра. В даний час розглядається можливість виділення мереж 5G мереж діапазонів 24,25 -27,5 ГГц і 37-43,5 ГГц для досягнення пропускної здатності 20 Гбіт / с, хоча застосування спектру частот нижче 6 ГГц вже погоджено Всесвітньою конференцією радіозв'язку (ВКР), оскільки він забезпечує хороше радіопокриття без високих інвестицій. Це означає меншу кількість базових станцій, краще проникнення крізь перешкоди і є найкращим варіантом для додатків, які потребують надвисоких швидкостей.

Для передачі експоненціально більше даних в бездротовій мережі, для більшої швидкості, зменшення заторів та меншої затримки, яка є затримкою ще до

початку передачі даних після вказівки, було розроблено технології, що надалі були об'єднані під загальною назвою “нове радіо 5G”, 5G New Radio (5G NR).

Цей новий інтерфейс, що використовує міліметровий спектр хвиль, дозволяє використовувати більше пристроїв у межах однієї географічної області (4G підтримує близько 4000 пристроїв на один квадратний кілометр, а 5G - близько мільйона). Отже, більше потокової передачі, голосових дзвінків та стримінгових сервісів здійснюватиметься безперебійно.

Теодор Сайзер, віце-президент по розробці бездротових технологій в Bell Labs відмічав, що в мережах 5G будуть працювати безліч різних пристроїв. Смартфони і планшети нікуди не дінуться, але окрім них в мережі з’явиться цілий “зоопарк” різних пристроїв, як камери відеоспостереження, погодні датчики, датчики “розумних” електричних мереж, “розумних” домів та автомобілів.

“Число пристроїв, що взаємодіють з інтернетом та між собою, постійно зростає. Необхідні більш удосконалені мережі, здатні забезпечити цю взаємодію максимально ефективним чином. Мережі нового покоління відкривають нові можливості в багатьох областях - від підвищення ефективності виробничих процесів, підвищення безпеки на дорогах та у місті в цілому до поліпшення комунальних сервісів та більш чистого навколишнього середовища,” - відмічав Джон Хілі, представник підрозділу Communication & Storage Infrastructure Group в Intel.

Також в 5G використовується ще одна нова цифрова технологія під назвою Massive MIMO, що розширюється для множинних вхідних сигналів з декількома вихідними сигналами. Massive MIMO використовує декілька націлених променів для прожектора та слідкуючих за користувачами навколо сайту стільникового зв'язку, покращуючи охоплення, швидкість та потужність. Сучасні мережеві технології функціонують як прожектори, висвітлюючи область, але з великою витратою світла / сигналу. Частина впровадження 5G включає встановлення Massive MIMO та 5G New Radio на всі базові станції мобільної мережі поверх існуючої інфраструктури 4G.

Таблиця 1.1 Порівняння 3G, 4G та 5G

	3G	4G	5G
Пропусна здатність	2 Мбіт / с	200 Мбіт / с	> 1 Гбт / с
Затримка	100-500 мілісекунд	20-30 мілісекунд	< 10 мілісекунд
Середня швидкість	144 Кбіт / с	25 Мбіт / с	200-400 Мбіт / с
Рік розгортання	2004-2005	2006-2010	2020+

За словами EE, 5G обіцяє швидкості пристроїв в 10-20п разів більше, ніж 4G, тобто миттєве завантаження двогодинного фільму ультрависокої роздільної здатності, відеодзвінки в 4К та багато іншого - стануть буденністю.

Швидкість та надійність мережі в режимі реального часу зможе дозволити галузям перевести машини від кабельного підключення до мережевого бездротового з'єднання. Це може значно підвищити продуктивність та зменшити витрати.

В проектах “розумних міст” 5G дозволить в режимі реального часу передавати інформацію з більшого числа сенсорів на різноманітних об'єктах. Старший директор Qualcomm з продуктового менеджменту мобільних технологій Санджив Аталі відмічає, що можна буде розгорнути тисячу сенсорів замість сотні, для обслуговування яких буде достатньо меншої кількості базових станцій, ніж при існуючих нині мережах. Це можуть бути як сенсори “розумного освітлення” чи сенсори звуку, котрі будуть встановлені в цілях безпеки, так і сенсори моніторингу стану об'єктів ЖКГ.

Нові сервіси з використанням 5G можуть бути реалізовані і в медицині. Так, наприклад, для організації віддаленого моніторингу стану пацієнтів. Лікар зможе

оперативно отримувати інформацію зі спеціальних сенсорів та слідкувати за станом пацієнтів цілодобово.

Завдяки дуже низькій затримці передача даних 5G також відкриє більше можливостей для віддаленого проведення операцій з використанням робота. Такий сервіс особливо актуальний для невеликих населених пунктів, де відсутні хірурги: керуючи роботом, спеціаліст може провести операцію, перебуваючи в зовсім іншому місці. Саме за рахунок 5G такий сервіс можливо буде розгорнути в бездротових мережах.

У великих виробничих компаніях, в логістиці 5G надасть можливість використовувати більше промислових роботів, виконуючих різноманітні функції замість людей, а також дронів. Останні вже зараз використовуються на деяких виробництвах, але частіше за все керуються з використанням мереж Wi-Fi. 5G дозволить охопити більшу дистанцію, ніж мережі Wi-Fi, а завдяки низькій затримці - підвищить стабільність роботи таких систем. Проект розгортання доставки товарів за допомогою дронів є, наприклад, у Amazon.

Ще один приклад - сервіс моніторингу транспорту в компаніях. Санджив Аталі з Qualcomm вважає, що з появою мереж нового покоління оператори, виступаючі провайдерами такого сервісу, зможуть знизити його ціну. Це стане можливим за рахунок того, що ціна однієї базової станції 5G буде нижче ціни станцій для існуючих мереж, а також за рахунок того, що одна базова станція зможе одночасно обслуговувати більшу кількість пристроїв, тобто для сервісу потрібно менше базових станцій.

1.2 Білінг та управління абонентами

Білінг (англ. billing) - автоматизована система обліку наданих послуг, їх тарифікації і виставлення рахунків для оплати. Також іменується “Автоматизована Система Розрахунків” (АСР). Тарифікація являє собою збір, обробку, передачу та оцінку інформації, що стосується тарифіційованому обєкту та виявляє об'єм

користування для виставлення рахунку абоненту. Цей процес також включає в себе отримання та запис платежів від абонентів.

Мета системи тарифікації в підрахунку об'єму інформації, переданої абонентом, та передача відомостей в білінг. Така система може здатися дуже простою, до неї ледве можливо пред'явити інші вимоги, крім продуктивності і точності обчислень. При цьому оператори, що надають абонентам широкосмуговий доступ в Інтернет, пред'являють різні вимоги в сфері тарифної політики, завдяки чому система тарифікації часто стає одним з найбільш складних та функціонально насичених елементів мережі.

Оператори мобільного зв’язку надаються послуги двома способами:

- передплата - тарифи, які щомісячно стягуються з клієнтів за надану послугу. Наприклад, плата за телефон буде становити 200 грн. незалежно від того, користуються ним чи ні.
- сплата за користування - це витрати, що беруться від клієнтів на основі використання послуг. Наприклад, з абонента буде стягнено плату за всі дзвінки, що були здійснені, або дані, завантажені за допомогою телефону.

Окрім передплат та сплат за використання, оператори можуть стягувати плату за ініціювання послуги, встановлення, призупинення або припинення послуги.

На ринку програмного забезпечення існують найсучасніші системи виставлення рахунків (білінгу), які дуже ефективно виконують завдання та надають велику гнучкість постачальникам послуг пропонувати свої послуги з різною ціновою структурою.

Системи білінгу часто розглядаються як дебіторська заборгованість, оскільки система виставлення рахунків сприяє збору (отриманню) коштів від абонентів. Платіжні системи також є частиною кредиторської заборгованості (для розрахунку між операторами), оскільки абоненти часто користуються послугами інших компаній, такими як роумінг, міжміські дзвінки через інші мережі.

Системи білінгу - висококласні, надійні та дорогі програмні засоби, які забезпечують різні функціональні можливості. Так, рейтингова система та білінг включають в себе рейтинг використання продуктів чи послуг та створення щомісячних рахунків. Обробка платежів передбачає проведення платежів абонента на його рахунок. Кредитний контроль та інкасація передбачають переслідування непогашених платежів та вживання відповідних дій для збору платежів. Спори та коригування передбачають запис спорів абонента проти рахунків та створення коригування для повернення суми щодо врегулювання суперечок. Передплата та післяплата передбачають підтримку як передплачених, так і послуг, що сплачуються по мірі їх використання.

Підтримка багатьох мов та валют також необхідна, якщо компанія розповсюджена по всьому світі та має багатонаціональних клієнтів, або якщо цього вимагають урядові постанови. Продукти та послуги передбачають забезпечення гнучких способів обслуговування різних товарів та послуг на їх продаж окремо або в пакетах. Існують і дисконтні програми - для залучення та збільшення кількості нових клієнтів, що передбачає визначення різних схем знижок.

Джонатан Кафтзан, глава відділу маркетингу в Amdocs, відмічає, що системи підтримки бізнесу (BSS) підтримують телекомунікаційні компанії протягом багатьох років. Такі системи. як білінг, CRM та система заказів, все ще необхідні провайдерам, але вони повинні удосконалюватись. Від білінгу до конвергентної тарифікації в режимі реального часу, від CRM та системи заказів до справжніх одноканальних цифрових систем взаємодії з абонентами. Ця еволюція підтримую трансформацію провайдера, надаючи абонентам змогу додавати та адаптувати комунікаційні послуги із всіх областей телефонного зв'язку, “в дорозі” та в режимі реального часу, та вони дозволяють абонентам взаємодіяти зі своїм провайдером по будь-якому каналу за допомогою будь-якого способу зв'язку. Ця еволюція системи підтримки бізнесу має рішуче значення для задоволення потреб, розширенні прав і можливостей, що потребують абоненти

Сфера телекомунікацій вже багато років рухається до конвергентного білінгу. Проблема криється в тому, що по мірі руху телекомунікацій до цієї цілі, по мірі розвитку потреб економіки, продовжують розвиватись визначення того, що означає конвергенція в контексті справляння плати та виставлення рахунків. Це почалось як засіб уніфікації систем передплати та післяплати. Потім система білінгу стала в конвергенцію напрямків, а тепер - розвивається і поширюється на нові цифрові послуги, такі як інтернет речей та OTT (англ. Over-The-Top), в режимі реального часу для всіх клієнтів. Але не можна сказати, що система тарифікації та білінгу дійсно конвергентна, якщо вона не може обробляти всі нові послуги в єдиній системі.

За словами Amdocs, головною вимогою галузі є можливість управління всіма доходами та абонентськими процесами в єдиній системі, а не створення системи для кожної нової послуги.

Беручи до уваги, що інтернет речей був вперше задуманий близько 30 років тому, почали впроваджувати його лише останні кілька років. Зараз, коли наступила ера 5G мереж мобільного зв’язку, кількість можливостей зростає все більше. Для виробників пристроїв, а також компаній, які забезпечують мережеве підключення для всіх цих пристроїв. Для отримання максимальної вигоди з нових технологічних досягнень, провайдерам необхідно переглянути свій підхід щодо монетизації.

За словами Ericsson, розвиток 5G мережі відкрило безліч нових можливостей для постачальників телекомунікаційних послуг (CSP) в таких галузях, як промислова автоматизація, безпека, охорона здоров'я та машинобудування. Для того, щоб використовувати їх, CSP повинні мати системи підтримки бізнесу (BSS), котрі були розроблені для управління складними ланцюгами створення цін та підтримки нових бізнес-моделей. Для роботи з величезною кількістю пристроїв через відкриті інтерфейси потрібні оптимізовані інформаційні моделі та високий ступінь автоматизації.

Таблиця 1.2 Опис основних можливостей кожного рівня BSS

Рівень BSS	Можливості
5G включно	● Підтримка 5G інтерфейса на основі послуг (SBI) (функція тарифікації) ● Підтримка розподілу мережі в BSS та OSS ● Роумінг партнери ● Контейнеризація та мікро послуги ● Загальний технологічний стек
Інтернет речей та його монетизація	● Управління ідентифікатором та кореляції ● Управління життєвим циклом пристроїв Інтернету речей ● Управління бізнес-клієнтами та 5G/IoT підприємствами ● Тарифікація в багаторівневих ієрархіях ● Договір поставки ● Гнучка організація процесів заказів ● Сервіс для управління пристроями ● Відкритий вплив API ● Неперервна інтеграція/неперервна поставка (CI/CD) для впливу на послуги ● Приватні мережі ● Партнерські платформи ● Контракти на послуги, не пов'язані з телефонним зв'язком (IoT/edge enabled) ● Моделі сплати за послуги, не пов'язані з телефонним зв'язком ● Нарахування та виставлення рахунків ● Послуги щодо місцезнаходження ● Блокчейн для розумного

	заключення контрактів
Повна екосистема 5G	● Управління відносинами партнера ● Каталог партнера ● Розподіл доходів партнера ● Узгодження та регулювання ● Гнучке виставлення рахунків ● Платформа як послуга та роздільні хмарні сервіси ● Послуги edge-платформи ● BSS як послуга ● Неперервний контроль ● Штучний інтелект і автоматизація машинного навчання ● CI/CD

Один за одним ці можливості доповнюють одне одного, постійно підвищуючи якість BSS та перетворюючи BSS в систему, здатну підтримати всі нові сценарії використання та бізнес-моделі, що характеризують 5G/IoT екосистему.

Перший етап еволюції - “5G включно” в таблиці 1.2 - забезпечує підтримку нових стандартів та концепцій 5G, що дозволяє різко підвищити пропускну здатність передачі даних при збереженні акценту на традиційних користувачах. Використання контейнеризації та загального технологічного стеку забезпечить масштабність рішення BSS для задоволення підвищення вимог щодо пропускної здатності мережі.

На наступному етапі рішення - Інтернет речей та монетизація - основна увага приділяється абонентам. Ці нові можливості дозволяють CSP забезпечувати розширену підтримку для підприємств, коли мова заходить про сценарії використання 5G та Інтернету речей, за рахунок управління пристроями Інтернету речей, підтримки тарифікації послуг, не пов'язаних з телефонною компанією, та багатостороння тарифікація, а також монетизація Інтернету речей. Крім того, вплив послуг дозволяє підприємствам самостійно обслуговуватись, а також розробляти

додатки для оптимізації пристроїв Інтернету речей. Кількість випадків використання 5G/IoT значно зростає на цьому етапі.

Додавання можливостей партнерів на рівні всієї екосистеми 5G дозволяє CSP розглянути зовсім нові сегменти користувачів, виходячи за рамки телекомунікацій, та надавати галузеві рішення для систем. CSP може створювати нові послуги (навіть надавати BSS як послугу) та пропонувати ці послуги на ринку для охвату нових сегментів користувачів. Безліч партнерств потребують підтримки нових бізнес-моделей, котрі дозволяють гнучко стягувати плату, розподіляти доходи та виставляти рахунки.

Висновок до розділу 1

За останні кілька десятиліть ми звикли до лінійного прогресу швидкості бездротового зв'язку. В 2G мова йшла про голос та повідомлення, 3G додав дані та мобільний інтернет, в той час як 4G значив відеотрансляції та вибух економіки додатків. Але 5G відрізняється. Це втягує нас в епоху продвинутого машинного навчання, змішаних віртуальних та фізичних реалій, а також повного мережевого підключення. Це буде відчуватись в різних галузях, регіонах та середовищах.

5G забезпечить швидкість передачі даних в 10-100 разів більше, ніж 4G. Забезпечить значно меншу затримку, надаючи можливість використання додатків у реальному часі, а також вбудовані захист та надійність. Це означає, що технології, що залежать від використання нескінченного потоку даних (наприклад, безпілотні автомобілі) будуть мати більш широкий та швидкий канал, через який будуть передаватись дані. Будинки стануть розумніше, лікарні зможуть надавати більш інтелектуальну допомогу, Інтернет речей та ін. - наслідки 5G масштабні.

РОЗДІЛ 2. АНАЛІЗ МЕТОДІВ ТА АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

2.1 Загальні відомості

Машинне навчання (англ. machine learning) - це метод аналізу даних, що автоматизує будування аналітичних моделей. Це відгілля від штучного інтелекту, побудоване на ідеї, що системи можуть вчитись на даних, ідентифікувати та приймати рішення з мінімальним втручанням людини.

Машинне навчання - це також наука змусити комп’ютери працювати, не будучи запрограмованими. Машинне навчання настільки розповсюджене, що ми навіть не знаємо, що використовуємо його в тих чи інших задачах. Багато дослідників вважають, що це кращий спосіб досягти прогресу в дорозі до штучного інтелекту на рівні з людиною.

Оскільки об'єм даних зростає кожного дня, то і проаналізувати їх з високою швидкістю та точністю неможливо. Більше 80% даних неструктуровані - це аудіо, відео, фотографії, документи, графіки та ін. Знайти закономірності в даних таких об'ємів дуже складно для мозку людини. Масивні дані, час, витрачений на обчислення нескінченно збільшується - саме тут вступає в дію машинне навчання, допомагаючи людям з обробкою значних даних за мінімальний період часу. Використовуючи штучний інтелект, люди хотіли побудувати більш якісні та інтелектуальні машини. Якщо поглянути на це зі сторони - дуже схоже наче дитина вчиться сама в себе. Так в машинному навчанні була розроблена нова можливість для комп'ютерів. І вже зараз машинне навчання існує в багатьох сегментах технологій, що ми навіть не розуміємо цього при його використанні.

Машинне навчання всюди - від медицинської діагностики на основі розпізнавання зображень до навігації безпілотних автомобілів. ML еволюціонує як дисципліна до такого ступеня, що в даний час дозволяє бездротовим мережам вчитись та отримувати знання, взаємодіючи з даними. Попередній інтерес та

обговорення доцільності розвитку 5G стандартів за допомогою протоколів ML привернули увагу інженерів та дослідників зі всього світу.

Ми стали свідками того, як мобільні та бездротові системи стали важливою частиною соціальної інфраструктури, мобілізуючи наше повсякденне життя та полегшуючи цифрову економіку різними способами. Однак, ML та бездротова мережа 5G сприймаються як різні області дослідження, незважаючи на потенціал, котрий вони можуть мати, коли використовуються разом. Фактично, вплив мобільного та бездротового мережевого зв'язку з підтримкою ML вже проявився в послугах на основі місцезнаходження, мобільного периферійного кешування, аналітиці Big Data та управлінні мережевим трафіком.

ML чудово підходить для складних проблем, коли значні рішення потребують ручного налаштування, і для проблем, вирішення яких взагалі відсутнє при використанні традиційних методів. Ці проблеми можливо вирішити вивчивши дані, змінивши звичайне програмне забезпечення, що містить довгі списки правил, підпрограмами ML, котрі автоматично навчаються на попередніх даних.

Важливим відмінністю ML від традиційних когнітивних алгоритмів є автоматичний витяг функцій. Завдяки цьому можна відмовитись від дорогих розробок функцій вручну. ML може знаходити аномалії, прогнозувати майбутні сценарії, адаптуватися до змін середовища, надавати представлення про складні проблеми з великими об'ємами даних та виявляти закономірності, котрі людина може пропустити.

Задачі, що стосуються машинного навчання, пропонують принципово оперативного визначення, а не тільки визначення поля в когнітивних термінах. Це слідує зі слів Алана Тьюринга в його статті “Обчислювальна техніка та інтелект”, в якій питання “Чи можуть машини думати?” замінено на “Чи можуть машини робити те, що і люди?”. В області аналізу даних машинне навчання використовується для розробки складних моделей та алгоритмів, котрі піддаються прогнозуванню. Що стосується комерційної сфери - це називають прогнозуючою аналітикою.

2.2 Задачі машинного навчання

Корінь проблеми полягає у виявленні шаблонів, котрі також повинні бути виявлені на основі мережевої діяльності 4G в якості джерела даних для виявлення поведінки мережі як системи та користувачів як компонентів, з можливістю власного вибору, котрий міг би змінити потреби в послугах, заснованих на їх діяльності. При обміні інформацією між мережею та пристроями користувачів можна отримати надзвичайно великий об'єм даних. Такий сценарій показує джерело різного роду інформації, яку можна об'єднати для отримання результатів.

В аналізованій системі необхідно брати до уваги безліч змінних, котрі можна визначити та реєструвати з журналів, виконуючих процеси реєстрації користувачів, викликів, передачі, призначення IP, потоку даних та ін. Кореляція між безліччю змінних, котрі знаходяться під управлінням в межах одного аналізу, не може бути виконана без використання обчислювальної допомоги. Процес вивчення нового сховища даних є новим завданням, оскільки на початку процесу немає чіткого розуміння критеріїв пошуку. Аналіз за часовою шкалою виявляє зміни в системі та надає представлення про діяльність в мережі на основі змін значень кожної змінної. При даному сценарії є необхідною підтримка алгоритмів машинного навчання для виявлення та ідентифікації шаблонів в мережі, підлеглих аналізу.

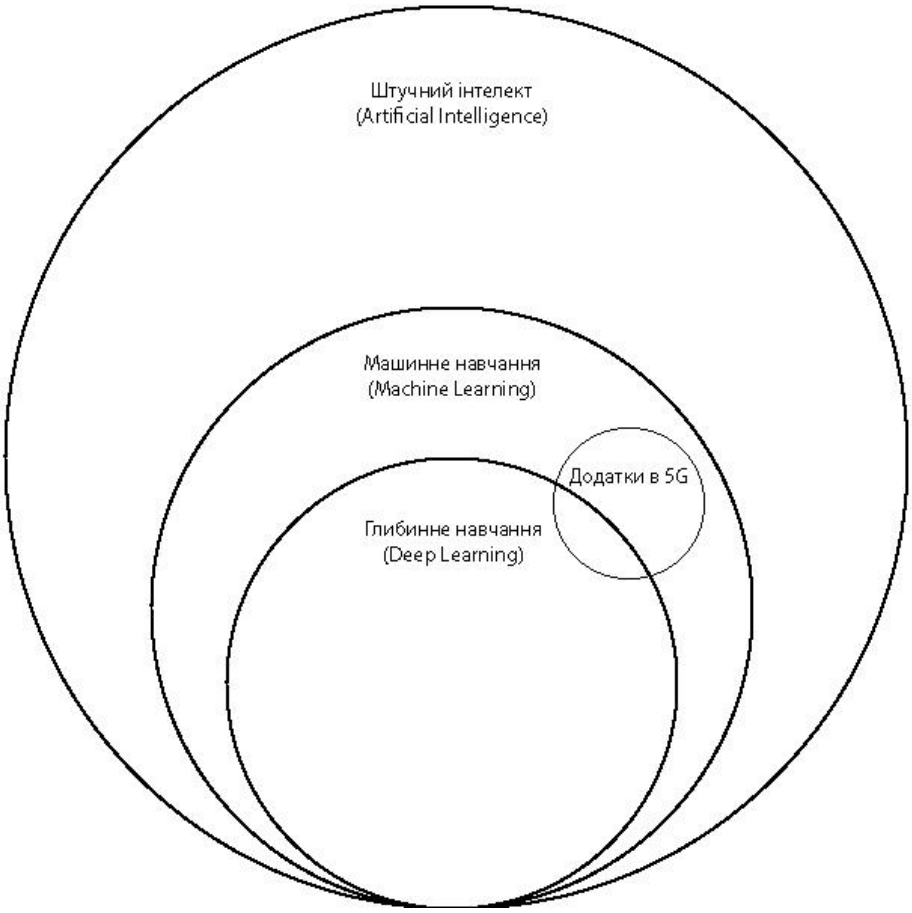


Рис. 2.1 Співвідношення між штучним інтелектом (AI), машинним навчанням (ML) та глибоким навчанням (DL).

В мобільних та бездротових мережах існує безліч параметрів, деякі з них встановлюються з використанням евристичних алгоритмів, оскільки це дозволяє пришвидшити рішення задачі в тих сценаріях, коли точне рішення не може бути знайдене. Для такого роду проблем алгоритм МН (наприклад, нейронна мережа) може вносити свій вклад, передбачаючи параметри та оцінюючи функції на основі наявних даних. Наступні покоління технологій мобільної та бездротової мереж також потребують використання оптимізації для мінімізації (або максимізації) деяких об'єктивних функцій.

З рисунку 2.1, машинне навчання є частиною штучного інтелекту. Тобто машинне навчання вважається штучним інтелектом, але не весь ШІ вважається МН. Наприклад, символічну логіку - механізми правил, експерті системи та графіки знань - можна назвати ШІ, але жодне з них не є МН. Одним із аспектів, що відділяє

машинне навчання від графіків знань та експертних систем, полягає в його особливостях змінювати себе під впливом великих об'ємів даних. Тобто машинне навчання є динамічним та не потребує втручання людини для внесення змін. Це робить його менш залежним від людини.

Машинне навчання пов'язане з аналізом даних. Аналіз даних - область математики та інформатики, що включає в себе розробку методів обробки даних. Гарним прикладом аналізу даних є підрахунок мінімального та максимального значень або відношення декількох величин, але при умові, що висновки нададуть важливі знання про дані та дозволять вирішити поставлені задачі. Але слід зазначити, що аналіз даних не є машинним навчанням в чистому виді, а важливою відмінністю є здатність вирішувати широкий спектр завдань без конкретного плану та опису алгоритму рішення.

Основними узагальненими задачами машинного навчання є:

- Регресії, що виконує функцію прогнозування значень мітки за набором пов’язаних компонентів. Таким чином, нехай X – безліч даних, що є описами деяких об’єктів, а Y – безліч можливих рішень для X. Мітка може приймати будь яке значення, а не лише обиратися з кінцевого набору значень – значення відомі тільки на об’єктах вибірки $XY = \{(x_1, y_1) \dots (x_n, y_n)\}, x \in X, y \in Y$. Алгоритми регресії моделюють залежність міток від пов’язаних компонентів, щоб визначити закономірність змін міток при різних значеннях компонентів. В такому випадку, на вхід алгоритму буде надходити набір прикладів з мітками відомих значень , а результатом буде функція $a: X \to Y$, котра вміє прогнозувати значення мітки залежно від набору вхідних компонентів. Прикладом для сценаріїв регресії можна вважати прогнозування продажу тарифу мобільного оператора в залежності від рекламного бюджету.
- Двійкова / багатокласова класифікація, що прогнозує розподіл елементів даних по двом чи більше категоріям (класам). В основі алгоритму класифікації є подання набору прикладів з мітками, кожна з яких представляє собою ціле число “0” або “1” для двійкової класифікації, де при багатокласовій класифікації

відбувається перетворення початкових текстових міток в числовий ключ. Але для класифікації також характерне розділення об'єктів на пересічні, непересічні та нечіткі класи, де для пересічних один об'єкт може належати кільком класам, непересічні - тільки одному, а для випадку нечіткого класифікування - об'єкт належить всім класам з певним ступенем належності. В результаті роботи алгоритму, буде отримано класифікатор, що вмітиме прогнозувати клас нових екземплярів без мітки. Прикладом двійкової класифікації є розподіл коментарів соціальної мережі Twitter за тональністю - позитивні чи негативні.

- Виявлення аномалій за допомогою аналізу головних компонентів. Виявлення аномалій на основі АГК дозволяє створювати моделі, коли отримати дані для навчання з одного класу не є складним завданням, але отримати достатню вибірку аномальних значень навпаки складно. Оскільки аномалії за своєю сутністю є рідкісною подією, то можуть виникати труднощі при сборі репрезентативної вибірки даних, які використовуються для моделювання.
- Ранжування створює засоби ранжування, беручи за основу набір прикладів з мітками. Ці набори містять в собі групи екземплярів, що можуть бути оцінені за заданими критеріями, а мітки ранжування для кожного екземпляру є числовим рядом, наприклад, $\{0, 1, 2, 3, 4\}$. Засіб ранжування навчається ранжувати нові групи екземплярів з невідомими оцінками для кожного екземпляру.
- Прогнозування використовує попередні дані часових рядів, щоб робити прогнози про поведінку в майбутньому. Прикладом прогнозування є звичайний прогноз погоди.
- Кластеризація, що проводить групування окремих екземплярів даних в кластери зі схожими характеристиками. Також дану задачу використовують для виявлення неможливих логічних зв'язків в наборі даних, які неможливо помітити переглядом чи при спостеріганні за даними. Описати кластеризацію можна наступним чином: нехай X є множиною даних, що містить опис деяких об’єктів; Y є безліччю кластерів, відмічених мітками; також визначена функція відстані між об’єктами з початкової множини X: $f(x, x)$, і є деяка навчальна вибірка об’єктів

$X_0 = \{x_n, y_n\}, x \in X$. Слід зазначити, що вхідні та вихідні дані напряму залежать від методу машинного навчання, рівно як і число кластерів заздалегідь невідомо і задається суб'єктивно. Надалі відбувається розбиття навчальної вибірки на кластери, де приписується номер кластера y_i для кожного x так, щоб близькі об’єкти належали одному кластеру, а об’єкти різних кластерів істотно відрізнялися за метрикою f. Будується алгоритм $a: X \to Y$, який ставить кожному $x \in X$ ставить у відповідність номер кластера $y \in Y$m що показано на рис.2.4 та рис.2.5.

Тобто немає чіткого критерію якості кластеризації, а існує лише ряд евристичних критеріїв, що виконують кластеризацію за одними даними, але з різними результатами. При цьому, незважаючи на описані складності, кластеризація допомагає досягти покращити розуміння даних простим розбиттям вибірки на групи схожих об’єктів, що спрощує подальшу обробку даних через застосування особливих методів аналізу до кожного окремого кластеру. Також вдається виявити аномалії та нові нетипові об'єкти, які не вдається віднести до жодного з кластерів. Прикладом сценаріїв для використання кластеризації є розподіл користувачів тарифів мобільного зв'язку на сегменти, беручи до уваги об'єм послуг обраних тарифів.

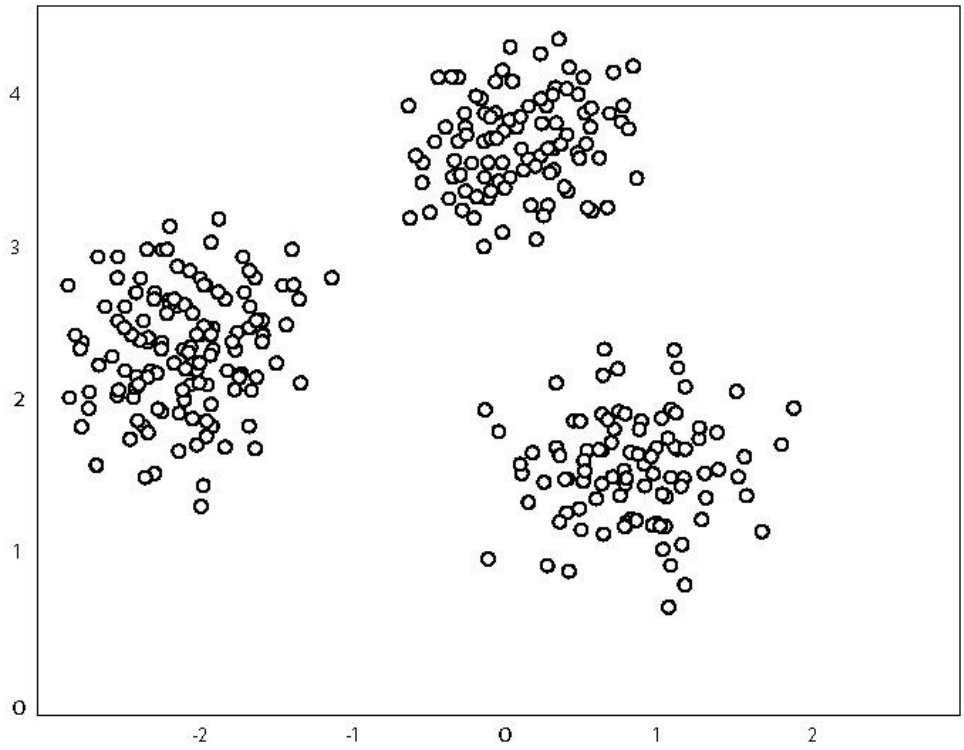


Рисунок 2.2 Кластеризація. Приклад заданих початкових даних

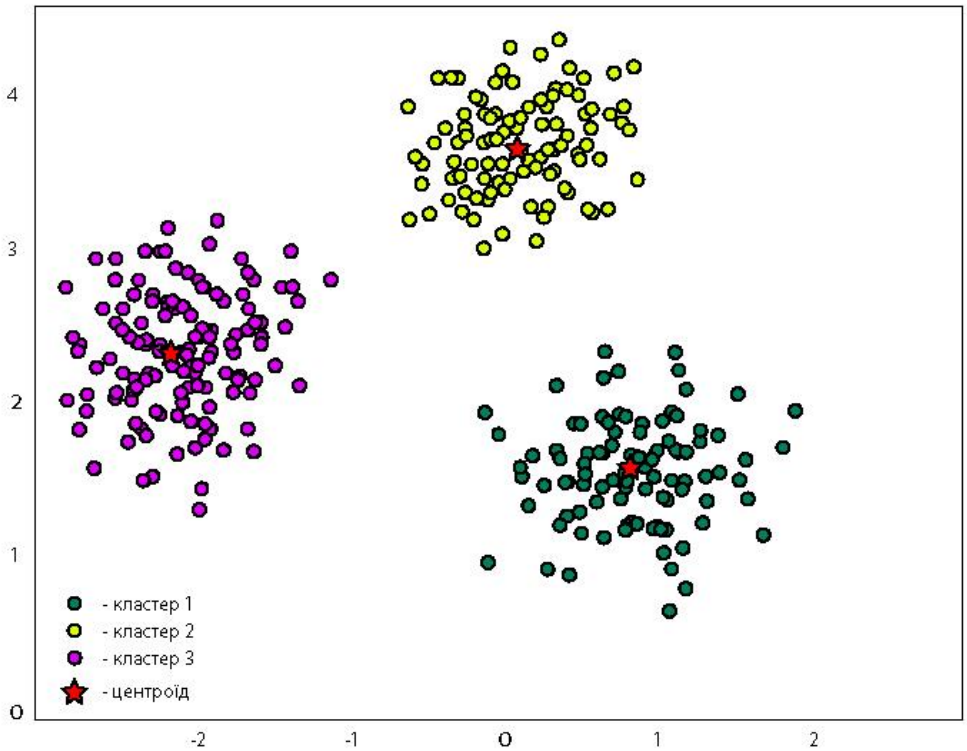


Рисунок 2.3 Кластеризація. Графічне зображення результату кластеризації

З вищевказаного опису задач можна зробити висновок, що для розробки машинному навчанню необхідні оцифровані дані, оскільки вони є ключем до якості та повноти рішення.

2.3 Моделі та алгоритми машинного навчання

Моделі описують переконання про те, як функціонує та чи інша система, вони пояснюють систему та вивчають наслідки різних компонентів на прогнозування поведінки.

2.3.1 Математичні моделі

Математична модель – це опис системи, що використовує математичні поняття та мову, а математичне моделювання – процес розробки математичної моделі. В загальному випадку, математичні моделі складаються з відносин та змінних. Тобто модель описує систему набором змінних та рівнянь, які встановлюють відносини між змінними. Змінні можуть бути багатьох типів, наприклад, цілі або дійсні числа, логічні значення або строки. Незважаючи на те, що змінні представляють деякі властивості системи, фактична модель являє собою набір функцій, описуючих відношення між різними змінними. Ці відношення можуть бути описані алгебраїчними операторами, функціями, диференціальними операторами і т.д. Вивчаючи моделі важливо виявляти широкі категорії з моделей. Класифікація окремих моделей за цими категоріями одразу вказують на деякі з основних елементів їх структури. За їх структурою можна використовувати декілька критеріїв класифікації для математичних моделей. Наприклад, статична та динамічна моделі або детерміністична або стохастична моделі. Статичні моделі не враховують варіації з часом, тоді як динамічні моделі в явному вигляді пов’язані на взаємодії з часом. В детерміністичних моделях всі математичні і логічні

відношення між елементами фіксовані, в результаті чого, всі відношення повністю визначають рішення. В стохастичній моделі хоча б одна змінна є випадковою.
В інженерії математичні моделі використовуються для максимізації
певного виходу. Наприклад, спостережувана система може потребувати певних вхідних даних, а відношення між входами та виходами може залежати від других змінних, таких як змінні прийняття рішення, змінні стану та випадкові змінні. Змінні рішень також називають незалежними змінними. Змінні не залежить одна від одної, оскільки змінні стану залежать від рішення, вхідних та вихідних змінних. Крім того, вихідні змінні залежать від стану системи.

На відміну від спостережуваних змінних, приховані змінні є змінними, котрі не спостерігаються безпосередньо, а визначаються з інших змінних. Математичні моделі, що направлені на пояснення спостережуваних змінних з точки зору латентних змінних, називаються латентними змінними моделями. Такі моделі використовуються в машинному навчанні / штучному інтелекті, фізиці, менеджменті та біоінформатиці.

Одною з переваг використання прихованих змінних є те, що вони можуть використовуватись для зменшення розмірності даних. Велика кількість спостережуваних змінних може бути агреговано в моделі для представлення базової концепції, що облегшує розуміння даних. Вони виступають функцією аналогічною до функції наукових теорій. Однак, приховані змінні пов’язують спостережувані реальні дані з символічними модульованими даними.

Проблеми математичного моделювання часто класифікуються на моделях “чорного ящика” та “білого ящика”, в залежності від того, наскільки доступна інформація про систему. Модель “чорного ящика” – система, про яку зовсім немає доступної інформації, в той час як в моделі “білого ящика” вся необхідна інформація доступна. Практично всі системи знаходяться між цими двома моделями, тому ця концепція існує тільки в якості інтуїтивного опори для прийняття рішення про підхід. Важливо використовувати якнайбільш апріорну інформацію для створення більш точної моделі. Правильне використання цієї

інформації дозволить моделі вести себе коректно. Зазвичай така інформація надходить в формі знань типу функцій, пов'язаних з різними змінними. В моделях "чорного ящику" відбувається оцінювання функціональної форми відношень між змінними та числові параметри в цих функціях.

Будь яка модель, що не є чистим "білим ящиком", містить деякі параметри, котрі можуть бути використані для підгону моделі до системи, яку вона повинна описати. Якщо моделювання відбувається за допомогою машинного навчання, то оптимізація параметрів буде називатись навчанням. В звичайному моделюванні через явно задані математичні функції параметри зазвичай визначаються апроксимацією кривої.

2.3.2 Статистична модель

Під статистичною моделлю для машинного навчання розуміють математичну модель, яка втілює набір статистичних припущень, що стосуються генерації вибіркових даних. Статистична модель являє собою процес генерації даних (в ідеальній формі). Статистична модель задається як математична залежність між одним або кількома випадковими змінними та іншими невипадковими змінними.

Статистична модель зазвичай розглядається як пара (S, P), де S – набір можливих спостережень (вибірка), а P – набір розподілу ймовірності на S. Нехай, існує істинний розподіл ймовірності, який виконується завдяки генеруванню даних для спостережень. P обирається для представлення набору, який містить розподіл, наближений до істинного. Слід зазначити, що немає потреби містити істинне значення розподілу в множині значень P. Набір $P = \{ P_\theta : \theta \in \Theta\}$. Набір Θ описує параметри моделі. Параметризація, як правило, потрібна для відокремлених значень параметрів, які приведуть до різних розподілів, $P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2$ (повинна бути ін'єкційною). Параметризація, задовольняюча вимоги, вважається ідентифікованою.

Якщо задано статистичну модель (S, P) з $P = \{ P_\theta : \theta \in \Theta\}$, то вважається, що вона параметризована Θ має кінцевий розмір. Береться до уваги, що $\Theta \subseteq R^k$, де k – додатне ціле число, що називають розмірністю моделі. Статистична модель є напівпараметричною, якщо має як кінцеві, так і нескінченні параметри, та є непараметричною, якщо набір параметрів Θ не має кінцевих розмірів.

2.3.3 Алгоритми машинного навчання

Алгоритмом в машинному навчанні називають процедуру, що виконується над даними для створення моделі машинного навчання. Алгоритми машинного навчання виконують функцію “розпізнавання шаблонів”. Алгоритми вчаться по даних або вкладаються в набір даних. Існує безліч алгоритмів машинного навчання. Так, наприклад, є алгоритми класифікації як k-найближчі сусіди.
Узагальнений алгоритм процесу машинного навчання зображено на рис.2.4.

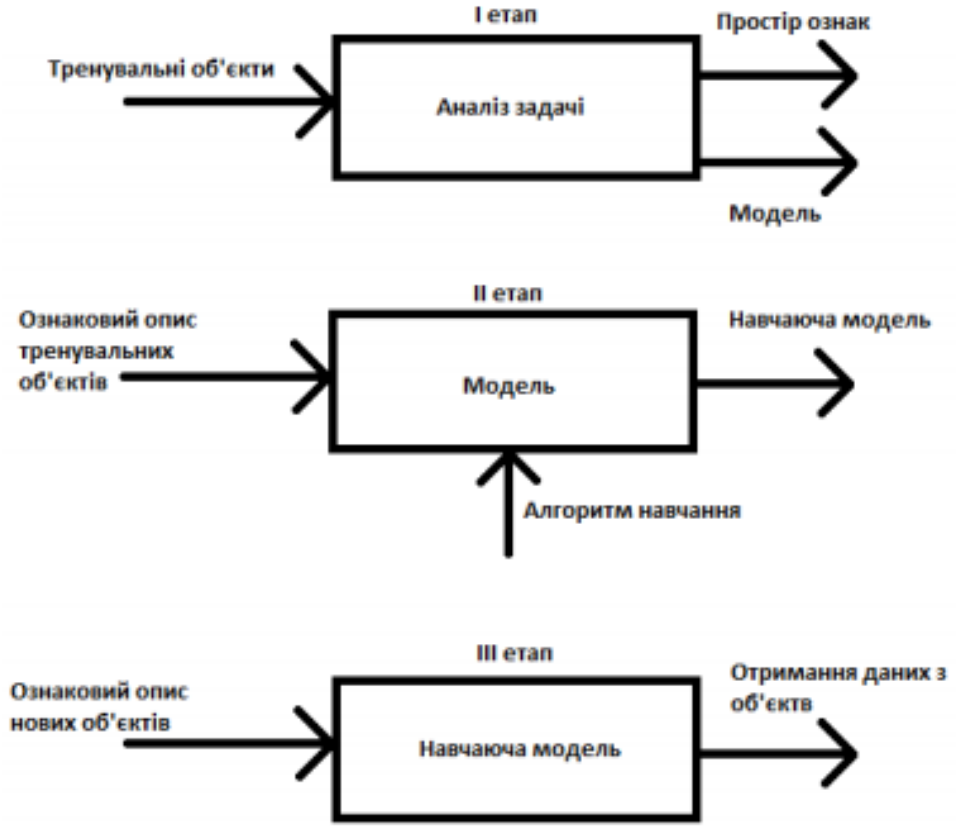


Рис.2.4 Узагальнений алгоритм проекту машинного навчання

Слід наголосити, що представлена ілюстрація відображає узагальнений випадок, в результаті якого отримано одразу працюючу модель. Але між другим та третім етапами є також чимало важливих дій – наприклад, оцінка якості моделі. Саме за результатами оцінки якості моделі буде прийнято рішення щодо переходу на наступний етап або повернення до попередніх.

Коли алгоритм асоційований з обробкою інформації, дані можуть можуть бути зчитувати з джерела вводу, записуватись на прилад виводу та зберігатись для подальшої обробки. Збережені дані розглядаються як частина внутрішнього стану об’єкту, виконуючого алгоритм. На практиці, вони зберігаються в одній або декількох структурах даних. Таким чином, алгоритм повинен бути чітко визначеним, а порядок розрахунків мати вирішальне значення для функціонування алгоритму.

В мовах програмування алгоритм призначений насамперед для вираження алгоритмів в формі, яку здатен виконати комп’ютер. Проектування алгоритмів є методом для рішення задач та інженерних алгоритмів. Є частиною багатьох теорій рішення операційних дослідів, як динамічне програмування. Методики проектування та реалізації алгоритмів також називають шаблонами. Один з найважливіших аспектів проектування алгоритму є створення алгоритму, час виконання якого є ефективним.

Алгоритми машинного навчання можна описати як навчання цільової функції f, яка найкращим чином співвідносить вхідні змінні X та вихідну змінну $cY : Y = f(x)$. Оскільки невідомо, що з себе буде представляти функція f, то необхідно навчити їх за допомогою різних алгоритмів.

Одним з найбільш використовуваних алгоритмів в машинному навчанні є лінійна регресія. Як вже описувалося раніше, моделювання в першу чергу стосується мінімізації помилки моделі або якомога більш точного прогнозування.

Лінійну регресію можна представити у вигляді рівнянь, які описують пряму, найбільш точно зображаючи зв'язок між вхідними змінними X та вихідними змінними Y.

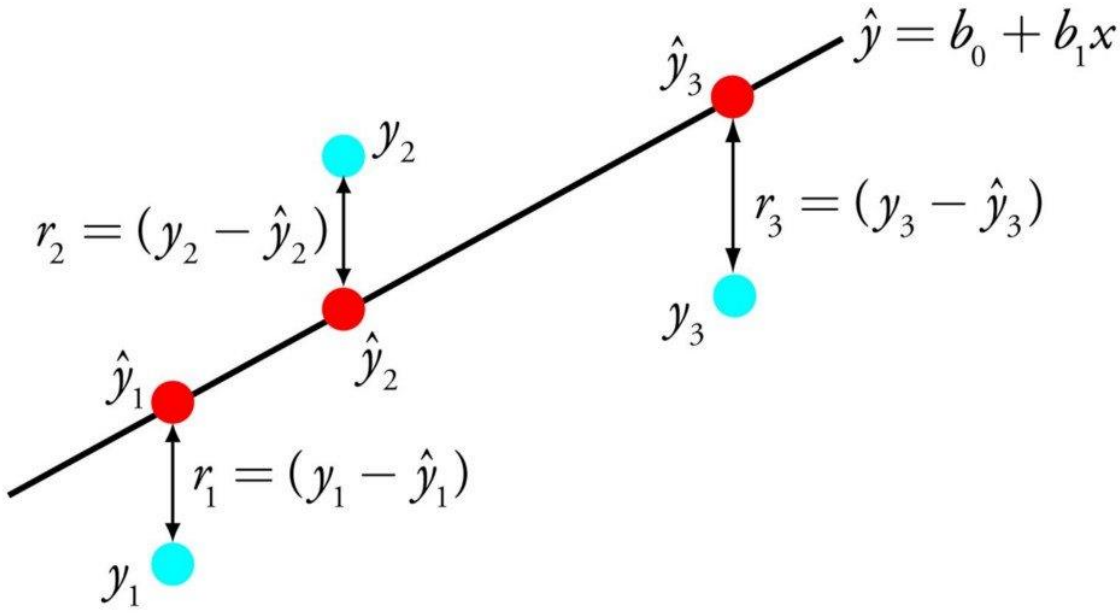


Рис.2.5 Графічне зображення використання лінійної регресії

Наприклад, $Y = B_0 + B_1 * X$. Знаючи X треба знайти Y. А мета лінійної регресії полягає в пошуку значень коефіцієнтів B_0 та B_1. Для оцінки регресивної моделі використовуються різні методи типу лінійної алгебри або методу найменших квадратів.

Також часто використовується дерево прийняття рішень. Дерево рішень можна представити у вигляді двійкового дерева, знайомого по алгоритмам та структурам даних. Кожен вузол представляє собою вхідну змінну і точку розподілу для кожної змінної, але тільки при умові, що змінна є числом. Тобто суть роботи полягає в послідовному розбитті безлічі даних на непересічні класи, які також піддаються розбиттю за будь-якими критеріями з оцінкою ефективності розбиття.

Дерево рішень складається з:

- “листя” - містять значення цільової функції;
- “гілок” - містять записи атрибутів, від яких залежить цільова функція;
- “вузлів” – містять інші атрибути, які використовуються при класифікації.

Дерева рішень найчастіше використовуються для класифікації (передбачається результат - клас, якому належать дані) та регресії (результатом є прогнозоване значення цільової функції).

Узагальнений алгоритм для побудови дерева прийняття рішень за навчальною вибіркою є наступним:

1. Береться атрибут та встановлюється в корінь дерева.
2. В “листі” даної “гілки” залишаються лише ті значення, які відповідають необхідній умові. Крок повторюється для кожного значення цього атрибута.
3. Продовжується побудова дерева із залишеного на попередньому кроці “листя”.

Дерева швидко навчаються та роблять прогнози. Крім того, вони є точними для широкого спектру задач та не потребують особливої підготовки даних.

Було б неправильно не розглянути й наступний алгоритм. Дуже простий та ефективний алгоритм – K-ближніх сусідів. Модель КНС представлена набором тренувальних даних. Прогнози для нової точки робляться виходячи з результатів пошуку K-ближніх сусідів в наборі даних та сумуванні вихідної змінної для цих K екземплярів. Для того, щоб визначити схожість між екземплярами даних необхідно використати евклідові відстані (при умові, що масштаб один і той самий для всіх параметрів) – числа, які можна вирахувати на основі відмінностей з кожною вхідною змінною.

Метод K-ближніх сусідів може вимагати багато пам’яті для зберігання всіх даних, але це компенсується швидкими прогнозами. Дані для навчання також можна оновлювати, щоб прогнози залишались точними з плином часу. Даний алгоритм може погано працювати з багатовимірними даними, що негативно вплине на ефективність алгоритму при вирішення задачі.

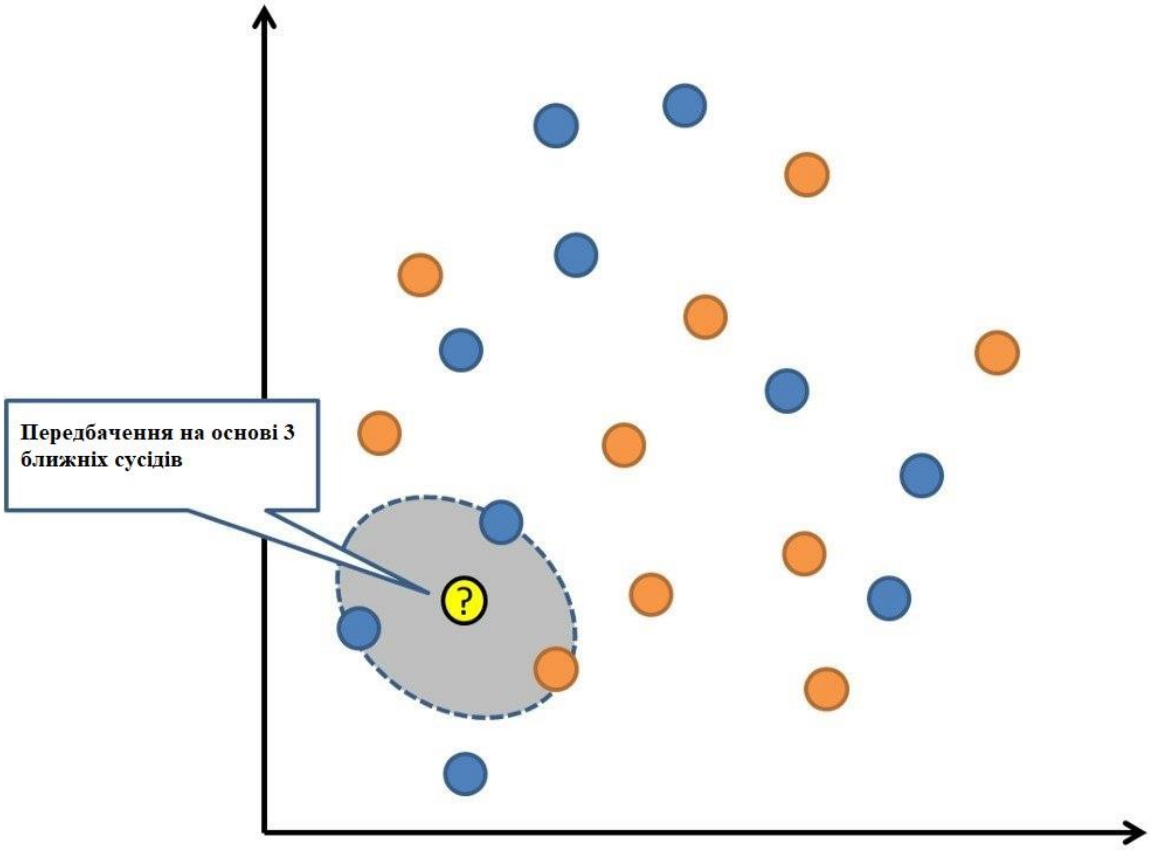


Рис. 2.5 Визначення подібності між екземплярами даних

2.4 Моделі кластеризації

Моделі кластеризацію можна класифікувати за наступними типами:

- На основі щільності: ці моделі розглядають кластери як щільно заповнену область, що має деякі спільні та відмінні якості в порівнянні з нижчою за щільністю областю. Цим моделям властиві точність та здатність до злиття двох кластерів.
- Ієрархічні моделі: кластери створюють деревовидну структуру на основі ієрархії. Нові кластери формуються виходячи з використання попередньо сформованого.
- Моделі секціонування: застосовується розділення об’єктів на k кластерів, де кожен розділ формує один кластер. Цей спосіб використовується для оптимізації функції подоби цільового критерію, коли відстань є основним параметром. Прикладом є розглянутий метод k-найближчих сусідів.

- Моделі на основі сіток: в таких моделях простір даних перетворюється в число осередків, які формують структуру сітки. Всі операції кластеризації виконуються швидко та незалежно від кількості об’єктів даних.

Розглянувши модель дерева рішень та K-найближчих сусідів, які були прикладами ієрархічної моделі та моделі секціонування, тепер звернемося до моделі кластеризації в квестах (англ. CLustering In QUEst).

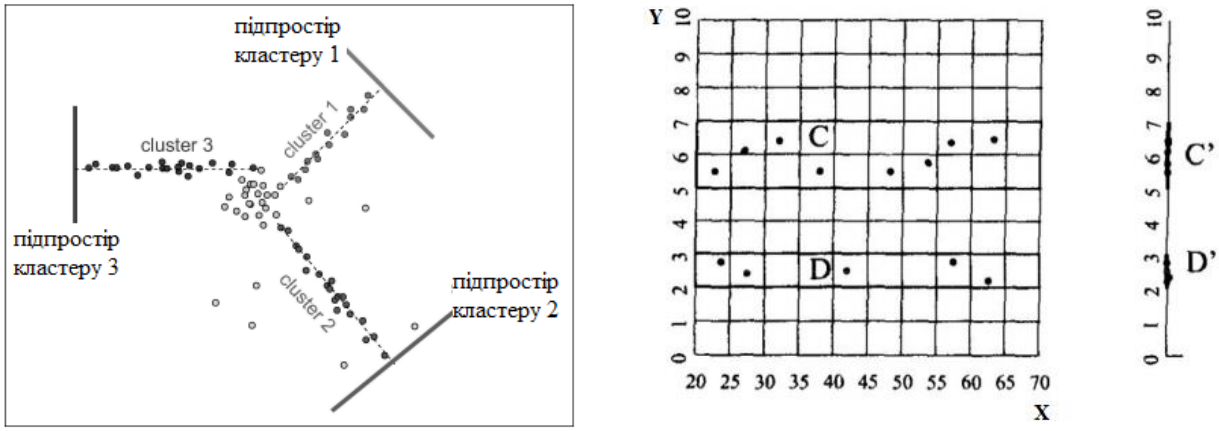
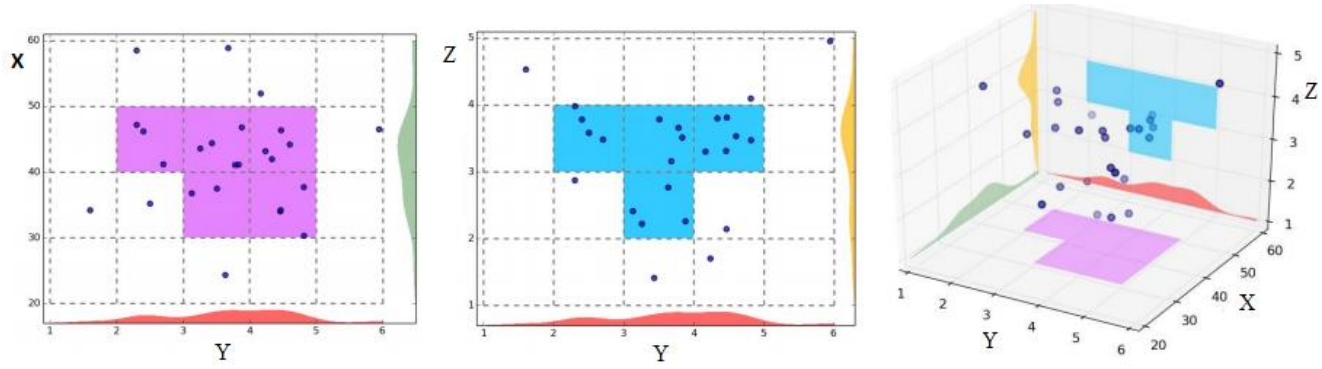


Рис.2.6 Представлення даних на основі щільності

CLIQUE – алгоритм кластеризації підпростору на основі щільності та сітки. По сітці проводиться дискретизація простору даних через сітку та оцінюється щільність, підраховуючи кількість точок в комірці сітки. Кластер на основі щільності - це максимальний набір пов’язаних за щільністю елементів в підпросторі. Елемент вважається щільним, якщо частка спільних точок даних, що містяться в елементі, перевищує вхідний параметр моделі. Кластером підпростору є набір сусідніх щільних комірок в довільному підпросторі. Він також виявляє деякі мінімальні описи кластерів. CLIQUE також автоматично визначає підпростори даних з великими розмірами, котрі дозволяють краще проводити кластеризацію, ніж початковий простір, з використанням апріорного принципу.



Апріорний принцип полягає в тому, якщо сукупність точок є кластером в k-мірному просторі, то ця сукупність також є частиною кластера в будь-яких (k-1)-мірних проекціях цього простору.

Основні етапи алгоритму CLIQUE:

- Визначення підпросторів, що містять кластери
  - Розбиття простору даних та пошук кількості точок, що лежать всередині кожної комірки розділу.
  - Визначення підпросторів, що містять кластери, з використанням апріорного принципу.
- Ідентифікація кластерів
  - Визначення щільних елементів в усіх підпросторах
  - Визначення пов’язаних щільних елементів в усіх підпросторах
- Створення мінімального опису для кластерів
- Визначення максимальних областей, котрі може покрити кожен кластер
- Визначення мінімального покриття кожного кластеру

Переваги алгоритму:

- Автоматичний пошук підпросторів найвищої розмірності при наявності кластерів високої щільності
- Ігнорування порядку запису у вхідні дані та не припускає канонічного розподілу даних
- Масштабується лінійно до розміру вхідних даних
- Простота методу та інтерпретування результатів

Недоліки:

- Як і у всіх підходах кластеризації на основі сітки, якість результатів значно залежить від вибору кількості та ширини секцій та комірок сітки.

Висновки до розділу 2

В результаті виконання даного розділу, було розглянуто основні задачі машинного навчання, його проблеми та методи боротьби з ними. Було проведено аналіз моделей машинного навчання, зокрема статистичної та математичної моделей. Також, в ході опрацювання даного розділу, було проведено аналіз моделей кластеризації та основних алгоритмів машинного навчання, серед яких дерево прийняття рішень, k-ближніх сусідів та статистичні моделі та методи.

РОЗДІЛ 3. АНАЛІЗ МОДЕЛЕЙ ПРОГНОЗУВАННЯ ВИБОРУ ТАРИФНОГО ПЛАНУ ДЛЯ ДВОХ КЛАСІВ КОРИСТУВАЧІВ

3.1 Вихідні дані

Алгоритм навчається на попередньо визначених даних, в даному випадку представлених у виді таблиці, де зазначені деякі параметри, за якими буде проводитись навчання. Оскільки, для отримання більш точного результату необхідні великі об’єми даних, то представлено буде лише невеликий витяг з використаних даних.

Для кожного елемента було встановлено необхідні значення параметрів, а саме обраний тарифний план – «chosenplan» (проміжні значення 1-9) та одинадцять анонімізованих параметрів з даними про абонента p1..11.

Таблиця 3.1 Вихідні дані абонентів

p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	chosen plan	class
5.9	0.180	0.28	1.0	0.037	24.0	88.0	0.99094	3.34	0.55	10.65	7	Sec
10.2	0.670	0.39	1.9	0.054	6.0	17.0	0.99760	3.19	0.47	10.0	5	First
8.4	0.715	0.2	2.4	0.076	10.0	38.0	0.99735	3.31	0.64	9.40	5	First
6.8	0.370	0.51	11.8	0.044	62.0	163.0	0.99760	3.19	0.44	8.80	5	Sec
8.9	0.750	0.14	2.5	0.086	9.0	30.0	0.99824	3.34	0.64	10.5	5	First

Для аналізу та обробки даних використовуються наступні бібліотеки (табл. 3.2).

Таблиця 3.2. Використані бібліотеки

Назва	Опис бібліотеки	Джерело
pandas	Використовується для обробки, організації та очистки даних	Стороння бібліотека
numpy	Використовується для підтримки великих багатомірних масивів та матриць, разом з функціями для операцій над цими масивами	Стороння бібліотека
scipy	Використовується для науково-технічних обчислень	Стороння бібліотека
io	Використовується для роботи з файлами (перегляд, читання, запис)	Стороння бібліотека
itertools	Використовується для створення власних ітераторів	Стороння бібліотека
matplotlib	Використовується для графічного аналізу	Стороння бібліотека
seaborn	Більш високорівневе API на базі matplotlib, використовується для візуалізації розподілу	Стороння бібліотека
statsmodels	Використовується для статистичного моделювання	Стороння бібліотека

```
import numpy as np
import pandas as pd
from io import BytesIO
from itertools import combinations
import matplotlib.pyplot as plt
from matplotlib import colors
import seaborn as sns
from sklearn import metrics
from scipy.spatial.distance import pdist
from scipy.cluster.hierarchy import fcluster, linkage, dendrogram
import statsmodels.api as sm
```

Рис.3.1. Вихідний код підключених бібліотек

Надалі необхідне підключення даних для їх подальшої обробки та аналізу. Дані знаходяться в “/trainings/intro_DS/data” на власному Google Drive. Виконується монтування, перевірка виконання команд та змісту даних.

```
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
```

```
Mounted at /content/drive
```

```
!ls /content/drive/'My Drive'/trainings/intro_DS/data
```

```
chosenplan-first.csv  chosenplan-second.csv
```

```
# Load first-class
dfr = pd.read_csv("/content/drive/My Drive/trainings/intro_DS/data/chosenplan-first.csv")
dfr['subs-class'] = 'first'
# Load second-class
dfw = pd.read_csv("/content/drive/My Drive/trainings/intro_DS/data/chosenplan-second.csv")
dfw['subs-class'] = 'second'
#Concatenate and shuffle data
df = pd.concat([dfr, dfw])
df = df.sample(frac=1, random_state=3).reset_index(drop=True)
```

```
df.head()
```

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	chosenplan	subs-class
0	5.9	0.180	0.28	1.0	0.037	24.0	88.0	0.99094	3.29	0.55	10.65	7	second
1	10.2	0.670	0.39	1.9	0.054	6.0	17.0	0.99760	3.17	0.47	10.00	5	first
2	8.4	0.715	0.20	2.4	0.076	10.0	38.0	0.99735	3.31	0.64	9.40	5	first
3	6.8	0.370	0.51	11.8	0.044	62.0	163.0	0.99760	3.19	0.44	8.80	5	second
4	8.9	0.750	0.14	2.5	0.086	9.0	30.0	0.99824	3.34	0.64	10.50	5	first

Рис. 3.2. Вихідний код монтування директорії з даними та перевірка їх змісту

3.2 Аналіз вихідних даних

Наступним кроком є аналіз даних. Але складною задачею для будь-якого аналітичного проекту є оцінка потенційного впливу аномалій. Тому необхідно звертати увагу на можливість появи артефактів даних, які можуть значно вплинути на створення моделей, але важливо й те, що артефакти можуть бути значущими в прогнозуванні результатів в особливий умовах.

Основним набором даних є обраний абонентом тарифний план. Оскільки, помітити наявність артефактів для використовуваного об’єму даних є складною задачею, то для кращого сприйняття побудуємо гістограму (рис.3.3) відповідно з даними.

```
_ = df.hist('chosenplan', bins=100, figsize=(10, 6))
```

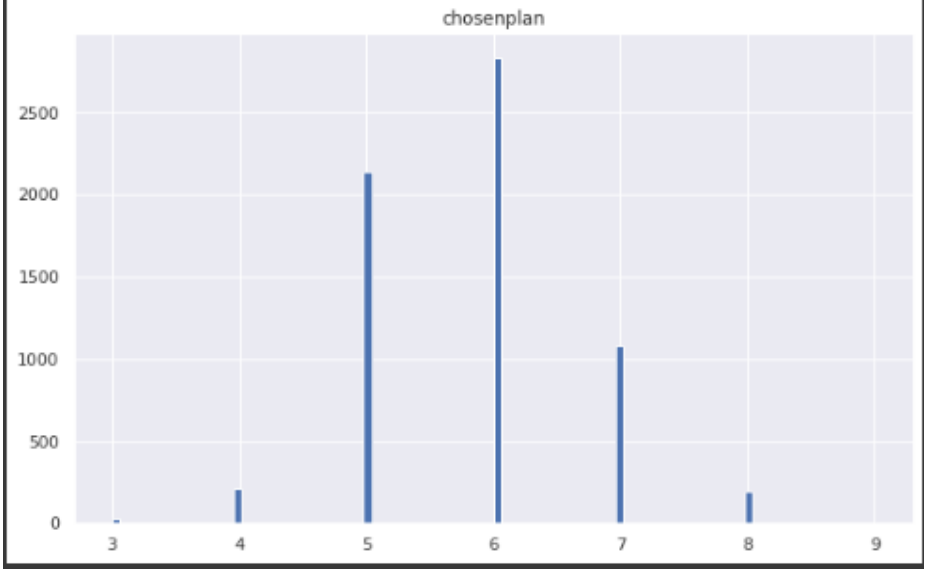


Рис.3.3. Вихідний код побудови гістограми та гістограма даних обраного тарифного плану

При дослідженні отриманої гістограми значень можна зробити висновок, що частка використання крайніх за позначеннями тарифів є дуже низькою. Це необхідно врахувати при виборі моделей прогнозування та оцінки ступеня взаємодії обраного тарифного плану та інших параметрів.

За таким же сценарієм проведемо аналіз розподілу інших параметрів:

```
df['class'] = df.chosenplan.apply(lambda x: "high" if x > 5 else "lower")
```

```
for column in df.columns.drop(['chosenplan', 'subs-class', 'class']):
    fig, (ax1, ax2) = plt.subplots(1,2, figsize=(20, 5))
    df.boxplot(column, 'class', ax=ax1, showfliers=False)
    ax1.set_title('')
    ax2.hist([df.loc[df['class'] == 'high', column], df.loc[df['class'] == 'lower', column]],
             label=['high', "lower"], color=['blue', 'orange'], density=True)
    ax2.legend()
    fig.suptitle('{} by class'.format(column), fontsize=14)
```

Рис.3.4. Вихідний код циклу побудови гістограми та діаграм розмаху параметрів p1..p11

Як видно з рисунку 3.4, додатково було створено два класи, критерієм яких є обраний тарифний план. Якщо в наборі даних окремого абонента значення параметра “chosenplan” перевищує 5, то такий абонент класифікується як “high”, інакше – “lower”.

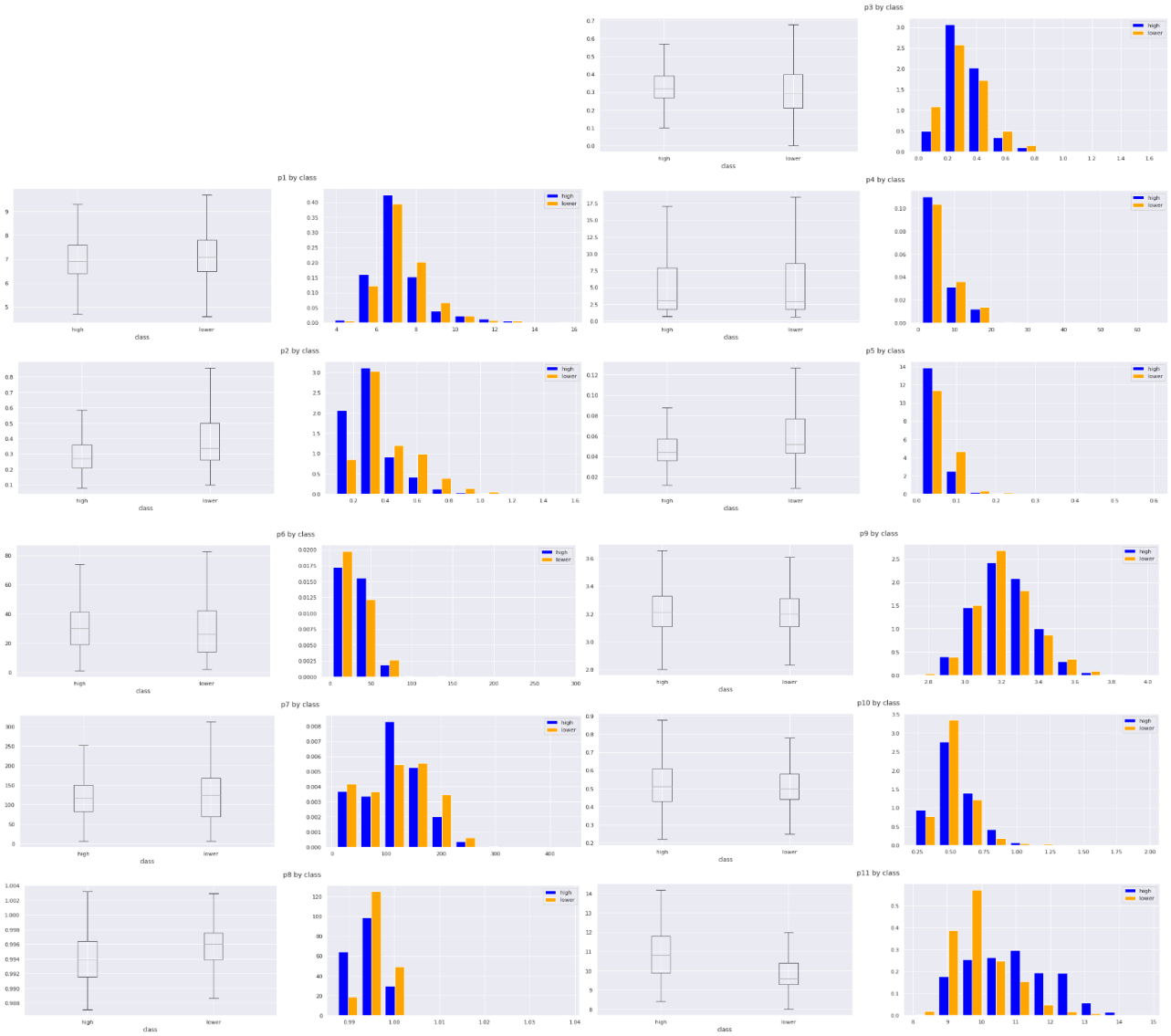


Рисунок 3.5. Гістограми даних параметрів p1..p11 та їх діаграми розмаху за класом

Оскільки є необхідним провести найбільш повний аналіз даних, щоб отримати повне уявлення, буде доречним також побудувати діаграми розмаху кожного параметру за їх значеннями та кореляцію з обраним тарифом.

```
for column in df.columns.drop(['chosenplan', 'subs-class', 'class']):
    fig, ax = plt.subplots(figsize=(7,5))
    df.boxplot(column, 'chosenplan', showfliers=False, ax=ax)
    ax.set_title('')
    fig.suptitle('{} by value'.format(column), fontsize=14)
```

Рисунок 3.6. Вихідний код циклу побудови діаграм розмаху за їх значеннями

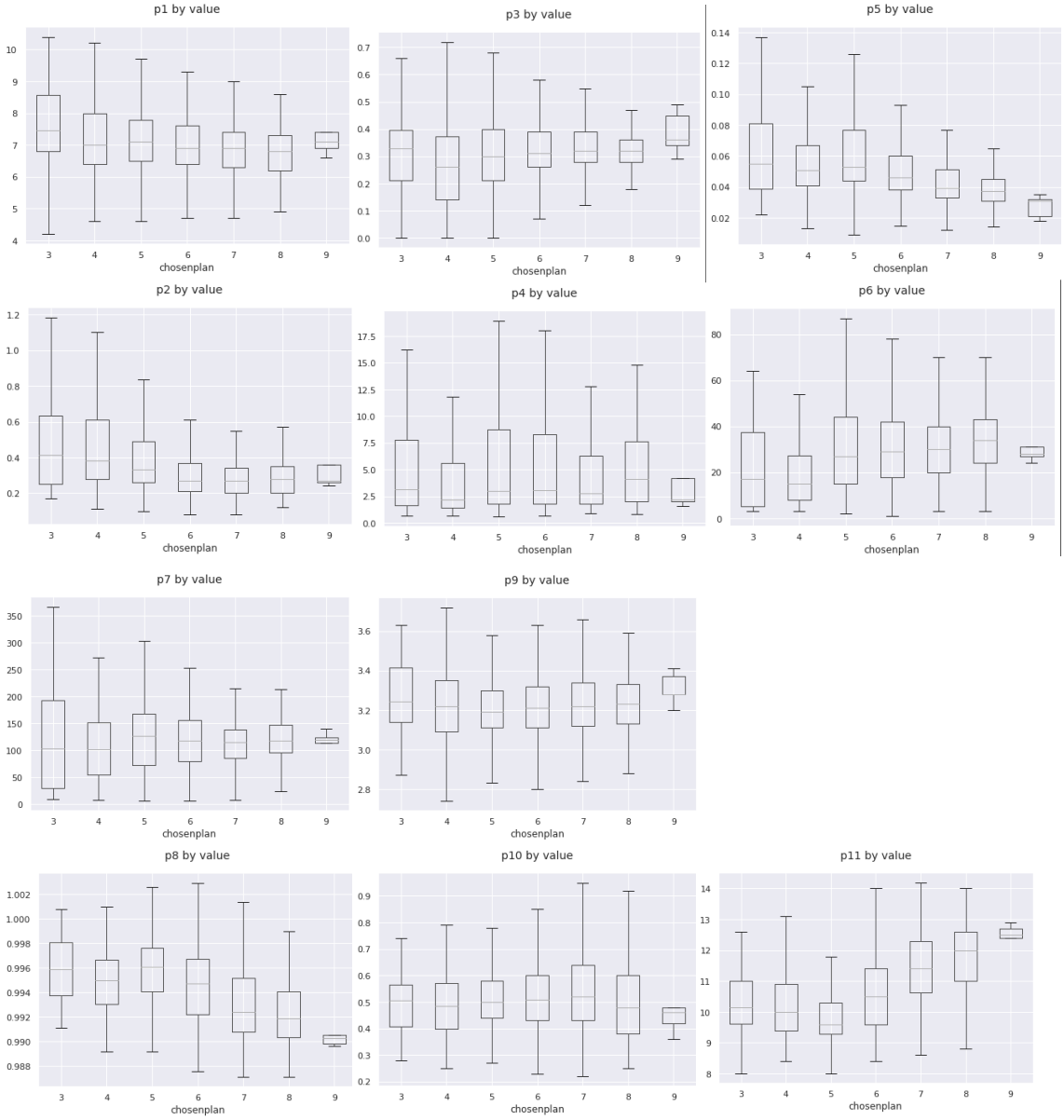


Рисунок 3.7. Діаграми розмаху параметрів p1..p11 за їх значеннями

Останнім етапом аналізу вихідних даних є кореляція та визначення предикторів, найбільш корельованих з обраним тарифом. Кореляція дозволяє виділити залежності між випадковими величинами, її коефіцієнт приймає значення

від -1 до 1. Якщо значення кореляції знаходяться в діапазоні від 0 до 1, то такі значення вважають позитивною кореляцією. Це означає, що в одному ряду даних значення збульшуються одночасно зі значеннями в іншому ряду, а коефіцієнт наближається до 1.

В свою чергу, коли значення кореляції в діапазоні від -1 до 0, тобто значення одного ряду збільшуються, а відповідні значення протилежного ряду зменшуються, такі значення кореляції називають негативно корельованими.

Таблиця 3.3 Значення коефіцієнту кореляції

Значення	Ступінь кореляції
0.8 – 1.0	Дуже сильна
0.6 – 0.8	Сильна
0.4 – 0.6	Середня
0.2 – 0.4	Слабка
0.0 – 0.2	Дуже слабка

```
df.corr().style.apply(background_gradient, axis=1) \
    .set_properties(**{'max-width': '120px', 'font-size': '12pt'})\
    .set_caption("Hover to magify")\
    .set_precision(2)\
    .set_table_styles(magnify())
```

Рис.3.8. Вихідний код кореляції з обраним тарифом

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	chosenplan
p1	1.00	0.22	0.32	-0.11	0.30	-0.28	-0.33	0.46	-0.25	0.30	-0.10	-0.08
p2	0.22	1.00	-0.38	-0.20	0.38	-0.35	-0.41	0.27	0.26	0.23	-0.04	-0.27
p3	0.32	-0.38	1.00	0.14	0.04	0.13	0.20	0.10	-0.33	0.06	-0.01	0.09
p4	-0.11	-0.20	0.14	1.00	-0.13	0.40	0.50	0.55	-0.27	-0.19	-0.36	-0.04
p5	0.30	0.38	0.04	-0.13	1.00	-0.20	-0.28	0.36	0.04	0.40	-0.26	-0.20
p6	-0.28	-0.35	0.13	0.40	-0.20	1.00	0.72	0.03	-0.15	-0.19	-0.18	0.06
p7	-0.33	-0.41	0.20	0.50	-0.28	0.72	1.00	0.03	-0.24	-0.28	-0.27	-0.04
p8	0.46	0.27	0.10	0.55	0.36	0.03	0.03	1.00	0.01	0.26	-0.69	-0.31
p9	-0.25	0.26	-0.33	-0.27	0.04	-0.15	-0.24	0.01	1.00	0.19	0.12	0.02
p10	0.30	0.23	0.06	-0.19	0.40	-0.19	-0.28	0.26	0.19	1.00	-0.00	0.04
p11	-0.10	-0.04	-0.01	-0.36	-0.26	-0.18	-0.27	-0.69	0.12	-0.00	1.00	0.44
chosenplan	-0.08	-0.27	0.09	-0.04	-0.20	0.06	-0.04	-0.31	0.02	0.04	0.44	1.00

Рис. 3.9. Значення кореляції параметрів з обраним тарифом

Залишається лише визначити предиктори, найбільш корельовані з обраним тарифним планом, якими є параметри “p2”, “p5”, “p8” та “p11” (рис.3.10).

```
vals = ['p2', 'p5', 'p8', 'p11']
cc = combinations(vals, 2)

for k,v in combinations(vals, 2):
    fig, (ax1, ax2, ax3) = plt.subplots(1,3, figsize=(20, 7))
    sns.regplot(k,v, data=df[df['class']=='lower'], fit_reg=False,  color="blue",
                scatter_kws={'alpha':0.1, 's':200}, ax=ax1, label='high')
    ax1.legend(loc='best')
    sns.regplot(k,v, data=df[df['class']=='high'], fit_reg=False,  color="green",
                scatter_kws={'alpha':0.1, 's':200}, ax=ax2, label='lower')
    ax2.legend(loc='best')
    sns.regplot(k,v, data=df[df['class']=='lower'], fit_reg=False,  color="blue",
                scatter_kws={'alpha':0.1, 's':200}, ax=ax3, label='high')
    sns.regplot(k,v, data=df[df['class']=='high'], fit_reg=False, color="green",
                scatter_kws={'alpha':0.1, 's':200}, ax=ax3, label='lower')
    ax3.legend(loc='best')
```

Рис.3.10. Вихідний код побудови графіків для кожного із предикторів

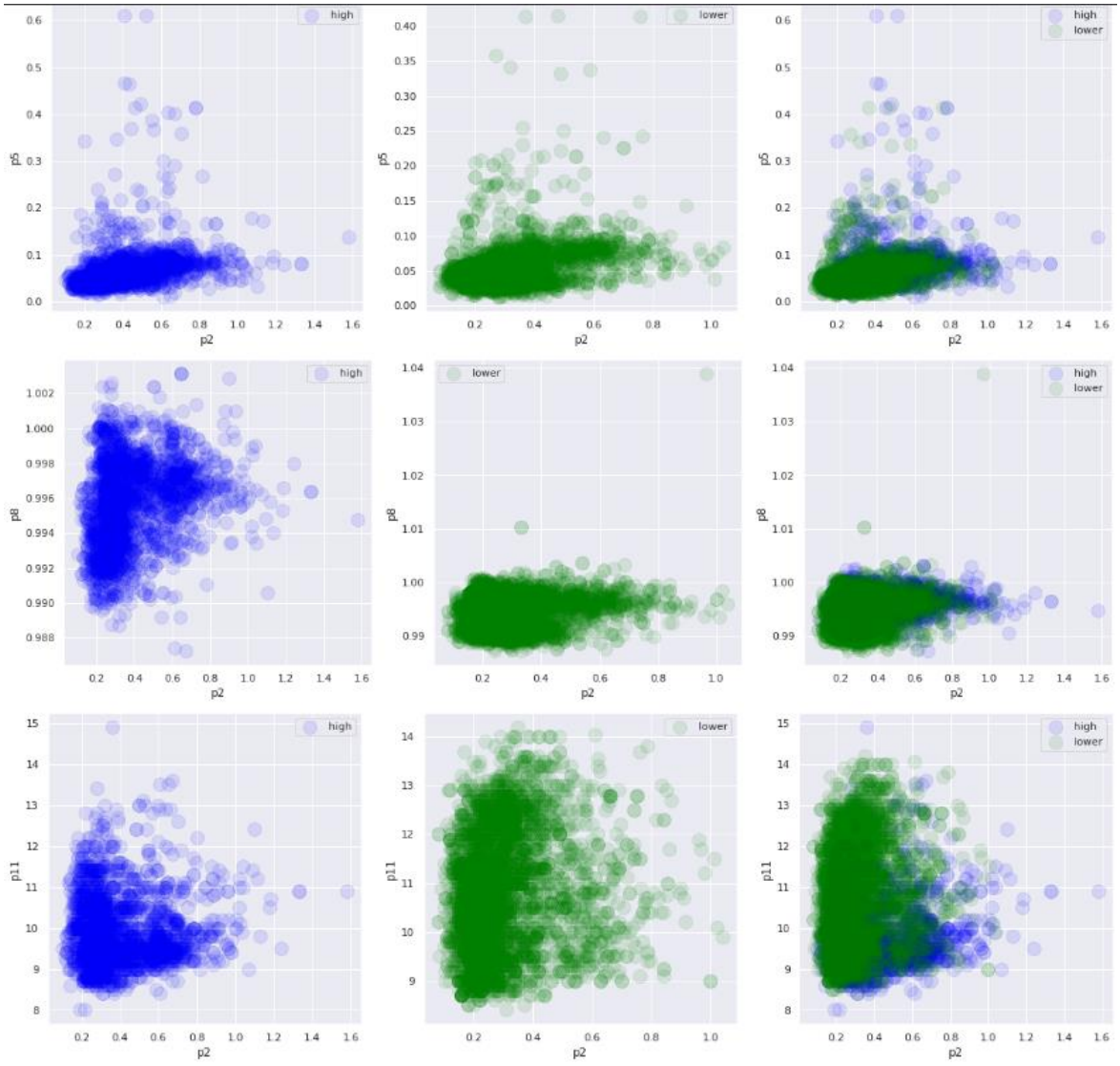


Рис. 3.11а

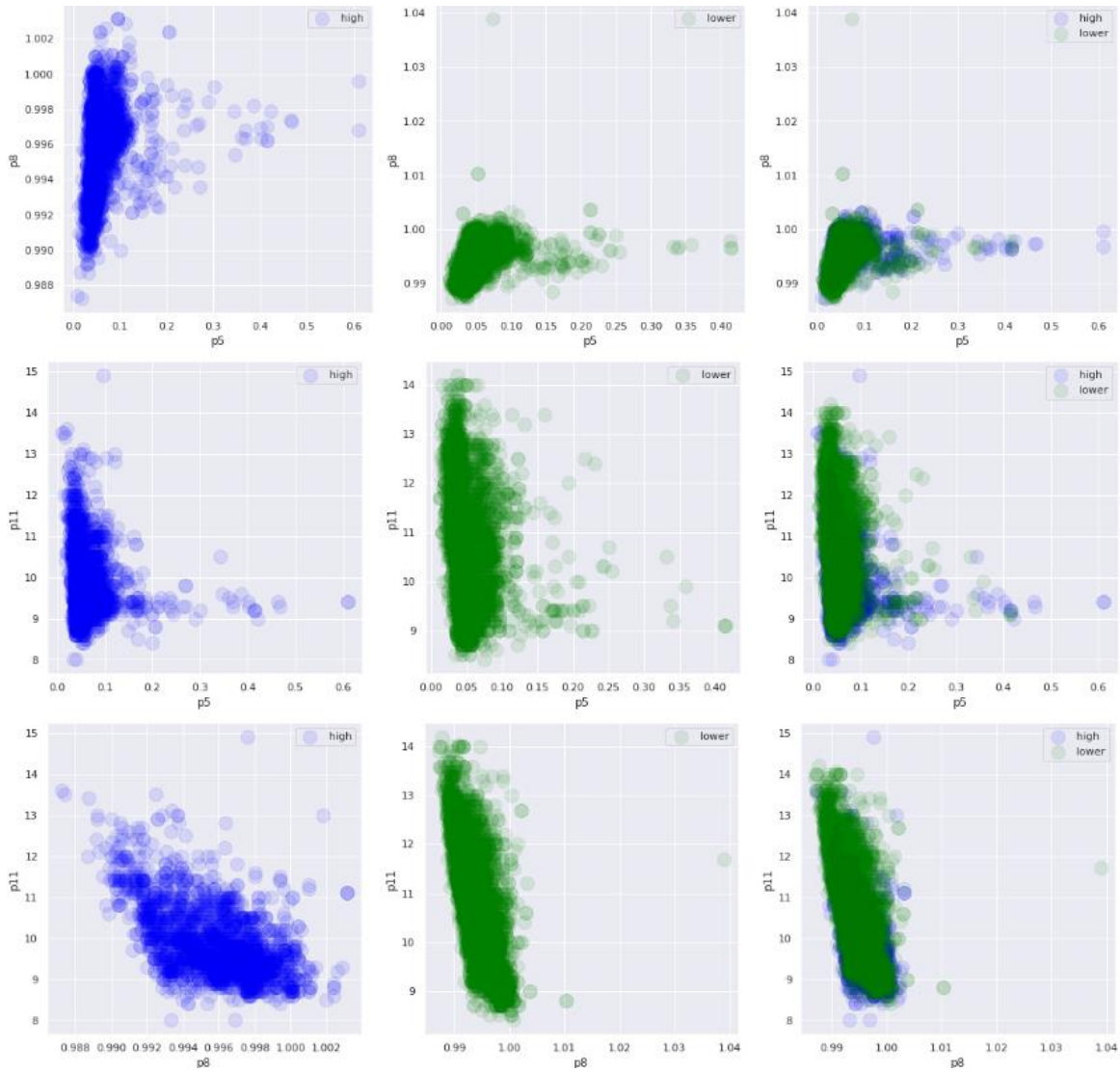


Рис.3.11б. Отримані графіки предикторів, найбільш корельованих з обраним тарифом

З наведених вище графіків (рис.3.11) видно, що прогнозуючі змінні показують хороший лінійний зв'язок.

3.3 Побудова моделей класифікації

3.3.1 Логістична регресія

Надалі будуть представлені моделі бінарної / двійкової класифікації, адже метою створення алгоритма є класифікація на дві категорії. Важливим моментом

бінарної класифікації є несиметричність двох класів за обсягом відмінних наборів даних з кожного класу.

Для створення моделі прогнозування та перевірки її здатності робити прогнози щодо вибору тарифного плану абонента будемо використовувати бібліотеку SciKit-Learn, яка є широко використовуваною для різних задач машинного навчання. SKLearn підтримує API “передбачень”, “тестування” та “підгону”, що значно спрощує її використанняв з багатьма методами та алгоритмами.

Першим кроком є підготовка даних, а саме конвертація та розділення даних. Для цього будемо використовувати функцію prepare_data() з модуля sklearn.model_selection, про що свідчить функція StratifiedShuffleSplit() на рис.3.12.

```
def prepare_data(X, y):
    """
    Convert and split data
    """
    split = StratifiedShuffleSplit(n_splits=1, test_size=0.3, random_state=42)
    for train_index, test_index in split.split(X, y):
        X_train, X_test = X[train_index], X[test_index]
        y_train, y_test = y[train_index], y[test_index]
    return X_train, X_test, y_train, y_test
```

Рис.3.12 Опис функції prepare_data()

Результатом виконання вищевказаної функції prepare_data() є розділення даних на набори для тестування (рис.3.13), де 30% призначається на тестування, 70% на навчання та параметру random_state призначається значення 42. Цей параметр контролює випадковість отриманих показників навчання та тестування.

```
X_train, X_test, y_train, y_test = prepare_data(df.drop(['class', 'subs-class', 'chosenplan'], axis=1).to_numpy(), df['class'].values)
```

Рис.3.13 Розділення даних

Наступною дією є побудова моделі логістичної регресії з використанням набору даних для навчання. Для цього використаємо функцію LogisticRegression (рис.3.14) та попередньо створену функцію evaluate_classification(), що

використовується для побудови графічного зображення отриманих результатів логістичної регресії.

```
lr_model = LogisticRegression(class_weight="balanced")
lr_model.fit(X_train, y_train)
lr_pred = lr_model.predict(X_test)
HTML(evaluate_classification(y_test, lr_pred, pos_label="high",
sample_weight=compute_sample_weight('balanced', y=y_test)))
```

Рис.3.14 Вихідний код побудови моделі логістичної регресії

Побудова графічного представлення результатів логістичної регресії, як було сказано вище, виконується за допомогою функції evaluate_classification(), в якій записано всі параметри побудови та стилі. Також в ній описані параметри, які визначають ступінь правильності прогнозу – F1 score та Accuracy score.

```
def evaluate_classification(actual, pred, pos_label=1, average='binary', sample_weight = None, fs=4):
    """
    Basic set of evaluation metrics for classification
    """
    return """
    <table style='width:100%'><tr><td style='text-align:left;vertical-align:top'><table style='font-size:14px'> \
    <tr><td><b>Accuracy score:</b></td><td>{1}</td></tr><tr><td><b>F1 score:</b></td><td>{2}</td></tr></table></td> \
    <td>{0}</td></tr></table>
    """.format(confusion_matrix_picture(actual, pred, fs=fs),
               metrics.accuracy_score(actual, pred, sample_weight=sample_weight),
               metrics.f1_score(actual, pred, pos_label=pos_label, average=average))
```

Рис.3.15 Опис функції evaluate_classification()

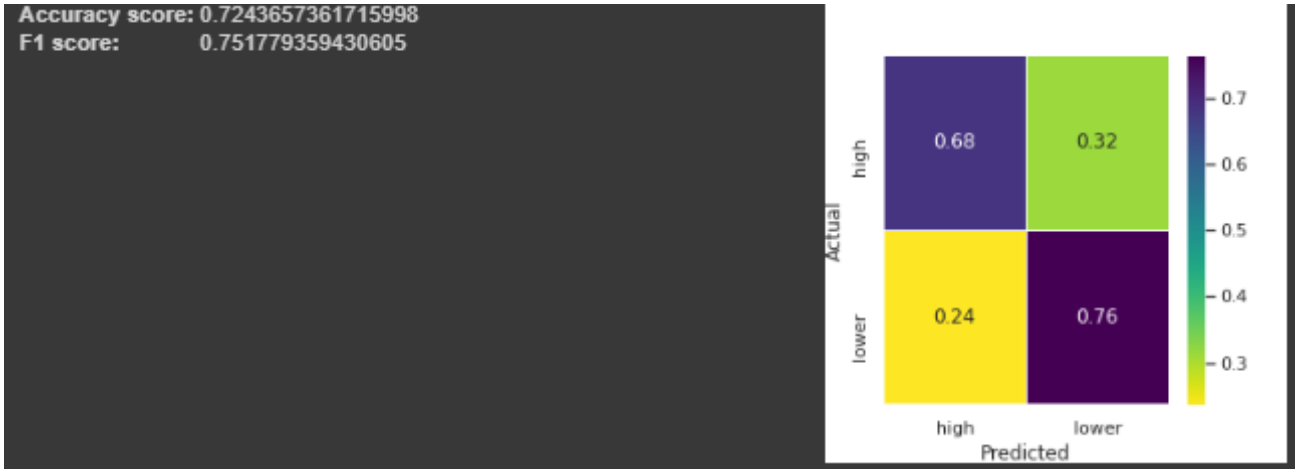


Рис.3.16 Модель логістичної регресії

Різниця між F1 score та Accuracy score полягає в тому, що в F1 score береться до уваги мінімізація невірних прогнозів. Також F1 score використовується в випадках, коли є значна перевага в об’ємах даних кожного з класів та більш

важливим є прогнозування невірних рішень, де Accuracy score використовується у випадках з пріорітетним прогнозом вірних рішень. Таким чином отримуємо, що завдяки моделі логістичної регресії 72.4% прогнозів будуть вірні.

3.3.2 Дерево рішень

Наступною моделю будемо розглядати дерево рішень. Для моледі дерева рішень також використовується бібліотека SKLearn, а саме функція DecisionTreeClassifier(). Параметр min_samples_leaf визначає мінімальне число вибірок, які повинні знаходитись у вузлі листа. Точки розподілення будуть оглянуті лише у випадку, коли вона складає не менше min_samples_leaf тренувальних елементів в кожній з правої та лівої «гілок».

Рис.3.17 Код побудови моделі та виведення результатів прогнозування

Слід додати, що побудова моделі також виконувалась зі застосування функції evaluate_classification(). В результаті побудови моделі дерева рішень, було визначено значення F1 score та Accuracy score, які складають 81.3% та 73.8% відповідно.

3.3.3 K-ближніх сусідів

Для побудови моделі також буде використовуватися бібліотека SKLearn, а саме функція KNeighborsClassifier(), зі встановленими параметрами n_neighbors=2 та algorithm = «auto». Перший параметр визначає кількість сусідів і має значення “2”, а algorithm визначає саме алгоритм для вираховування ближніх сусідів. Значення «auto» автоматично визначає накращий алгоритм на основі значень, поданих до методу fit() .

Рис. 3.18 Код побудови моделі k-ближніх сусідів та виведення результатів

З отриманих результатів видно, що при використанні моделі k-ближніх сусідів Accuracy score = 62.2%, а F1 score = 79.4%. Складемо загальну таблицю з результатами вірного прогнозування моделей.

Таблиця 3.4. Порівняння результатів вірного прогнозування

	Логістична регресі	Дерево рішень	k-ближніх сусідів
Accuracy score, %	72,4	73,8	62,2
F1 score, %	75,2	81,3	79,4

Виходячи з даних таблиці 3.4, можна зробити висновок, що при умові необхідності прогнозування вірного рішення, що в даному випадку є правильно підібраним тарифним планом для кожного з класів абонентів, прогнози краще складати за допомогою моделі дерева рішень, оскільки має найвищі показники.

Висновки до розділу 3

В результаті виконання даного розділу, було продемонстровано загальний процес аналітичного проекту, а саме обробка вихідних даних, дослідницький аналіз даних, вибір та побудова моделей з використанням аналітичних бібліотек, а також їх оцінка. Було продемонстровано вибір необхідних функцій, їх опис та способи використання, виведення графічних результатів моделей прогнозування з використанням бібліотек SciKit-Learn та StatsModels. Також, було показано виведення основних метрик класифікації «F1 score» та «Accuracy score», які дозволяють отримати загальне значення вірогідності правильного прогнозування для кожної моделі.

ЗАГАЛЬНІ ВИСНОВКИ

В даній роботі було проаналізовано використання алгоритмів машинного навчання в телекомунікаційних системах 5G. Зокрема було розкрито загальні проблеми систем 5G, білінгу та управління абонентами з можливістю впровадження технологій машинного навчання.

Було проведено загальний огляд нового покоління мобільного зв'язку 5G, визначено основнім проблеми та можливі рішення, проведено порівняння основних можливостей з минулими поколіннями. Було описано проблему створення загальної системи білінгу при впровадженні 5G та управління абонентами, розглянуто основні способи надання послуг абонентам.

При аналізі методів та алгоритмів машинного навчання також було проведено загальний огляд, виділено основні задачі, методи, моделі та алгоритми. Було розкрито способи його використання в промисловості. Слід додати, що також було розглянуто математичні та статистичні моделі машинного навчання, моделі кластеризації. Приведено узагальнений алгоритм проекту машинного навчання.

Було виділено основні моделі класифікації для проведення аналізу їх використання: логістична регресія, дерево прийняття рішень та модель k-ближніх сусідів. Також визначено сторонню бібліотеку SKLearn для побудови моделей на основі машинного навчання, оскільки вона має переваги у вигляді проведення тестування, прогнозування та підгону моделей.

Було описано роботу функцій LogisticRegression(), boxplot(), predict(), corr(), StratifiedShuffleSplit(), prepare_data(), evaluate_classification(), fit(), KNeighborsClassifier(), DecisionTreeClassifier(). Також було описано значення основних метрик класифікації «Accuracy score» та «F1 score», які визначають вірність проведених прогнозів моделей, було визначено доцільність використання цих метрик в окремих випадках.

На прикладі аналізу моделей прогнозування вибору тарифного плану двом класам користувачів можна зробити висновок про необхідність впровадження технологій машинного навчання для значного пришвидшення аналізу та обробки величезних об’ємів даних.

Результати даних досліджень можуть бути використані для підбору операторами необхідних послуг, тарифних планів абонентів в цілях отримання максимальної його якості для можливостей та потреб абонентів. Також є доцільним використання результатів даних досліджень в навчальних дисциплінах з машинного навчання.

Перелік посилань

1. Bloch D. Machine Learning: Models and Algorithms [Електронний ресурс] / Daniel Bloch // SSRN. – 17. – Режим доступу до ресурсу: ssrn.com/abstract=3307566.
2. MANUEL EUGENIO MOROCHO-CAYAMCELA. Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions [Електронний ресурс] / MANUEL EUGENIO MOROCHO-CAYAMCELA, HAEYOUNG LEE, WANSU LIM // IEEE Access – Режим доступу до ресурсу: https://www.researchgate.net/publication/335937022_Machine_Learning_for_5GB5G_Mobile_and_Wireless_Communications_Potential_Limitations_and_Future_Directions.
3. Friman J. 5G BSS: Evolving BSS to fit the 5G economy [Електронний ресурс] / J. Friman, M. Nilsson, E. Mueller // Ericsson Technology Review – Режим доступу до ресурсу: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/evolving-bss-to-fit-the-5g-economy.
4. Матеріали сайту tutorialspoint. [Електронний ресурс] – Режим доступу до ресурсу: https://www.tutorialspoint.com/telecom-billing/billing-introduction.htm.
5. Матеріали сайту scikit-learn. [Електронний ресурс] – Режим доступу до ресурсу: https://scikit-learn.org/stable/.
6. Матеріали вільної енциклопедії «Вікіпедія» [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/.
7. Матеріали сайту MSDN. [Електронний ресурс] – Режим доступу до ресурсу: https://docs.microsoft.com/en-us/dotnet/machine-learning/resources/tasks.
8. Géron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow / Aurélien Géron.. – 525 с.
9. A convergent billing system for the 5G era [Електронний ресурс] // Mobile World Live. – 2020. – Режим доступу до ресурсу: https://www.mobileworldlive.com/latest-stories/a-convergent-billing-system-for-the-5g-era/