Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями»

Виконала: студентка 2 курсу, групи 2ІСТ-24м
спеціальності 126 – «Інформаційні системи та технології»
_____ Анна СУДЕЦЬ

Керівник: к.т.н., асистент каф. САІТ
_____ Ігор ШТЕЛЬМАХ
«27» _____ 2025 р.

Рецензент: к.т.н., доц. каф. КН
_____ Володимир ОЗЕРАНСЬКИЙ
«03» _____ 2025 р.

Допущено до захисту

Завідувач кафедри САІТ
_____ д.т.н., проф. Віталій МОКІН
«28» _____ 2025 р.

Вінниця ВНТУ – 2025 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій
Рівень вищої освіти – другий (магістерський)
Галузь знань – 12 Інформаційні технології
Спеціальність – 126 Інформаційні системи та технології
Освітньо-професійна програма – Інформаційні технології аналізу даних та зображень

ЗАТВЕРДЖУЮ
Завідувач кафедри САІТ
_____ д.т.н., проф. Віталій МОКІН
«25» 09 2025 року

ЗАВДАННЯ
НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТЦІ
Судець Анні Олександрівні

1. Тема роботи: «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями»
керівник роботи: Ігор ШТЕЛЬМАХ, к.т.н., асистент каф. САІТ
затверджені наказом ВНТУ від «24» 09 2025 року №313
2. Термін подання студенткою роботи 28.11.2025 року
3. Вихідні дані до роботи:
Набір даних про рівні задоволеності пасажирів авіакомпаніями https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction.
4. Зміст текстової частини:
   1) Характеристика об'єкту досліджень;
   2) Розвідувальний аналіз даних;
   3) Розроблення інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями;
   4) Економічна частина.
5. Перелік ілюстративного матеріалу:
   1) Кореляційна матриця;
   2) Аналіз впливу класу та типу подорожі на задоволеність;
   3) Аналіз впливу віку на рівень задоволеності;
   4) Діаграма розгортання;
   5) Матриця плутанини моделі XGBoost;
   6) Матриця плутанини 5-шарової нейронної мережі;

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
3	Євгеній КРИЖАНОВСЬКИЙ, к. т. н., доцент каф. САІТ	(дата і підпис) 05.10.2025	(дата і підпис) 25.10.2025
4	Олександр ЛЕСЬКО, д. е. н., проф. каф. ЕПВМ	05.11.2025	15.11.2025

7. Дата видачі завдання «25» 09 2025 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання		Примітки
		початок	закінчення	
1	Характеристика об'єкту досліджень	(дата) 15.09.2025	(дата) 25.09.2025	Виконано
2	Вибір оптимальних інформаційних технологій	25.09.2025	05.10.2025	Виконано
3	Розвідувальний аналіз даних	05.10.2025	25.10.2025	Виконано
4	Розроблення інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями	25.10.2025	05.11.2025	Виконано
5	Економічна частина	05.11.2025	15.11.2025	Виконано
6	Оформлення матеріалів до захисту МКР	15.11.2025	25.11.2025	Виконано

Студентка _____ Анна СУДЕЦЬ

Керівник роботи _____ Ігор ШТЕЛЬМАХ

АНОТАЦІЯ

УДК 004.09

Судець А. О. Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями. Магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2025. 110 с.

На укр. мові. Бібліогр.: 28 назв; рис.: 49; табл.: 10.

У магістерській кваліфікаційній роботі розглянуто проблему автоматизації аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями, що є важливим завданням у розвитку сучасних технологій обслуговування та підтримки управлінських рішень. Досліджено методи машинного навчання та глибинного навчання, які забезпечують можливість опрацювання великих обсягів пасажирських даних і виявлення ключових чинників, що впливають на оцінку сервісу.

Запропонована інформаційна технологія ґрунтується на поєднанні моделей машинного навчання та багатошаровими нейронними мережами різної глибини, що дозволяє підвищити точність класифікації рівнів задоволеності. Розроблені методи спрямовані на покращення процесів аналізу та підвищення точності передбачення.

Ілюстративна частина складається з 6 плакатів із результатами моделювання.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями.

Ключові слова: аналіз, передбачення, задоволеність, технологія, автоматизація.

ABSTRACT

Sudets A.O. Information technology of analysis and transfer of equal passenger requests by airlines. Master's qualification work on specialties 126 - information systems and technologies, introductory professional program - information technologies of data and image analysis. Vinnytsia: VNTU, 2025. 110 p.

In Ukrainian speech Bibliography: 28 titles; Fig.: 49; tab.: 10.

The master's qualification paper now discusses the issue of automating the analysis and transfer of equal requirements of passengers by airlines, which is important to consider in the development of modern technologies for service and management support. solutions The methods of machine and deep penetration have been developed, which exclude the possibility of working with large volumes of passenger data and attracting key clients that influence the service site.

The information technology used for matching machine learning models with large neural capabilities of different depths is proposed, which allows to increase the accuracy of level classifications. satisfaction Methods aimed at accelerating the analysis process, forming forecasts and increasing the efficiency of aviation services have been developed.

The result of the work can be applied in information technologies of airlines, passenger file monitoring services and planning tools for improvement in the field of aviation services.

The object of consideration is the process of developing information technology for the analysis and transfer of equal passenger satisfaction.

The field of application is information technologies for passenger service, analytical services and fast reception systems in the aviation industry.

Illustrative partial assembly of 6 posters as a result of modeling.

In the section of the economic part, information is published about the details of the details and the introduction of information technology for the analysis and transfer of equal satisfaction of passengers by airlines.

Key words: analysis, prediction, programming, technology, automation

ЗМІСТ

ВСТУП

Актуальність дослідження зумовлена трансформацією української авіаційної галузі в умовах повномасштабної війни. Попри закритий повітряний простір, українські авіакомпанії продовжують роботу за кордоном, виконують міжнародні рейси та зберігають українську реєстрацію. Завдяки цьому вони й надалі сплачують податки в Україні та підтримують національну економіку. Це підкреслює важливість розвитку сучасних інформаційних технологій для підвищення якості обслуговування та задоволеності пасажирів у майбутньому відновленні галузі.

Мета і завдання роботи. Метою дослідження є підвищення точності передбачення рівнів задоволеності авіакомпаніями, шляхом створення інформаційної технології та застосування сучасних методів машинного навчання і багатошарових нейронних мереж.

Для досягнення поставленої мети потрібно виконати такі завдання:

– Провести аналіз предметної області та оцінити актуальність задачі передбачення задоволеності пасажирів авіакомпаній;

– Здійснити вибір оптимальних інформаційних технологій, моделей та методів машинного навчання для розв’язання задачі класифікації;

– Провести підготовку даних та розвідувальний аналіз;

– Розробити архітектурну модель та алгоритм роботи інформаційної технології;

– Розробити інформаційну технологію, яка використовує моделі машинного навчання та багатошарові нейронні мережі для здійснення передбачення рівнів задоволеності пасажирів авіакомпаніями;

– Оцінити результати роботи моделей.

Об’єктом дослідження магістерської кваліфікаційної роботи є процес розроблення інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями.

Предметом дослідження магістерської кваліфікаційної роботи є методи інформаційних технологій, що застосовуються для аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями.

Новизна одержаних результатів дістала подальший розвиток шляхом удосконалення інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями. Подальший розвиток забезпечено шляхом комплексного використання методів машинного навчання інтелектуальних моделей та багатошарових нейронних мереж різної глибини. Така інтеграція класичних алгоритмів і глибинних моделей дала змогу підвищити точність передбачення та ефективно обробляти комплексні залежності у даних пасажирських оцінок.

Практичне значення. Результати магістерської кваліфікаційної роботи можуть бути використані для автоматизації процесу аналізу та передбачення рівнів задоволеності пасажирів авіакомпаній. Запропонована технологія дозволяє підвищити точність оцінювання якості обслуговування та оптимізувати роботу авіасервісів. Застосування моделей машинного навчання та нейронних мереж дає можливість авіакомпаніям своєчасно виявляти проблемні аспекти, приймати обґрунтовані управлінські рішення та покращувати взаємодію з пасажирами.

Апробація та публікації результатів магістерської кваліфікаційної роботи. Опубліковано тези на «LV Всеукраїнській науково-технічній конференції підрозділів Вінницького національного технічного університету (ВНТКП ВНТУ)» [1].

1. ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ

1.1 Аналіз предметної області

Сфера авіаперевезень є однією з найбільш динамічних та конкурентних галузей сучасної економіки, що постійно зазнає впливу глобалізації, технологічного прогресу та зміни споживчих потреб. У цій сфері якість обслуговування пасажирів та рівень їх задоволеності безпосередньо впливають на репутацію авіакомпанії, її конкурентоспроможність, фінансові показники та довгострокову стійкість на ринку. Саме тому аналіз і передбачення факторів, що визначають ступінь задоволеності пасажирів, є одним із ключових завдань у системі управління якістю послуг та стратегічного планування діяльності авіаперевізників [2].

Рівень задоволеності пасажирів формується під впливом комплексу взаємопов'язаних чинників. Серед них можна виділити такі основні групи:

– Комфорт під час польоту – зручність і ергономіка сидінь, чистота салону, наявність сучасних розважальних систем, якість харчування та напоїв, температурний режим у салоні. Комфорт безпосередньо впливає на загальне враження пасажира від перельоту та формує його емоційний стан;

– Якість обслуговування персоналом – доброзичливість, професійність, оперативність реагування на запити, здатність вирішувати конфліктні ситуації та індивідуальний підхід до кожного пасажира. Служба обслуговування є обличчям авіакомпанії і безпосередньо впливає на формування лояльності клієнтів;

– Технічні та організаційні аспекти – дотримання розкладу, своєчасна реєстрація, обробка багажу, наявність додаткових сервісів (Wi-Fi, розваги на борту), забезпечення безпеки та дотримання правил авіаційного регламенту. Надійність і точність організації польоту значно впливають на задоволеність пасажирів, особливо в умовах високої конкуренції;

– Цінова політика та система лояльності – адекватність тарифів, прозорість цін, наявність бонусних програм, акцій та знижок для постійних

клієнтів. Вартість квитка та економічна вигідність перельоту формують пряме сприйняття цінності послуги;

– Загальні враження від взаємодії з авіакомпанією до, під час і після польоту – включають процес покупки квитка, роботу контакт-центрів, доступність інформації про рейси, підтримку при виникненні непередбачених ситуацій та зворотний зв’язок після перельоту.

Оцінка задоволеності пасажирів здійснюється за допомогою різних методів, серед яких анкетування на борту та онлайн, опитування після завершення рейсу, аналіз відгуків у соціальних мережах та на спеціалізованих платформах. Зібрані дані можуть мати як числові, так і текстові характеристики, що дозволяє враховувати як кількісні, так і якісні аспекти сприйняття послуги. Детальний аналіз цих даних дозволяє виділити ключові фактори, що впливають на позитивне чи негативне ставлення пасажирів, та визначити пріоритети для поліпшення сервісу.

Сучасні інформаційні технології аналізу задоволеності пасажирів спрямовані на автоматизацію обробки великих обсягів даних, виявлення закономірностей та тенденцій у поведінці клієнтів, а також на побудову прогнозних моделей, які дозволяють передбачати рівень задоволеності на основі визначених параметрів. Такі технології дозволяють авіакомпаніям приймати обґрунтовані управлінські рішення, оптимізувати внутрішні процеси, підвищувати якість обслуговування та покращувати клієнтський досвід [3].

Таким чином, розвиток і впровадження сучасних інформаційних технологій для аналізу та передбачення рівнів задоволеності пасажирів є важливою складовою конкурентної стратегії авіакомпаній. Вони сприяють підвищенню рівня сервісу, зміцненню довіри клієнтів, формуванню лояльності та забезпечують довгостроковий розвиток і стійкість авіаційної галузі навіть у умовах економічних і соціальних викликів сучасності.

1.2 Аналіз актуальності

Повномасштабна війна, розпочата Російською Федерацією проти України у 2022 році, завдала значних втрат усім галузям економіки, зокрема й авіаційній. З перших днів вторгнення повітряний простір України було повністю закрито для цивільної авіації, що фактично зупинило діяльність вітчизняних авіакомпаній, призвело до фінансових збитків, втрати літаків та робочих місць. Попри це, український авіаційний ринок не зник, а трансформувався та почав шукати нові можливості для розвитку.

Закриття неба змусило авіакомпанії перебудовувати логістичні ланцюги, активно використовуючи європейські хаби. Аеропорти сусідніх держав, таких як Польща, Угорщина, Словаччина та Румунія, отримали значні переваги від перерозподілу пасажиропотоків, а обсяги перевезень у прикордонних регіональних аеропортах, зокрема Жешув, Люблін та Будапешт, за останні роки відчутно зросли. Це дало поштовх до розвитку інфраструктури, залучення інвестицій та створення нових робочих місць, у тому числі для українських фахівців, які змушені були виїхати за кордон. Зростання попиту на авіаперевезення в цих країнах створює нові економічні зв’язки, а українські працівники авіаційної сфери активно долучаються до їх реалізації [4].

Попри закриття неба над Україною, частина українських авіакомпаній змогла продовжити роботу на міжнародному рівні. Найуспішнішою серед них є SkyUp, яка перенесла операційну діяльність до країн ЄС і виконує чартерні та ACMI-рейси, перевозячи пасажирів для європейських перевізників під власним екіпажем і флотом. У 2024 році компанія здійснила понад 6,5 тисяч рейсів і перевезла понад мільйон пасажирів у 67 країн світу, сплачуючи при цьому податки в Україні. Інші компанії, такі як Windrose та Skyline Express, також зберегли українську реєстрацію, що сприяє наповненню державного бюджету та підтримці економіки навіть під час війни. Динаміка українських авіакомпаній після 2022 року демонструє гнучкість і здатність адаптуватися до нових умов. Незважаючи на складнощі, український авіаційний бізнес поступово інтегрується в європейський

ринок, розширює партнерські зв’язки та формує позитивну репутацію завдяки професійності персоналу та високому рівню обслуговування. Це свідчить про потенціал галузі до відновлення після завершення воєнних дій.

Зміни торкнулися й структури вантажних потоків. Близько 45 % перевезених вантажів становлять гуманітарні та медичні матеріали, ще близько 30 % — військова техніка та запчастини. Авіаційна логістика залишилася ключовим каналом доставки критично важливих вантажів, що забезпечує стабільність постачання навіть в умовах війни. Крім того, авіакомпанії активно інвестують у підвищення кваліфікації персоналу, оскільки підтримка пілотів, бортпровідників та технічних фахівців є критичною для виживання галузі. Наприклад, SkyUp проводить підготовку екіпажів на міжнародних базах, а значна частина персоналу пройшла європейську атестацію.

Незважаючи на всі виклики, перспективи української авіації залишаються значними. Компанії не лише зберігають флот та персонал, але й формують нові бізнес-моделі, інтегруючись у європейський ринок і підвищуючи міжнародну конкурентоспроможність. Авіація залишається каналом для гуманітарних перевезень, критичних логістичних потоків і міжнародної співпраці. Водночас, дослідження задоволеності пасажирів та оптимізація сервісу набувають ще більшої актуальності, адже якість обслуговування може стати ключовою конкурентною перевагою при відновленні українського авіаринку [5].

На світовому рівні, за даними Міжнародної асоціації повітряного транспорту (IATA), у 2024–2025 роках спостерігається стабільне зростання попиту на авіаперевезення, а середній прибуток на одного пасажира зріс із 6,4 до 7 доларів, а у деяких регіонах, таких як Близький Схід, — до 23,9 доларів[6]. Це підтверджує загальну тенденцію до відновлення авіаційної галузі після пандемії та воєнних потрясінь.

На рисунку 1.1 зображено передбачення для світового ринку пасажирських авіаперевезень.

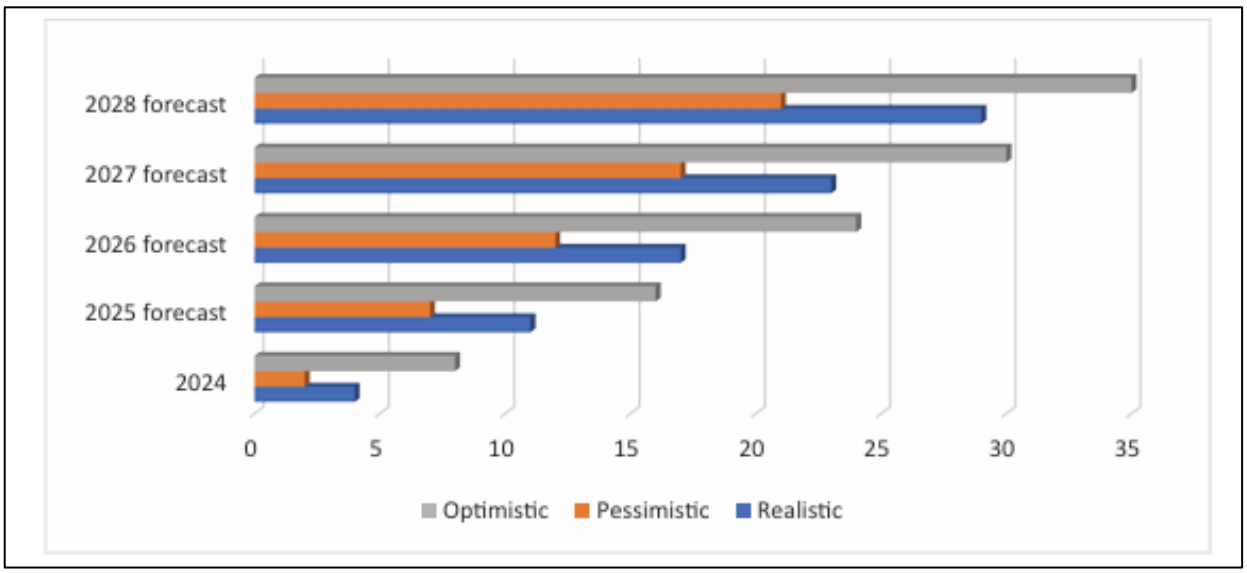


Рисунок 1.1 – Передбачення для світового ринку пасажирських авіаперевезень

Відповідно, зростає потреба у впровадженні нових інформаційних технологій для підвищення ефективності, якості обслуговування та задоволеності пасажирів, що є актуальним і для майбутнього розвитку українського авіаційного сектору після відкриття повітряного простору.

1.3 Вибір оптимальних інформаційних технологій

Для реалізації поставленого завдання було обрано платформу Kaggle, яка є одним із провідних середовищ для аналізу даних, розробки моделей штучного інтелекту та глибокого навчання. Kaggle забезпечує доступ до потужних обчислювальних ресурсів, включаючи GPU та TPU, що дозволяє прискорити процес навчання складних моделей і виконувати розрахунки на великих обсягах даних. Платформа має зручне середовище для розробки, інтегроване з Jupyter Notebook, що спрощує написання та тестування коду, а також активну спільноту фахівців з усього світу, яка ділиться рішеннями, кодом та підходами до вирішення практичних задач [7].

Крім цього, Kaggle пропонує широкі можливості для роботи з наборами даних, їх обробки та візуалізації результатів. Платформа дозволяє завантажувати власні набори даних, а також користуватися загальнодоступними датасетами з

різних сфер, що є особливо корисним при виконанні аналітичних завдань або тестуванні моделей машинного навчання на різних типах даних. Високий рівень доступності інструментів для побудови та навчання нейронних мереж робить Kaggle оптимальним вибором для проведення даного дослідження.

На рисунку 1.2 зображено головну сторінку платформи.

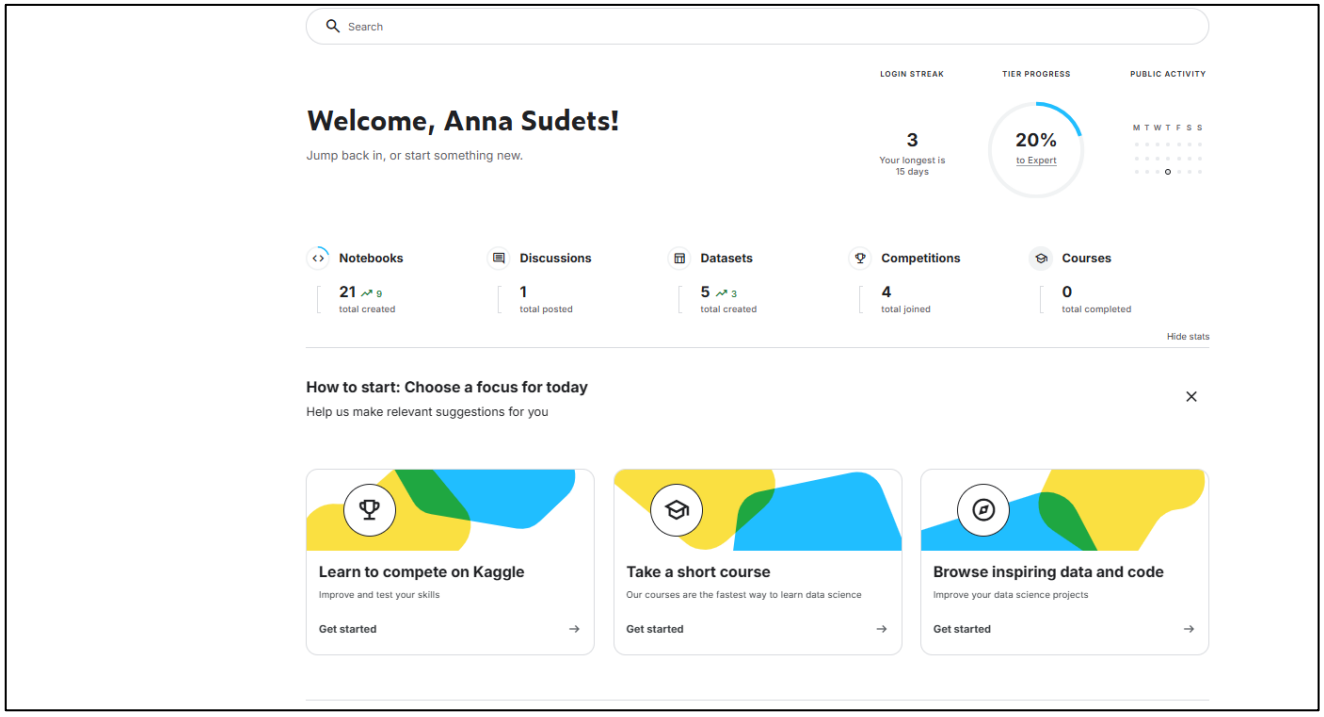


Рисунок 1.2 – Головна сторінка Kaggle

До переваг використання Kaggle слід віднести можливість роботи з великими обсягами даних без необхідності купівлі власного потужного обладнання, інтеграцію з сучасними бібліотеками машинного навчання та глибокого навчання, наявність готових прикладів рішень та спільнот для обговорення, а також зручний інтерфейс для візуалізації результатів, що спрощує аналіз даних та побудову графіків. Крім того, платформа підтримує збереження версій ноутбуків і результатів експериментів, що дозволяє відстежувати прогрес роботи та повторно використовувати коди при подальших дослідженнях [8].

Серед недоліків можна виділити обмеження в безкоштовному тарифі щодо обчислювальних ресурсів, що може бути критично для навчання дуже великих

нейронних мереж, а також обмеження на час виконання одного сеансу коду, яке вимагає оптимізації процесів та розбиття задач на етапи. Крім того, робота з зовнішніми базами даних іноді потребує додаткових зусиль для інтеграції з платформою, а також доступ до приватних корпоративних даних обмежений [8].

Python – основна мова програмування, яка використовується на платформі Kaggle для побудови моделей машинного навчання та аналізу даних. Вона є високорівневою мовою загального призначення, яка активно застосовується у науці про дані, штучному інтелекті та глибокому навчанні. Python забезпечує простий та зрозумілий синтаксис, широкий вибір бібліотек і модулів для роботи з даними, статистики та нейронних мереж, а також дозволяє швидко створювати, тестувати та вдосконалювати моделі. Завдяки Python користувачі Kaggle можуть легко інтегрувати різні інструменти і бібліотеки (TensorFlow, Keras, PyTorch, scikit-learn) у свої проєкти та ефективно реалізовувати задачі передбачення та класифікації.

Серед основних переваг використання Python слід виділити:

– простий і зрозумілий синтаксис, що полегшує навчання та швидке написання коду;

– велика кількість бібліотек і готових модулів для аналізу даних, машинного навчання та візуалізації;

– активна спільнота розробників, яка створює документацію, приклади коду та навчальні матеріали;

– кросплатформеність, що дозволяє запускати програми на різних операційних системах;

– інтеграція з популярними платформами для аналізу даних, такими як Kaggle, Google Colab, Jupyter Notebook;

– можливість швидкого прототипування моделей і експериментів, що особливо важливо для наукових досліджень та тестування нових підходів.

Серед недоліків Python можна відзначити:

– нижчу швидкість виконання порівняно з компільованими мовами (C/C++), що іноді обмежує застосування у великих обчислювальних завданнях;

– високе споживання пам’яті при роботі з великими наборами даних;

– динамічне типізування може призводити до помилок, які складніше відслідковувати на етапі компіляції;

– залежність від сторонніх бібліотек для реалізації багатьох спеціалізованих функцій, що може ускладнювати встановлення та оновлення середовища.

В цілому, Python залишається найбільш зручною та гнучкою мовою для роботи з аналізом даних та побудовою моделей машинного навчання на платформі Kaggle, поєднуючи простоту використання з широкими можливостями для розробки ефективних рішень [9].

Нейронні мережі є одним із найпотужніших інструментів сучасного штучного інтелекту. Вони імітують роботу біологічних нейронів, дозволяючи системі навчатися на прикладах і самостійно виявляти закономірності у великих обсягах даних. Кожен нейрон приймає вхідні сигнали, зважує їх за певними коефіцієнтами та передає далі до наступного шару. Після багаторазового проходження даних мережею відбувається навчання — налаштування ваг так, щоб результати відповідали очікуванням. Це робить нейронні мережі особливо ефективними для вирішення складних задач передбачення та класифікації, де класичні статистичні методи можуть давати недостатньо точні результати.

Одним із найпоширеніших типів нейронних мереж є багатошаровий перцептрон (MLP, Multi-Layer Perceptron). Це модель, що складається з вхідного, одного або кількох прихованих та вихідного шарів. Кожен прихований шар дозволяє моделі навчатися більш складним залежностям і взаємозв’язкам між даними. Багатошарові мережі особливо ефективні для задач класифікації, передбачення та розпізнавання шаблонів, що робить їх доцільним вибором для задачі визначення рівня задоволеності пасажирів авіакомпаніями [10].

Для реалізації нейромережевих моделей у даному дослідженні було обрано дві популярні бібліотеки. Keras, як високорівневий інтерфейс до TensorFlow, значно спрощує побудову, тренування та оцінку моделей глибокого навчання. Вона забезпечує зручний синтаксис для створення архітектур нейронних мереж різної

складності, а також підтримує використання GPU для прискорення обчислень. Scikit-learn, у свою чергу, є бібліотекою машинного навчання, що містить широкий набір інструментів для класифікації, регресії, кластеризації, оцінки якості моделей і попередньої обробки даних. Вона ідеально підходить для реалізації моделей типу MLP, проведення порівняльного аналізу та визначення оптимальних параметрів навчання [10].

Важливо зазначити, що у бакалаврській дипломній роботі для передбачення рівнів задоволеності пасажирів авіакомпаніями використовувалися різноманітні методи машинного навчання, серед яких Naive Bayes, Decision Tree, Random Forest, Extra Trees, K-Nearest Neighbors (KNN), Logistic Regression, AdaBoost, Gradient Boosting та LGBM.

На рисунку 1.3 зображено результати бакалаврської дипломної роботи.

	Model	Train Precision	Test Precision
0	Naive Bayes	0.843150	0.840626
1	Decision Tree	0.918550	0.913897
2	Random Forest	0.922290	0.916335
3	Extra Trees	0.926855	0.918003
4	KNN	0.938789	0.907994
5	Logistic Regression	0.872023	0.868215
6	AdaBoost	0.916075	0.913127
7	GradientBoost	0.924655	0.915822
8	LGBM	0.982511	0.949313

Рисунок 1.3 – Результати бакалаврської дипломної роботи

Використання цих методів дозволило оцінити точність передбачення та визначити моделі, які найбільш ефективно справлялися із поставленим завданням класифікації.

Для магістерської роботи було обрано додатково три моделі машинного навчання: XGBoost, Gaussian Process Classifier (GPC) та Ridge Classifier, що

дозволить розширити порівняльний аналіз та підвищити точність передбачення задоволеності пасажирів.

XGBoost (Extreme Gradient Boosting) — це один із найефективніших сучасних алгоритмів градієнтного бустингу, який поєднує велику кількість слабких моделей, найчастіше дерев рішень, у потужну ансамблеву модель. Основна ідея XGBoost полягає у поетапному побудуванні дерев, де кожне наступне дерево намагається компенсувати помилки попередніх, поступово покращуючи якість передбачення. На відміну від базових методів бустингу, XGBoost використовує оптимізовані механізми роботи з даними, включно з розумною обробкою пропусків, роботою з нерівномірно розподіленими ознаками та можливістю автоматично визначати напрямок для відсутніх значень. Крім того, алгоритм підтримує регуляризацію, що допомагає уникати перенавчання, роблячи модель більш узагальнюючою і стійкою до шуму. Його архітектура орієнтована на високу продуктивність: XGBoost використовує паралельні обчислення, оптимізовані структури даних та спеціальні методи для прискорення навчання навіть на великих і складних датасетах. Серед ключових переваг XGBoost можна виділити здатність моделювати дуже складні, нелінійні взаємозв’язки між ознаками, гнучкість у налаштуванні великої кількості параметрів, а також високу точність, яка часто перевершує інші алгоритми. Водночас цей метод має і певні недоліки: він може бути обчислювально важким при неправильному налаштуванні, потребує досвіду у виборі гіперпараметрів, а також менш інтерпретований порівняно з лінійними моделями або простими деревами [11].

На рисунку 1.4 зображено структуру моделі XGBoost.

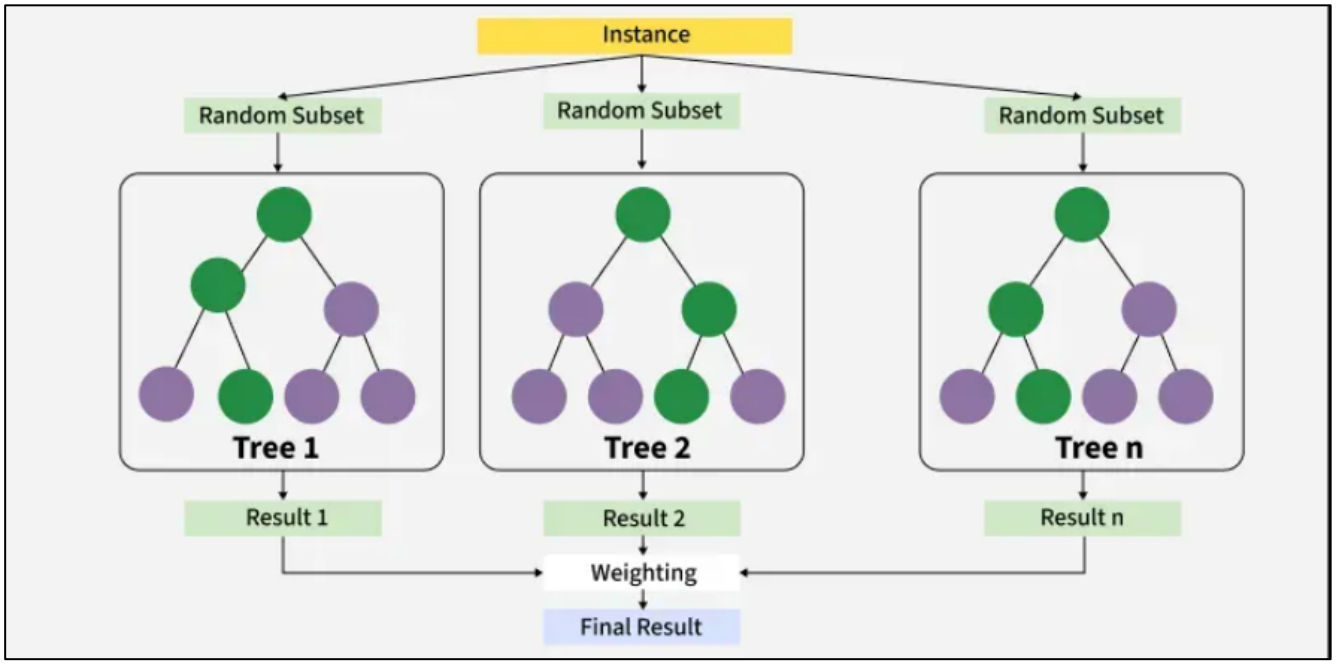


Рисунок 1.4 – Структура моделі XGBoost

У контексті класифікації рівня задоволеності пасажирів авіакомпаній XGBoost проявляє себе як надзвичайно ефективний інструмент, здатний точно виявляти складні закономірності у поведінці та характеристиках клієнтів. Завдяки своїй здатності глибоко аналізувати структурні взаємозв'язки між ознаками модель може виявляти приховані закономірності, які залишаються непоміченими більш простими алгоритмами. Це дозволяє суттєво підвищити точність передбачення, підтримує боротьбу з перенавчанням і робить XGBoost одним із найрезультативніших методів у задачах аналізу задоволеності пасажирів та побудови рекомендацій для покращення сервісу [11].

Gaussian Process Classifier (GPC) — це ймовірнісний метод машинного навчання, який використовує апарат гаусівських процесів для моделювання прихованої функції, що відображає ймовірність належності об'єкта до певного класу. Основна ідея полягає в тому, що замість жорсткого визначення межі між класами, як це роблять класичні класифікатори, GPC припускає наявність нескінченновимірного простору можливих функцій і обирає серед них ті, що найбільш узгоджуються з даними. Завдяки цьому GPC генерує не просто рішення «клас 0» або «клас 1», а повний розподіл ймовірностей, що дозволяє оцінити

ступінь невизначеності прогнозу. Під час навчання GPC використовує коваріаційні (kernel) функції для аналізу подібності між об’єктами, що дає змогу моделювати складні, нелінійні залежності в даних. Такий підхід робить Gaussian Process Classifier особливо корисним у задачах, де важлива не лише точність передбачення, але й розуміння того, наскільки модель упевнена у своїх рішеннях. До ключових переваг GPC належить висока інтерпретованість у частині невизначеності прогнозів, гнучкість завдяки використанню різних ядрових функцій, а також здатність добре працювати на малих вибірках, де інші алгоритми часто втрачають стабільність. Водночас метод має і певні недоліки: він є обчислювально дорогим, оскільки складність навчання зростає кубічно зі збільшенням кількості об’єктів, що обмежує застосування на дуже великих датасетах; потребує ретельного підбору ядра та його параметрів; а також може погано масштабуватися при високій розмірності простору ознак [12].

У контексті дослідження задоволеності пасажирів авіакомпаній GPC є цінним інструментом, оскільки дозволяє не лише класифікувати пасажирів на задоволених і незадоволених, але й оцінити ступінь упевненості моделі у кожному прогнозі. Така інформація є корисною при аналізі ризиків, формуванні рекомендацій та підтримці прийняття рішень, оскільки дає змогу виявляти групи пасажирів, для яких модель найбільш або найменш впевнена, а отже — потенційно проблемні або неоднозначні випадки.

Ridge Classifier є лінійним методом класифікації, який працює на основі підходу найменших квадратів із використанням L2-регуляризації. Його ключова ідея полягає в тому, щоб знайти оптимальні ваги для кожної ознаки, але водночас обмежити їхнє надмірне зростання. Завдяки регуляризації модель стає стійкішою до шуму, випадкових коливань у даних та мультиколінеарності, що часто виникає при використанні великої кількості категоріальних ознак після кодування. Ridge Classifier фактично перетворює задачу класифікації у задачу регресії, прогнозує проміжне числове значення, а потім відносить його до певного класу, що забезпечує стабільність та надійність результатів.

Структурно цей метод належить до простих та інтерпретованих моделей, де кожній ознаці відповідає певний коефіцієнт, що відображає силу її впливу на результат. Це робить Ridge Classifier корисним інструментом у тих випадках, коли важливо не лише отримати прогноз, а й зрозуміти, які саме фактори впливають на рішення моделі. Особливо ефективним він є на високорозмірних наборах даних, де кількість ознак значно перевищує кількість спостережень, що є типовою ситуацією при роботі з даними, отриманими після кодування текстових та категоріальних параметрів.

До переваг Ridge Classifier належать простота реалізації, висока швидкість навчання та стійкість до перенавчання, що досягається завдяки регуляризації. Він добре справляється з даними, у яких присутні корельовані ознаки, та зберігає стабільність навіть у випадку незначного шуму. Крім того, модель легко інтерпретується, що дозволяє зрозуміти внесок кожного параметра в підсумковий прогноз.

Разом із тим Ridge Classifier має і певні недоліки. Оскільки це лінійна модель, вона не може ефективно відтворювати складні, нелінійні залежності між ознаками, що може знижувати її точність у порівнянні з більш гнучкими методами, такими як XGBoost або багатошарові нейронні мережі. Також модель є чутливою до масштабу ознак, тому перед навчанням необхідно проводити їх стандартизацію. У разі сильно незбалансованих класів Ridge Classifier може зміщуватися у бік домінуючої групи, якщо не застосовувати додаткові техніки балансування [13].

У контексті дослідження задоволеності пасажирів Ridge Classifier виступає надійною базовою моделлю, яка дозволяє швидко отримати стабільні результати та провести порівняльний аналіз із більш складними алгоритмами. Такий підхід допомагає визначити роль окремих ознак і оцінити, наскільки вони впливають на формування оцінки пасажирами авіаційного сервісу.

Враховуючи, що завдання дослідження полягає у класифікації рівня задоволеності пасажирів, у магістерській роботі було застосовано поєднання нейронних мереж і сучасних алгоритмів машинного навчання. Зокрема, розроблено моделі нейронних мереж із різною кількістю прихованих шарів на базі бібліотеки

Keras, а також реалізовано багатошаровий перцептрон (MLP) із використанням scikit-learn. Для розширення порівняльного аналізу та підвищення точності класифікації додатково використано три продуктивні методи машинного навчання — XGBoost, Ridge Classifier та Gaussian Process Classifier (GPC). Кожен із цих алгоритмів має власні структурні особливості та переваги: XGBoost забезпечує високоточне моделювання складних нелінійних залежностей, Ridge Classifier виступає стабільною та інтерпретованою лінійною моделлю, а GPC дозволяє оцінювати не лише прогноз, а й ступінь невизначеності рішення [13].

Застосування такого різнопланового інструментарію дає змогу комплексно оцінити поведінку моделей за різних характеристик даних і визначити найефективніші підходи до передбачення рівня задоволеності пасажирів.

Використання платформи Kaggle для роботи з даними, мови програмування Python та бібліотек Keras і scikit-learn разом із сучасними алгоритмами XGBoost, Ridge Classifier і GPC створить оптимальне технічне середовище, яке забезпечить високу якість, точність і надійність отриманих результатів дослідження.

1.4 Огляд готових рішень для передбачення

У даному розділі представлено огляд існуючих рішень та підходів, реалізованих іншими дослідниками й розробниками для задачі передбачення. Аналіз таких методів є важливою складовою наукової роботи, оскільки дозволяє сформувати цілісне уявлення про вже накопичений досвід у цій сфері, а також оцінити ефективність різних алгоритмів, технік обробки даних та стратегій моделювання. Вивчення попередніх рішень допомагає визначити рівень складності типових задач, зрозуміти, які інструменти показали себе найкраще, а які підходи продемонстрували обмеження або невисоку точність.

Огляд існуючих робіт також дає змогу виявити найпоширеніші проблеми, з якими стикаються дослідники під час побудови моделей: це можуть бути питання дисбалансу класів, недостатньої кількості якісних даних, складність у виділенні інформативних ознак або ризик перенавчання на складних багатовимірних

вибірках. Крім того, порівняння різних методів дозволяє побачити їхні переваги та недоліки, зокрема інтерпретованість, обчислювальну вартість, стійкість до шуму, здатність моделювати нелінійні залежності чи працювати на великих наборах даних.

Такий аналіз є корисним не лише з точки зору систематизації знань, але й для формування власної методології дослідження. Він дає можливість обґрунтувати вибір конкретних моделей і технологій, визначити, які інструменти доцільно застосовувати у межах магістерської роботи, а яких варто уникати, враховуючи недоліки, виявлені у попередніх дослідженнях. У результаті огляд наявних рішень допомагає побудувати більш ефективну та обґрунтовану модель передбачення, спираючись на сильні сторони існуючих підходів та уникаючи повторення типових помилок, що значно підвищує якість і надійність кінцевих результатів.

У якості прикладу проаналізовано декілька готових рішень, реалізованих за допомогою нейронної мережі.

Розглянемо одне з таких рішень під назвою ANN From Scratch using Numpy | 81% | EDA (рис. 1.5) [14].

```
def forward_propagation(X, parameters):
    W1 = parameters["W1"]
    b1 = parameters["b1"]
    W2 = parameters["W2"]
    b2 = parameters["b2"]

    Z1 = np.dot(W1, X) + b1
    A1 = np.tanh(Z1)

    Z2 = np.dot(W2, A1) + b2
    A2 = 1 / (1 + np.exp(-Z2))

    cache = {"Z1": Z1, "A1": A1, "Z2": Z2, "A2": A2}
    return A2, cache

hidden_size = 25
output_size = 1
```

Рисунок 1.5 – Приклад коду

На рисунку 1.6 зображено результат передбачення.

```
train_accuracy = accuracy(y_train, predictions_train.squeeze())
print(f"Training Accuracy: {train_accuracy}")
```

```
Training Accuracy: 0.8145307206652295
```

```
predictions_test = predict(parameters, X_test_processed_boxcox.T)
```

```
test_accuracy = accuracy(y_test, predictions_test.squeeze())
print(f"Training Accuracy: {test_accuracy}")
```

```
Training Accuracy: 0.8147135817677856
```

Рисунок 1.6 – Результат передбачення

Незважаючи на те, що код конкурента реалізує повний цикл роботи штучної нейронної мережі — від ініціалізації параметрів до передбачення та оцінки точності — і демонструє базове розуміння архітектури моделей, у ньому бракує кількох важливих аспектів, що роблять модель більш надійною й адаптивною. Наприклад, використання функції активації «tanh» у прихованому шарі дійсно може бути виправданим, однак відсутність Dropout, а також фіксоване число нейронів (25) без обґрунтування — знижує гнучкість моделі. Тренувальна та тестова точність вийшли приблизно однаковими, але це могло трапитися випадково, оскільки автор не навів графіків втрат чи валідаційної точності, а отже — не аналізував можливе перенавчання. Тобто, попри коректну реалізацію структури мережі, відсутність належного аналізу її поведінки на різних етапах навчання та відсутність засобів боротьби з перенавчанням залишають цю реалізацію базовою, з потенціалом до покращення [14].

Далі розглянемо готове рішення під назвою 93% Accuracy with Neural Network (рис. 1.7) [15].

```
model = Sequential()
model.add(Input(shape=(X_train.shape[1],)))
# Add more layers here
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(32, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(16, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(8, activation='relu'))

# Output layer
model.add(Dense(1, activation='relu'))
```

```
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

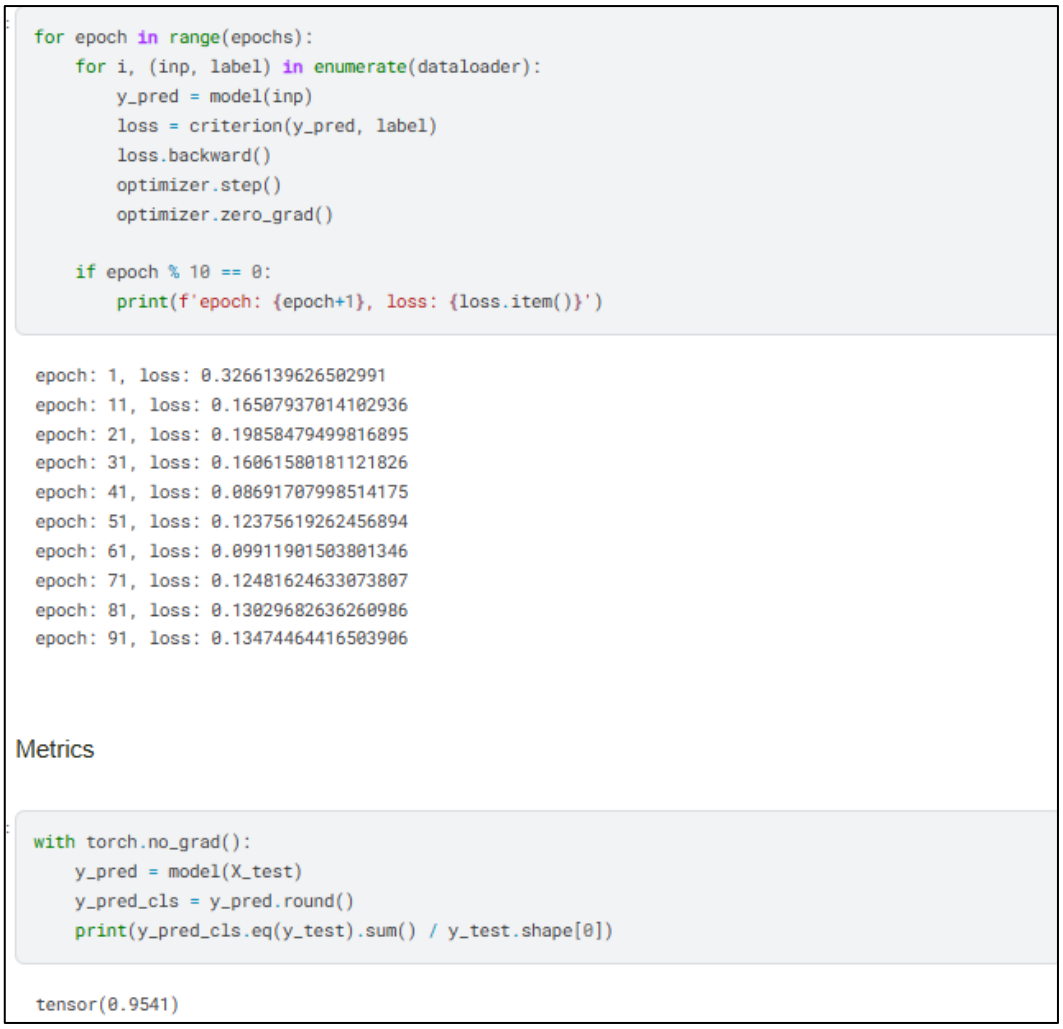
```
history = model.fit(X_train, y_train, epochs=50, batch_size=10, validation_data=(X_test, y_test))
```

Рисунок 1.7 – Приклад коду

Код конкурента демонструє використання багатошарової нейронної мережі на основі Keras з поступовим зменшенням кількості нейронів у кожному шарі та активацією “relu”, що є ефективним підходом для побудови глибокої моделі. Також у кожному прихованому шарі використовується Dropout зі значенням 0.5, що знижує ризик перенавчання та покращує узагальнювальну здатність моделі. Модель тренується з використанням валідаційного набору, що дозволяє контролювати її якість під час навчання. Отримана точність на тестовому наборі даних є високою — близько 93%, що свідчить про достатню ефективність обраної архітектури.

Проте в коді допущена помилка у вихідному шарі: використано функцію активації “relu” замість “sigmoid”, яка є більш доречною для бінарної класифікації, оскільки нормалізує вихід у межах від 0 до 1. Це може призвести до некоректної інтерпретації результатів класифікації. Також відсутні графіки динаміки втрат або точності, що ускладнює оцінку стабільності навчання.

Далі розглянемо ще один приклад під назвою Airline Satisfaction Prediction w/ PyTorch (рис. 1.8) [16].



```
for epoch in range(epochs):
    for i, (inp, label) in enumerate(dataloader):
        y_pred = model(inp)
        loss = criterion(y_pred, label)
        loss.backward()
        optimizer.step()
        optimizer.zero_grad()

    if epoch % 10 == 0:
        print(f'epoch: {epoch+1}, loss: {loss.item()}')
```

```
epoch: 1, loss: 0.3266139626502991
epoch: 11, loss: 0.16507937014102936
epoch: 21, loss: 0.19858479499816895
epoch: 31, loss: 0.16061580181121826
epoch: 41, loss: 0.08691707998514175
epoch: 51, loss: 0.12375619262456894
epoch: 61, loss: 0.09911901503801346
epoch: 71, loss: 0.12481624633073807
epoch: 81, loss: 0.13029682636260986
epoch: 91, loss: 0.13474464416503906
```

Metrics

```
with torch.no_grad():
    y_pred = model(X_test)
    y_pred_cls = y_pred.round()
    print(y_pred_cls.eq(y_test).sum() / y_test.shape[0])
```

```
tensor(0.9541)
```

Рисунок 1.8 – Приклад коду та результат

У даному прикладі конкурент створив модель на основі PyTorch. Сам процес навчання реалізований вручну: модель отримує дані, рахує помилку, оновлює свої ваги й так робить багато разів. На тестових даних точність склала 95,4%, що є дуже хорошим результатом. Але в коді не показано, як саме виглядає структура нейромережі: скільки в ній шарів і що в них відбувається. Також немає графіків, які показували б, як змінюється помилка чи точність під час навчання — це важливо, щоб зрозуміти, чи модель не "завчила" дані занадто сильно. Ще один недолік — не використано методи, які допомагають уникнути перенавчання (наприклад, Dropout).

Проведений аналіз існуючих рішень, реалізованих іншими розробниками, демонструє широкий спектр підходів до побудови нейронних мереж та значні відмінності у якості їх реалізації. Розглянуті приклади показали, що навіть за наявності правильно побудованої архітектури критично важливо враховувати нюанси навчання моделі — такі як вибір функцій активації, кількість нейронів у шарах, використання регуляризаційних технік, аналіз навчальних кривих та боротьба з перенавчанням. Навіть невеликі помилки або спрощення можуть суттєво вплинути на якість передбачення, що підкреслює важливість уважної та обґрунтованої побудови нейромережевих моделей.

Разом із тим огляд показав, що нейронні мережі далеко не єдиний ефективний підхід до задачі класифікації задоволеності пасажирів. У багатьох випадках альтернативні методи машинного навчання — такі як градієнтний бустинг, гаусівські процеси чи лінійні класифікатори з регуляризацією — можуть забезпечувати не меншу, а інколи й вищу точність, бути обчислювально вигіднішими або більш інтерпретованими. Це створює підґрунтя для розширення дослідження за межі виключно нейромережевих рішень та порівняння їх з іншими підходами.

Саме тому наступним кроком є детальний огляд сучасних методів машинного навчання, що можуть бути застосовані для задачі передбачення рівня задоволеності пасажирів. Розглянемо ключові моделі, їх властивості, переваги й обмеження, а також обґрунтуємо доцільність використання кожної з них у межах магістерської роботи.

На рисунку 1.9 зображено наступний приклад коду аналогу, який демонструє обробку даних, перетворення категоріальних ознак та застосування базових алгоритмів машинного навчання для класифікації рівня задоволеності пасажирів [17]. Цей приклад ілюструє підхід іншого розробника до підготовки даних, вибору моделей та оцінки точності передбачень.

```
gender=pd.get_dummies(df['Gender'])
df.drop('Gender',axis=1,inplace=True)
df=pd.concat([df,gender],axis=1)
df.head()
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['Class']=le.fit_transform(df['Class'])
df.head()
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(X_train,y_train)
yhat=lr.predict(X_test)from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,f1_score
accuracy_score(y_test,yhat)
0.85993840626203
from sklearn.neighbors import KNeighborsClassifier
k_value=[i for i in range(1,11)]
accuracy=[]
for i in k_value:
    knn=KNeighborsClassifier(n_neighbors=i)
    knn.fit(X_train,y_train)
    accuracy.append(accuracy_score(y_test,knn.predict(X_test)))
print('The maximum accuracy {} can be oobtained with k value ={}'.format(max(accuracy),k_value[accuracy.index(max(accuracy))]))
The maximum accuracy 0.9317336070832799 can be oobtained with k value =9
knn=KNeighborsClassifier(n_neighbors=9)
knn.fit(X_train,y_train)
yhat=knn.predict(X_test)
accuracy_score(y_test,yhat)
 0.9317336070832799
from sklearn.svm import SVC
svc=SVC(kernel='rbf')
svc.fit(X_train,y_train)
yhat=svc.predict(X_test)

accuracy_score(y_test,yhat)

0.9482869241627101
```

Рисунок 1.9 – Приклад коду

Поданий конкурентом код демонструє стандартний підхід до побудови моделей машинного навчання для задачі класифікації рівня задоволеності пасажирів. Позитивними сторонами роботи є коректна попередня обробка даних, зокрема перетворення категоріальних ознак у числові, заповнення пропусків середнім значенням та масштабування ознак перед тренуванням моделей. Також автор перевірив кілька класичних алгоритмів — логістичну регресію, KNN та SVM, що дає змогу оцінити продуктивність традиційних методів на цьому датасеті [17].

Разом з тим, деякі аспекти реалізації залишають простір для подальшого вдосконалення. Наприклад, обробка категоріальних ознак частково виконана через get_dummies, частково через LabelEncoder, що можна систематизувати для більшої узгодженості.

Також у коді відсутнє налаштування гіперпараметрів і системна валідація моделей, що є звичайною практикою для покращення точності та надійності передбачень. Важливо зазначити, що використані моделі — базові алгоритми машинного навчання — забезпечують надійну стартову точку для аналізу, а застосування більш складних методів, таких як нейронні мережі або ансамблеві моделі, може додатково підвищити ефективність передбачення.

У підсумку, приклад конкурента добре ілюструє базові кроки підготовки даних, вибору моделей і оцінки точності передбачень. Він може слугувати відправною точкою для побудови більш складних та глибоких моделей, які включають нейронні мережі, сучасні ансамблеві алгоритми та системну оцінку якості моделей.

Розглянемо останній приклад, у якому автор порівнює ефективність кількох класичних алгоритмів машинного навчання для передбачення рівнів задоволеності пасажирів авіакомпаніями. У цьому рішенні використано логістичну регресію, наївний Байєс, Random Forest та алгоритм найближчих сусідів. Кожна з цих моделей має свої сильні та слабкі сторони, що безпосередньо впливає на точність класифікації та можливість практичного застосування в інформаційних технологіях аналізу сервісної якості [18].

На рисунку 1.10 зображено приклад першої частини коду аналогу.

```
x = df.drop('satisfaction', axis=1)
y = df['satisfaction']

x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.2, random_state=42)


model = LogisticRegression()
model.fit(x_train, y_train)
log_Y_pred = model.predict(x_test)


gaussian = GaussianNB()
gaussian.fit(x_train,y_train)
gaussian_Y_pred = gaussian.predict(x_test)


random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(x_train, y_train)
random_forest_Y_pred = random_forest.predict(x_test)


knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(x_train,y_train)
knn_Y_pred = knn.predict(x_test)
```

Рисунок 1.10 – Перша частина коду аналогу

У першій частині коду здійснюється формування вибірок для навчання та тестування моделі, після чого відбувається тренування чотирьох алгоритмів класифікації: Logistic Regression, Naive Bayes, Random Forest та K-Nearest Neighbors. Спочатку з датафрейму виділяються ознаки та цільова змінна “satisfaction”. За допомогою функції train_test_split набір даних поділяється на тренувальну та тестову частини, що дозволяє об’єктивно оцінити якість моделей. Далі кожна модель навчається на тренувальних даних і одразу використовується для передбачення значень на тестовій вибірці. Такий підхід забезпечує швидке порівняння базових алгоритмів і створює основу для подальшого аналізу точності [18].

Цей етап має свої переваги. Використання чотирьох різних моделей дозволяє оцінити, наскільки різні підходи машинного навчання здатні відтворювати залежності в даних щодо задоволеності пасажирів. Поділ вибірки гарантує

об’єктивність оцінки, а застосування класичних алгоритмів забезпечує швидке отримання результатів та можливість порівняння різних типів моделей — лінійної, ймовірнісної, ансамблевої та метричної. Втім, є й недоліки.

На цьому етапі ще не проводиться попередня обробка ознак, масштабування або оптимізація гіперпараметрів, що може впливати на якість роботи KNN і логістичної регресії. Крім того, моделі використовуються “за замовчуванням”, без налаштувань, тому отримані передбачення можуть не відображати максимально можливу точність цих методів. Таке порівняння є коректним лише як базове, але потребує подальшого вдосконалення для застосування в реальній інформаційній системі передбачення задоволеності пасажирів [18].

На рисунку 1.11 зображено другу частину коду аналогу.

```
models = [model,gaussian,random_forest,knn]
model_names = ['Logistic Regression', 'Naive Bayes', 'Random Forest', 'KNN']

metrics = {
    "Accuracy": accuracy_score,
    "Precision": precision_score,
    "Recall": recall_score,
    "F1 Score": f1_score
}

results = []

for model, name in zip(models, model_names):
    model_preds = model.predict(x_test)
    model_results = {"Model": name}
    for metric_name, metric_func in metrics.items():
        model_results[metric_name] = metric_func(y_test, model_preds)
    results.append(model_results)

results_df = pd.DataFrame(results)
print(results_df)

plt.figure(figsize=(10, 6))
plt.bar(results_df['Model'], results_df['Accuracy'], color='skyblue')
plt.xlabel('Model')
plt.ylabel('Accuracy')
plt.title('Model Comparison')
plt.show()
```

Рисунок 1.11 – Друга частина коду аналогу

У другій частині коду здійснюється узагальнення результатів роботи кожної моделі, формування таблиці з метриками та побудова простого графічного порівняння. Такий підхід має певні переваги, оскільки він дозволяє в компактній формі переглянути значення основних показників для всіх алгоритмів, не дублюючи код і забезпечуючи зручність порівняння [18]. Структура через цикл

робить процес універсальним — за потреби можна легко додати нову модель або метрику, а побудова діаграми створює наочне уявлення про роботу алгоритмів.

Водночас цей фрагмент має й певні недоліки: метрики подаються без додаткових налаштувань чи глибшої обробки, тому результати можуть бути надто узагальненими. Візуалізація подається у спрощеному вигляді та не розкриває повної поведінки моделей, зокрема їх помилок або особливостей класифікації різних груп даних [18]. Отримані значення використовуються лише під час виконання коду, без подальшого аналізу чи збереження, що обмежує можливість розширеної інтерпретації та повторного використання результатів.

1.5 Висновки

У даному розділі було проведено детальний аналіз предметної області, актуальності та доступних інформаційних технологій для аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями. Виявлено, що сучасні інформаційні технології дозволяють автоматизувати збір і аналіз великих обсягів даних, виявляти закономірності та тенденції, а також будувати прогностичні моделі для оптимізації сервісу та підвищення лояльності пасажирів.

Аналіз актуальності показав, що українська авіаційна галузь, незважаючи на складні умови воєнного часу та закриття повітряного простору, продовжує функціонувати, адаптуватися до нових ринкових умов та інтегруватися у європейський ринок. Попит на авіаперевезення стабільно зростає, що підкреслює важливість впровадження сучасних технологій для підвищення ефективності та якості обслуговування пасажирів.

У роботі обрано платформу Kaggle і мову програмування Python як основні інструменти для аналізу даних, застосовано багатошарові нейронні мережі та сучасні методи машинного навчання, зокрема XGBoost, Gaussian Process Classifier і Ridge Classifier, що дозволяє підвищити ефективність класифікації та оцінки задоволеності пасажирів.

Огляд готових рішень показав сильні та слабкі сторони існуючих підходів, зокрема реалізацій нейронних мереж на базі Numpy, Keras та PyTorch, і дав змогу визначити напрямки для удосконалення моделі, враховуючи питання перенавчання, вибору функцій активації, Dropout та оцінки стабільності навчання.

Таким чином, проведений аналіз предметної області, оцінка актуальності та огляд сучасних технологій дозволили обґрунтувати вибір інструментів і методів для побудови моделі передбачення задоволеності пасажирів. Це створює надійну основу для подальшого експериментального дослідження та розробки ефективного рішення для оптимізації сервісу авіакомпаній.

2. РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ

2.1 Підготовка даних та розвідувальний аналіз

На початковому етапі роботи був виконаний імпорт датасету «Airline Passenger Satisfaction», до складу якого входять файли train.csv та test.csv**.** Для перевірки успішності завантаження даних та ознайомлення зі структурою навчального набору було здійснено виведення перших п’яти записів. Такий підхід дозволив отримати первинне уявлення про типи змінних, їхні значення та формат даних, що стало важливим кроком перед подальшою обробкою, очищенням та аналізом.

На рисунку 2.1 зображено статистичний опис кількісних (числових) стовпців для тренувального набору даних.

	Unnamed: 0	id	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking
count	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.0000
mean	51951.500000	64924.210502	39.379706	1189.448375	2.729683	3.060296	2.756901
std	29994.645522	37463.812252	15.114964	997.147281	1.327829	1.525075	1.398929
min	0.000000	1.000000	7.000000	31.000000	0.000000	0.000000	0.000000
25%	25975.750000	32533.750000	27.000000	414.000000	2.000000	2.000000	2.000000
50%	51951.500000	64856.500000	40.000000	843.000000	3.000000	3.000000	3.000000
75%	77927.250000	97368.250000	51.000000	1743.000000	4.000000	4.000000	4.000000
max	103903.000000	129880.000000	85.000000	4983.000000	5.000000	5.000000	5.000000

Рисунок 2.1 – Статистичний огляд даних набору train

Для ознайомлення з числовими даними датасету було використано метод, що дозволяє отримати основні статистичні характеристики кожного стовпця. До таких характеристик належать кількість спостережень, середнє значення, стандартне відхилення, мінімальне та максимальне значення, а також квартилі (25%, 50%, 75%). Наприклад, аналіз стовпця "Age" показав, що середній вік пасажирів становить близько 39 років, стандартне відхилення дорівнює приблизно 15 рокам, при цьому мінімальний вік зафіксовано на рівні 7 років, а максимальний досягає 85 років. Така статистична оцінка дозволяє отримати чітке уявлення про розподіл

числових змінних, виявити потенційні аномальні значення, а також зробити перші висновки щодо тенденцій і варіативності даних.

У рамках підготовки даних до подальшого аналізу та навчання моделей машинного навчання було виконано кілька ключових етапів. По-перше, цільова змінна була закодована за допомогою LabelEncoder, що дозволило перетворити категоріальні значення у числовий формат, необхідний для більшості алгоритмів. По-друге, були видалені рядки з пропущеними значеннями як у навчальній, так і в тестовій вибірках, щоб забезпечити коректність та однорідність даних. По-третє, числові ознаки були масштабовані за допомогою методу StandardScaler, що дозволяє нормалізувати дані і покращує стабільність та швидкість навчання моделей. Завдяки цим крокам підготовка даних стала більш системною, що сприяє підвищенню ефективності подальшого аналізу, побудови моделей та оцінки результатів передбачення.

На рисунку 2.2 зображено код масштабування ознак за допомогою StandardScaler.

```
# Масштабування
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Рисунок 2.2 – Приклад коду

Розвідувальний аналіз даних (Exploratory Data Analysis, EDA) є одним із фундаментальних етапів роботи з будь-якими наборами даних і має вирішальне значення для успішного проведення аналітичних досліджень. Основна мета EDA полягає у всебічному вивченні структури та властивостей даних, що дозволяє отримати глибоке розуміння їхньої сутності та поведінки. На цьому етапі дослідник аналізує ключові характеристики набору даних, включаючи розподіл значень, наявність пропущених даних, аномалії, тенденції та взаємозв’язки між змінними.

Для досягнення цих цілей використовуються різноманітні статистичні методи, а також візуальні засоби, такі як гістограми, діаграми розсіяння, коробкові

діаграми, теплові карти та інші графічні представлення. Візуалізація даних допомагає наочно показати закономірності, виявити потенційні залежності та аномалії, які не завжди помітні при простому числовому аналізі. Такий підхід дозволяє сформулювати гіпотези для подальшого дослідження, оцінити доцільність використання конкретних моделей машинного навчання та прийняти обґрунтовані рішення щодо підготовки та обробки даних.

EDA також є важливим кроком для оцінки якості даних і їхньої готовності до моделювання. Завдяки розвідувальному аналізу можна виявити проблеми, пов’язані з пропущеними або некоректними значеннями, та визначити оптимальні стратегії їх обробки.

Крім того, він дає змогу оцінити вплив окремих змінних на результат, що є ключовим для побудови точних і стабільних моделей. У комплексі це забезпечує більш ефективний і свідомий підхід до машинного навчання, передбачення та прийняття рішень на основі даних, підвищуючи якість аналітичних висновків та дозволяючи розробляти моделі, адаптовані до реальних умов [19].

Першим кроком було побудовано кореляційну матрицю для дослідження зв’язків між числовими та закодованими категоріальними ознаками (рис. 2.3).

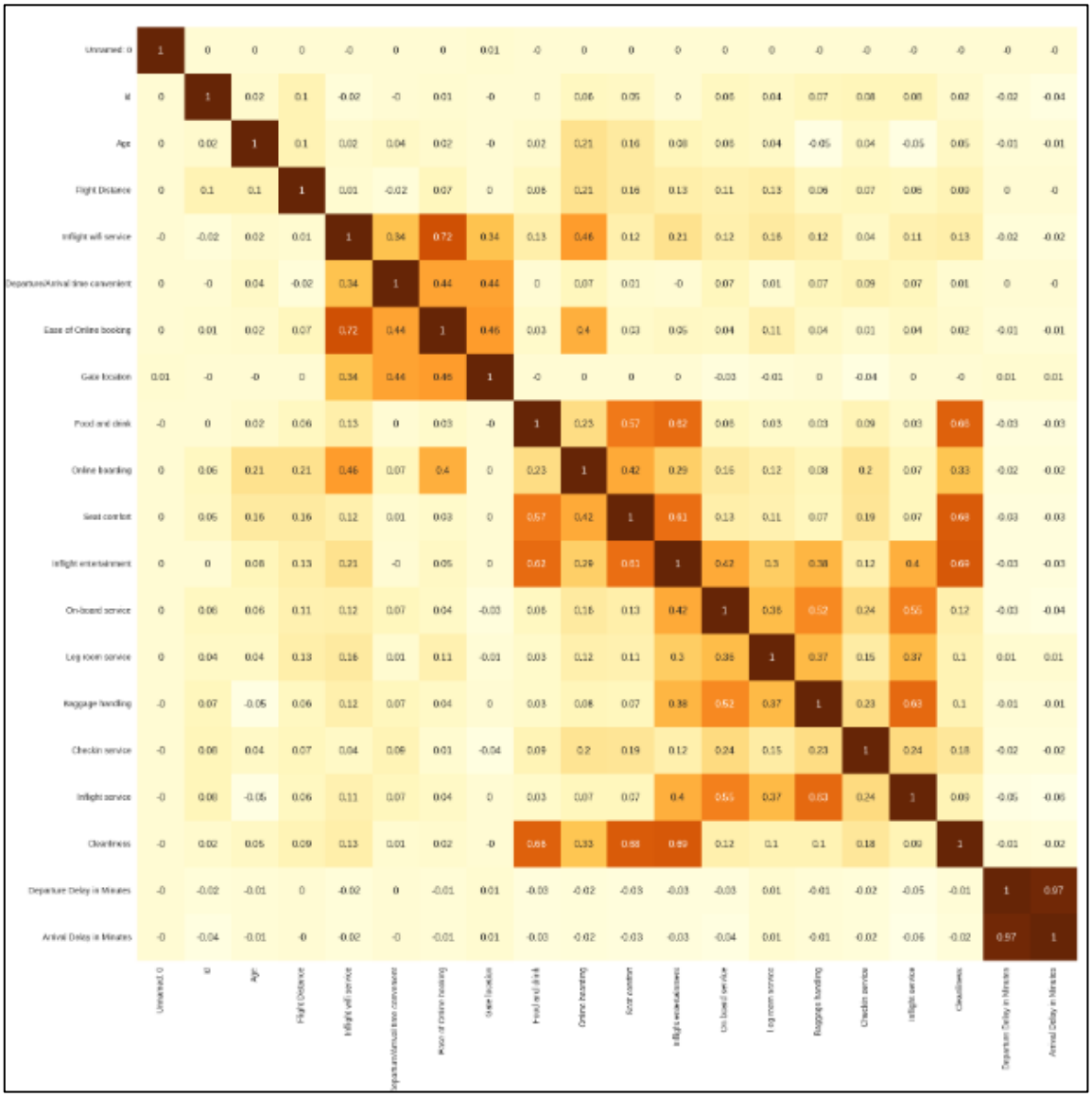


Рисунок 2.3 – Кореляційна матриця

Матриця кореляцій демонструє численні сильні зв’язки між різними змінними у наборі даних. Зокрема, змінні "Затримка відправлення в хвилинах" та "Затримка прибуття в хвилинах" мають надзвичайно високий коефіцієнт кореляції, що вказує на пряму залежність між цими параметрами. Це логічно, оскільки будь-яка затримка при відправленні рейсу, як правило, безпосередньо впливає на затримку після прибуття.

Крім того, спостерігається висока кореляція між "Ease of Online Booking" та "Inflight Wi-Fi Service", що може свідчити про загальне позитивне сприйняття пасажирами цифрових та онлайн-сервісів авіакомпанії. Іншими словами, пасажири,

які задоволені процесом онлайн-бронювання, часто також оцінюють високою якість Wi-Fi під час польоту.

Додаткові сильні кореляції помітні між змінними, що оцінюють комфорт та обслуговування під час польоту: наприклад, "Seat Comfort", "Food and Drink", "Inflight Entertainment" та "Cleanliness". Ці взаємозв’язки свідчать про те, що різні аспекти обслуговування пасажирів тісно пов’язані і взаємно впливають один на одного. Така інформація є важливою для подальшого побудови моделей машинного навчання, оскільки дозволяє враховувати взаємозалежність факторів задоволеності пасажирів та підвищує точність передбачення їхніх оцінок.

Загалом, аналіз кореляцій показує, що існують чіткі закономірності у сприйнятті різних сервісних аспектів авіаперевезень, і це дає змогу робити більш обґрунтовані висновки про поведінку та очікування пасажирів.

На рисунку 2.4 зображено розподіл за статтю та лояльністю.

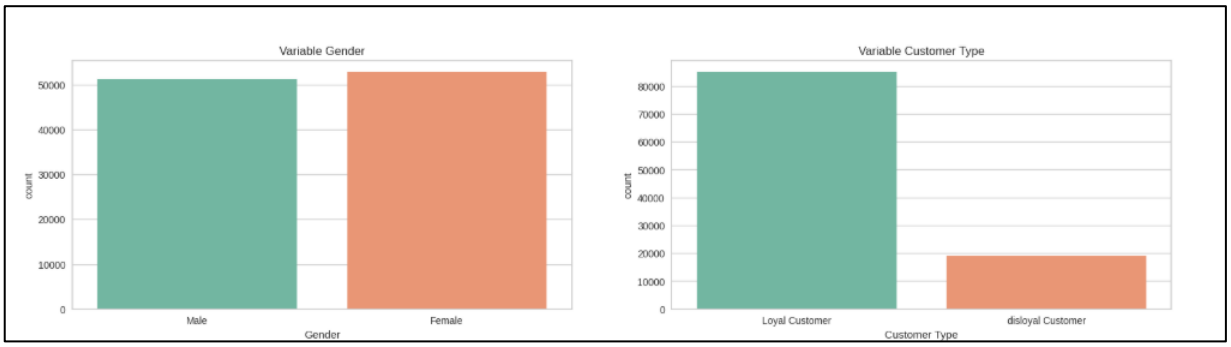


Рисунок 2.4 – Розподіл за статтю та лояльністю

Аналіз розподілу пасажирів за статтю показує, що кількість чоловіків і жінок практично рівна, що свідчить про гендерну збалансованість даних. Завдяки цьому можна зробити висновок, що стать респондентів навряд чи чинить значний вплив на загальний рівень задоволеності пасажирів послугами авіакомпанії. Іншими словами, оцінки та відгуки не схильні до упередженості щодо чоловіків чи жінок, і будь-які загальні тенденції у рівні задоволеності відображають реальний досвід пасажирів без врахування статі.

Натомість розподіл пасажирів за категоріями клієнтів демонструє явну перевагу лояльних, або постійних, клієнтів над тимчасовими чи новими пасажирами. Це означає, що більшість отриманих відгуків і оцінок надходить саме від тих, хто часто користується послугами авіакомпанії та має сформовану думку на основі регулярного досвіду взаємодії з її сервісом. Через це думка лояльних клієнтів відіграє ключову роль у формуванні загальної картини щодо якості обслуговування. Саме їхні оцінки є визначальними для висновків, що стосуються рівня задоволеності та потенційних напрямів покращення сервісу, оскільки вони відображають стабільний та об'єктивний досвід користування послугами авіакомпанії.

На рисунку 2.5 зображено розподіл за віком.

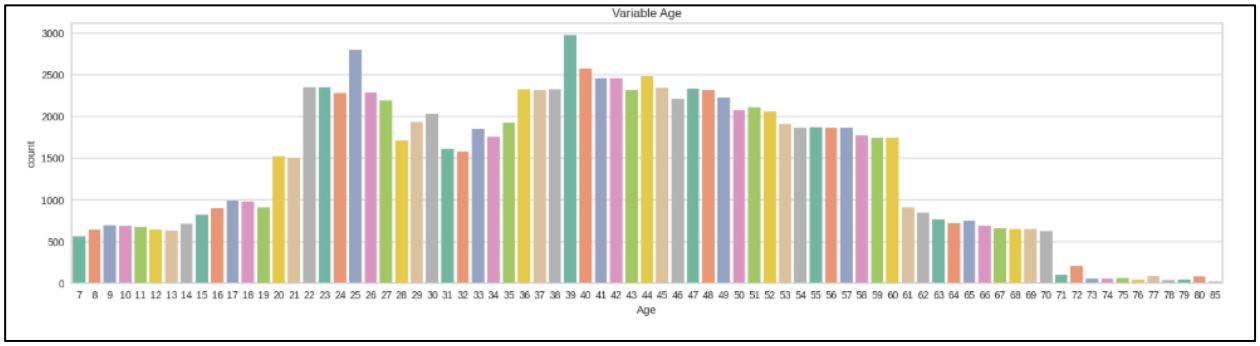


Рисунок 2.5 – Розподіл за віком

Аналіз вікової структури пасажирів показує, що найбільшу активність у користуванні послугами авіакомпанії демонструють люди віком від 25 до 50 років. Саме представники цієї вікової категорії складають більшість у вибірці, причому помітний пік припадає приблизно на 39 років. Молодші пасажири, зокрема ті, кому менше 20 років, а також старші пасажири старше 65 років, зустрічаються у набагато меншій кількості.

Це дозволяє зробити висновок, що основна частина аналізу рівня задоволеності пасажирів має бути зосереджена саме на дорослому, працездатному населенні. Саме ця вікова група формує більшість відгуків та оцінок, а також

найчастіше користується авіаперевезеннями, що робить її ключовою для визначення загальних тенденцій задоволеності обслуговуванням.

На рисунку 2.6 зображено аналіз задоволеності пасажирів за відстанню перельоту.

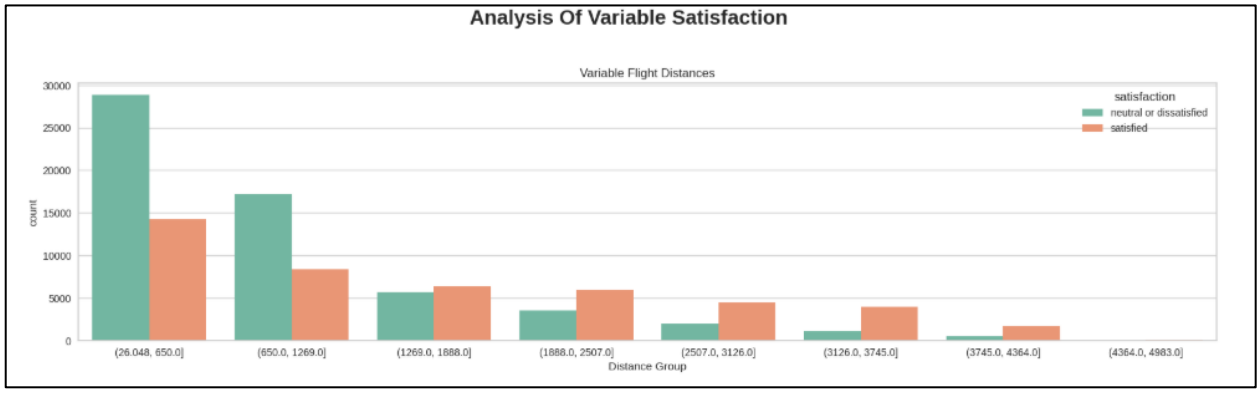


Рисунок 2.6 – Аналіз задоволеності пасажирів за відстанню перельоту

Аналіз задоволеності пасажирів залежно від відстані перельоту показує цікаві закономірності. З графіка видно, що більшість польотів у вибірці — це короткі рейси до 650 км, і саме на цих маршрутах значна частина пасажирів залишає відгуки, що характеризують їхній досвід як незадовільний або нейтральний. Натомість зі збільшенням відстані перельоту рівень задоволеності зростає: на середніх та довгих рейсах кількість позитивних оцінок значно більша.

Ця тенденція може свідчити про те, що на коротких маршрутах пасажири більш чутливі до якості обслуговування або мають високі очікування щодо швидкості та комфорту поїздки, тоді як на довгих рейсах сервіс стає більш помітним і комплексним, що сприяє вищому рівню задоволеності.

Переваги такого аналізу полягають у наступному:

– Виявлення проблемних сегментів – дозволяє визначити, на яких маршрутах якість обслуговування потребує покращення;

– Фокус на довгострокові покращення – дає змогу авіакомпанії концентрувати ресурси на тих рейсах, де підвищення комфорту та сервісу матиме найбільший ефект для пасажирів;

– Підтвердження важливості досвіду пасажира — показує, що задоволеність пасажирів залежить не лише від обслуговування, а й від тривалості перельоту та сприйняття комфорту;

– Об’єктивна оцінка сервісу — дозволяє аналізувати тенденції на основі реальних відгуків різних груп пасажирів, що підвищує точність висновків щодо якості обслуговування.

На рисунку 2.7 зображено вплив класу та типу подорожі на задоволеність пасажирів.

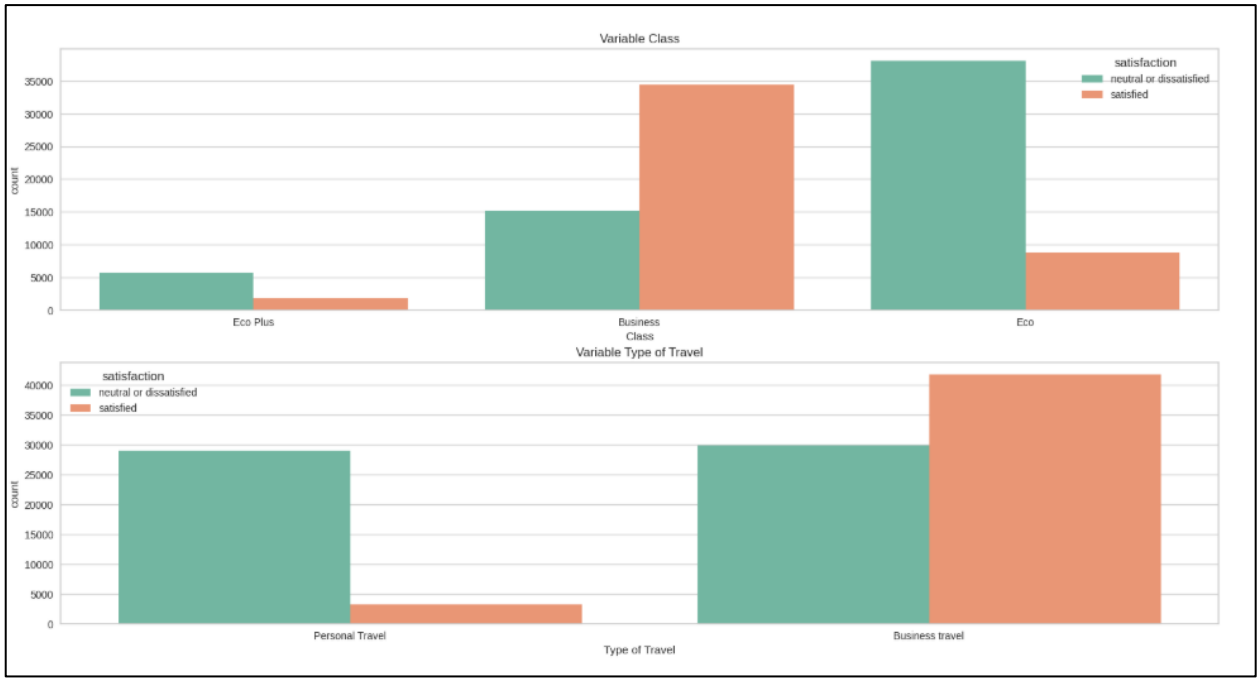


Рисунок 2.7 – Вплив класу та типу подорожі на задоволеність пасажирів

Аналіз рівня задоволеності пасажирів залежно від класу обслуговування та мети подорожі демонструє чіткі закономірності. Пасажири бізнес-класу найчастіше залишаються задоволеними сервісом авіакомпанії, тоді як у економ-класі переважають нейтральні або незадовільні оцінки. Це може свідчити як про більш високі очікування пасажирів економ-класу, так і про обмеженість доступних сервісів у цьому сегменті.

Що стосується мети подорожі, пасажири, які летіли з робочих причин, частіше залишалися задоволеними, тоді як ті, хто подорожував у приватних

справах, здебільшого висловлювали незадоволення або нейтральну оцінку. Така закономірність вказує на те, що рівень задоволеності пасажирів формується не лише класом обслуговування, але й контекстом поїздки та очікуваннями пасажирів.

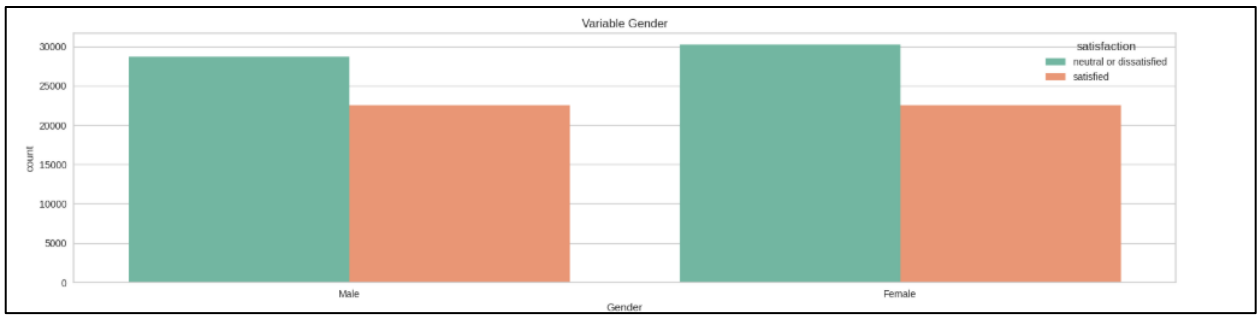


Рисунок 2.8 – Вплив статі на рівень задоволеності пасажирів

Хоча загальний аналіз показував, що гендер не чинить суттєвого впливу на рівень задоволеності пасажирів, детальніший розгляд за допомогою графіка дозволяє виявити певні відмінності між чоловіками та жінками. Зокрема, серед чоловіків спостерігається більша частка задоволених пасажирів, тоді як серед жінок переважають нейтральні або негативні оцінки.

Ця тенденція може свідчити про те, що жінки демонструють більш критичне сприйняття якості обслуговування під час перельотів або мають інші очікування від сервісу. Наприклад, вони можуть більше звертати увагу на деталі комфорту, точність обслуговування чи обслуговування клієнтського сервісу, що впливає на їхню оцінку. Водночас чоловіки, можливо, оцінюють сервіс більш схематично, концентруючись на основних аспектах поїздки, таких як своєчасність та базовий комфорт.

Отже, детальніший аналіз за статтю дозволяє зрозуміти, що хоча загальні висновки можуть не показувати суттєвої різниці між гендерами, у межах конкретних груп все ж існують відмінності у сприйнятті сервісу. Це дає змогу авіакомпанії більш точно адаптувати свої послуги, враховуючи очікування різних категорій пасажирів.

На рисунку 2.9 зображено вплив віку на рівень задоволеності пасажирів.

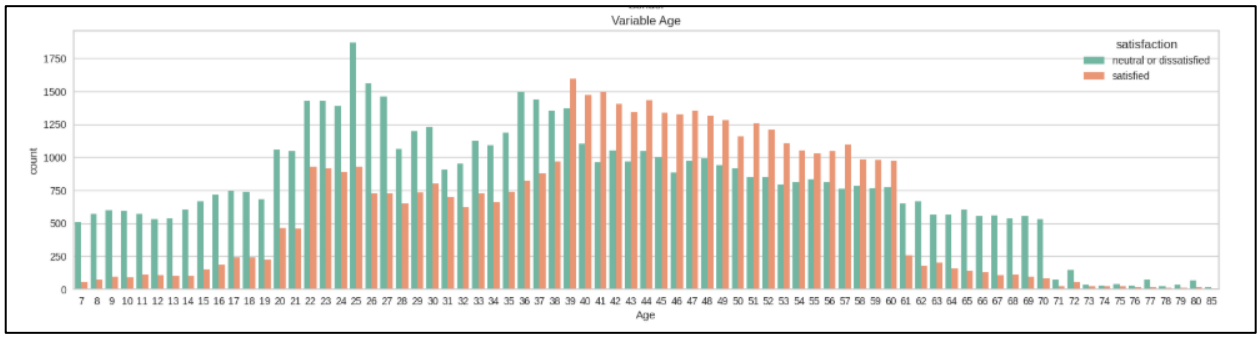


Рисунок 2.9 – Вплив віку на рівень задоволеності пасажирів

Аналіз задоволеності пасажирів за віковими категоріями показує цікаві закономірності. З графіка видно, що серед молодих пасажирів віком від 7 до 39 років частіше зустрічаються негативні або нейтральні оцінки польотів. У віковій групі 39–60 років, навпаки, переважають задоволені пасажири, що свідчить про більш позитивне сприйняття сервісу серед людей середнього віку.

Що стосується старшої вікової категорії від 61 до 80 років, результати є більш змішаними, проте незадоволені пасажири все ж переважають. Така тенденція може вказувати на те, що люди середнього віку, ймовірно, мають більший досвід подорожей авіаційним транспортом і, можливо, більш об’єктивно оцінюють якість обслуговування, у той час як молодші або старші пасажири можуть бути більш критичними або мати специфічні очікування від сервісу.

Отже, середній вік пасажирів асоціюється з вищим рівнем задоволеності, що може бути корисним для авіакомпанії при розробці цільових стратегій покращення обслуговування різних вікових груп.

Рисунок 2.10 – Розподіл затримок при відправленні та прибутті (у хвилинах)

Аналіз розподілу затримок рейсів при відправленні та прибутті у хвилинах показує, що більшість польотів здійснюються без значних відхилень від розкладу. Найчастіше рейси відправляються та прибувають вчасно або з мінімальною затримкою — кілька хвилин. Це підтверджується різким піком на графіках для 0 хвилин затримки, що вказує на те, що найбільша кількість пасажирів не зазнає незручностей через очікування. Затримки понад 10 хвилин трапляються значно рідше, що свідчить про загальну стабільність виконання рейсів авіакомпанією.

Такі дані дозволяють стверджувати, що авіакомпанія здебільшого дотримується графіка польотів, а пасажири можуть очікувати своєчасного обслуговування.

На рисунку 2.11 зображено розподіл пасажирів за типом подорожі та класом обслуговування.

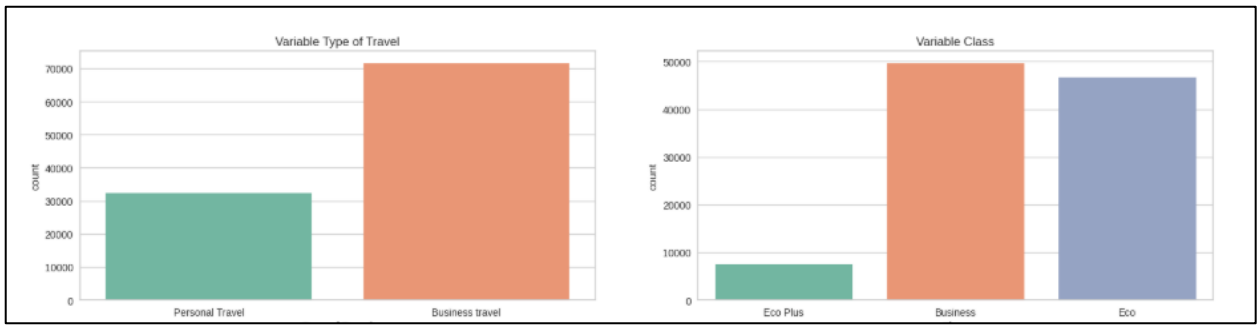


Рисунок 2.11 – Розподіл пасажирів за типом подорожі та класом обслуговування

Аналіз типу подорожей та обраного класу обслуговування показує, що більшість пасажирів подорожують з робочих причин, а не у приватних справах. Що стосується класів, найпопулярнішими залишаються бізнес-клас та економ-клас, тоді як проміжний клас «Eco Plus» практично не обирають.

Ця тенденція може свідчити про те, що пасажири прагнуть або максимального комфорту та сервісу, який пропонує бізнес-клас, або ж віддають перевагу більш економічному варіанту, намагаючись заощадити, обираючи економ-клас. Менший попит на «Eco Plus» може вказувати на те, що проміжний клас не відповідає очікуванням більшості пасажирів, які шукають явні переваги або значну економію.

Такий аналіз дозволяє авіакомпанії краще розуміти структуру попиту на різні класи обслуговування та адаптувати пропозиції відповідно до уподобань клієнтів.

На рисунку 2.12 зображено розподіл дистанцій польотів пасажирів.

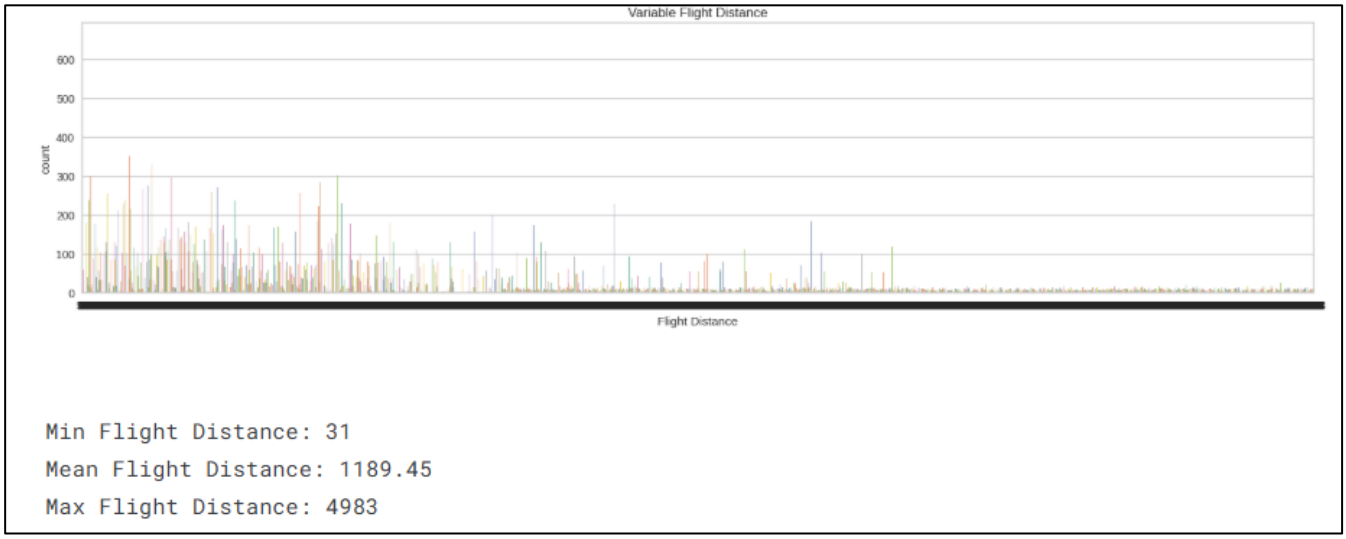


Рисунок 2.12 – Розподіл пасажирів за відстанню польоту

Аналіз тривалості рейсів показує, що більшість польотів у вибірці мають коротку дистанцію, тоді як кількість далеких рейсів поступово зменшується, що свідчить про експоненціальний характер розподілу. Мінімальна відстань перельоту становить 31 км, середня — близько 1189 км, а максимальна досягає 4983 км. Такі дані підтверджують, що короткі рейси переважають у загальному наборі даних і формують основну частку пасажиропотоку, тоді як довгі маршрути зустрічаються значно рідше.

Цей розподіл дозволяє авіакомпанії орієнтуватися на обслуговування переважно коротких рейсів, водночас оцінюючи потреби пасажирів на дальніх маршрутах для можливого покращення сервісу.

Отже, проведений аналіз різних факторів — статі, віку, типу клієнта, класу обслуговування, мети подорожі, тривалості рейсу та затримок — дозволяє сформувати комплексне розуміння досвіду пасажирів авіакомпанії. Виявлені закономірності показують, які групи пасажирів частіше залишаються задоволеними, а на яких сегментах слід звернути увагу для покращення якості обслуговування.

Такий підхід дозволяє авіакомпанії об’єктивно оцінювати сильні та слабкі сторони сервісу, визначати пріоритетні напрямки для вдосконалення, а також адаптувати послуги під очікування різних категорій пасажирів.

Використання структурованих даних забезпечує більш точні та надійні висновки, що можуть слугувати основою для прийняття управлінських рішень і підвищення задоволеності клієнтів у майбутньому.

2.2 Архітектура та алгоритм роботи

Блок-схема є графічним засобом подання алгоритмів або процесів у вигляді послідовності дій, що виконуються у системі. Вона складається з блоків різних типів, які позначають початок і кінець процесу, операції, рішення або підпроцеси, а також стрілок, що показують послідовність виконання. Блок-схеми дозволяють наочно уявити логіку роботи технології, спростити аналіз алгоритмів, виявити потенційні помилки та оптимізувати послідовність виконання дій [20].

Для реалізації програмного модуля було розроблено алгоритм, що описує послідовність виконання основних етапів обробки та аналізу даних. На рисунку 2.13 представлено блок-схему алгоритму роботи програми «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями», яка відображає основні етапи обробки даних, аналізу та формування прогнозу рівня задоволеності пасажирів.

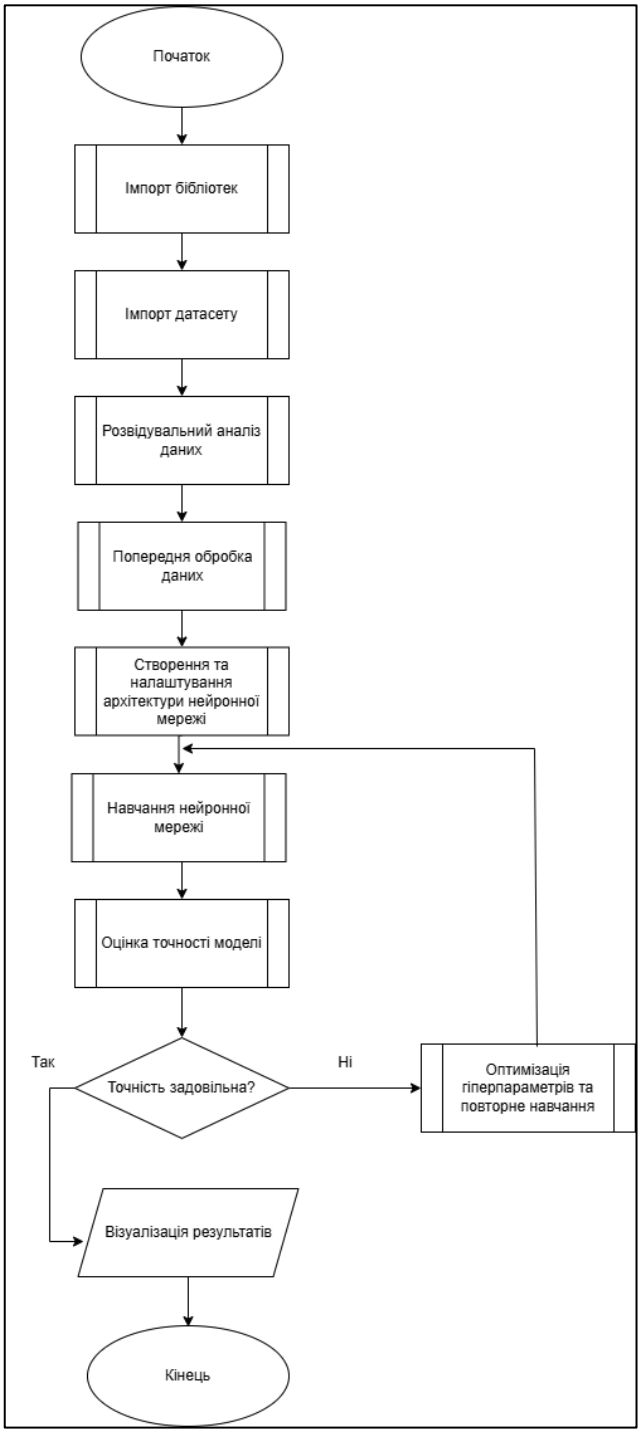


Рисунок 2.13 – Блок-схема алгоритму роботи програми

Спочатку відбувається імпорт бібліотек, далі імпортується датасет, після чого проводиться розвідувальний аналіз даних. Потім здійснюється попередня обробка даних (видалення нульових значень, кодування, масштабування), створюється та налаштовується архітектура нейронної мережі і відбувається її навчання.

Далі оцінюється точність моделі, і якщо вона задовільна, процес завершується, а якщо ні — проводиться оптимізація гіперпараметрів та повторне навчання до досягнення задовільного результату.

Діаграма розгортання (Deployment Diagram) є одним із типів діаграм UML і використовується для моделювання фізичного розміщення компонентів програмної технології на апаратному забезпеченні. Вона показує, які програмні об’єкти або модулі розгортаються на яких вузлах (сервери, комп’ютери, пристрої), як ці вузли взаємодіють між собою та які ресурси використовуються для виконання технології.

Діаграма розгортання дозволяє наочно відобразити архітектуру технології, зрозуміти її інфраструктурні вимоги, планувати взаємодію між апаратними та програмними компонентами, а також передбачати потенційні вузькі місця у роботі технології. Вона є важливим інструментом для системних інженерів та розробників, оскільки допомагає забезпечити ефективне та надійне розміщення програмних компонентів і сервісів [21].

У рамках даної дипломної роботи також була побудована діаграма розгортання, яка ілюструє структуру технології, взаємодію компонентів та розташування програмних і апаратних модулів для забезпечення процесу передбачення факторів задоволеності пасажирів авіакомпаніями (рис. 2.14).

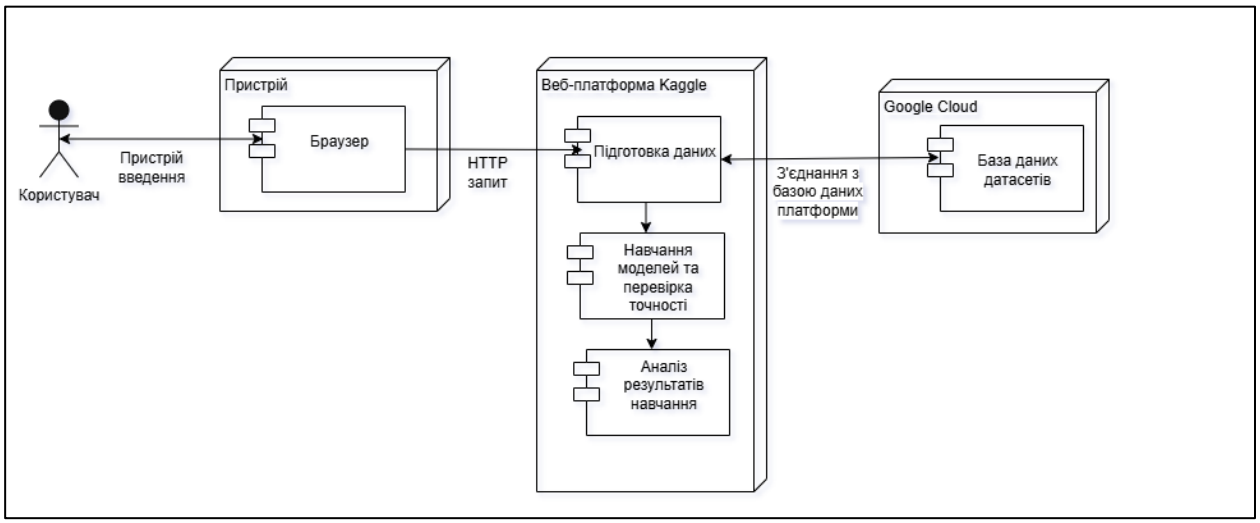


Рисунок 2.14 – Діаграма розгортання

Користувач взаємодіє через пристрій введення та браузер, який надсилає HTTP-запити до веб-платформи Kaggle. На платформі відбувається підготовка даних, навчання моделей і перевірка точності, а також аналіз результатів навчання. Дані для цього отримуються з бази даних датасетів у Google Cloud через підключення до платформи.

Для кращого розуміння функціональних можливостей розробленої інформаційної технології було побудовано діаграму варіантів використання (Use Case діаграму).

Діаграма варіантів використання (Use Case Diagram) є одним із основних засобів моделювання функціональних вимог технології в UML. Вона дозволяє наочно відобразити, які дії або сервіси доступні користувачам технології (акторам), а також як ці дії реалізуються через взаємодію з основними компонентами технології.

Use Case діаграма допомагає визначити ролі користувачів, окреслити межі технології, а також показати зв’язки між окремими функціональними можливостями. Вона використовується для формалізації вимог на ранніх етапах проєктування, забезпечує чітке розуміння того, що повинна виконувати технологія, і слугує основою для подальшого детального проектування та розробки [22].

Вона дозволяє візуально відобразити взаємодію між користувачами технології та її основними функціями. Такий підхід дає змогу визначити ролі користувачів, описати їхні дії та окреслити межі технології (рис. 2.15).

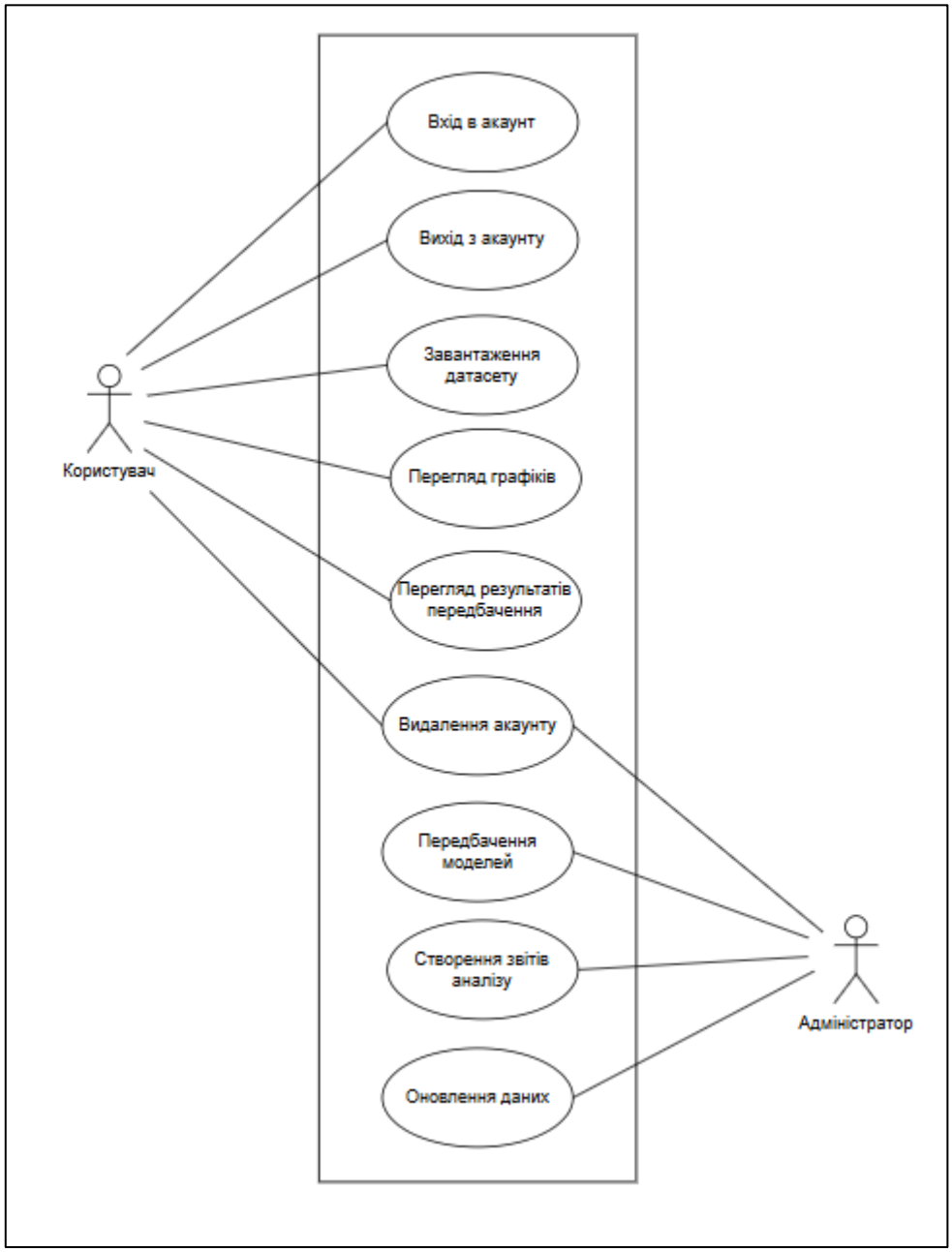


Рисунок 2.15 – Use Case діаграма

Діаграма відображає основні взаємодії користувачів із технологією, а також функціональні можливості, що реалізуються. Користувач має можливість виконувати базові операції — вхід і вихід із акаунту, завантаження датасету, перегляд графіків і результатів передбачення. Адміністратор, крім цих функцій, має розширені можливості — оновлення даних, створення звітів аналізу та керування процесом передбачення моделей.

Діаграма класів є одним із ключових інструментів мови UML (Unified Modeling Language) і використовується для графічного відображення структури програмної технології. Вона показує класи, їхні атрибути та методи, а також взаємозв'язки між об'єктами, що дозволяє зрозуміти організацію технології на рівні її компонентів. Діаграми класів застосовуються для проектування програмного забезпечення, документування існуючих систем, планування модульної структури та забезпечення узгодженості між різними частинами проєкту [23].

У рамках даної дипломної роботи було побудовано UML-діаграму класів основних компонентів технології аналізу та передбачення факторів задоволеності пасажирів авіакомпаніями, що дозволяє наочно продемонструвати структуру об'єктів, їхні атрибути та методи, а також взаємозв'язки між ними (рис. 2.16).

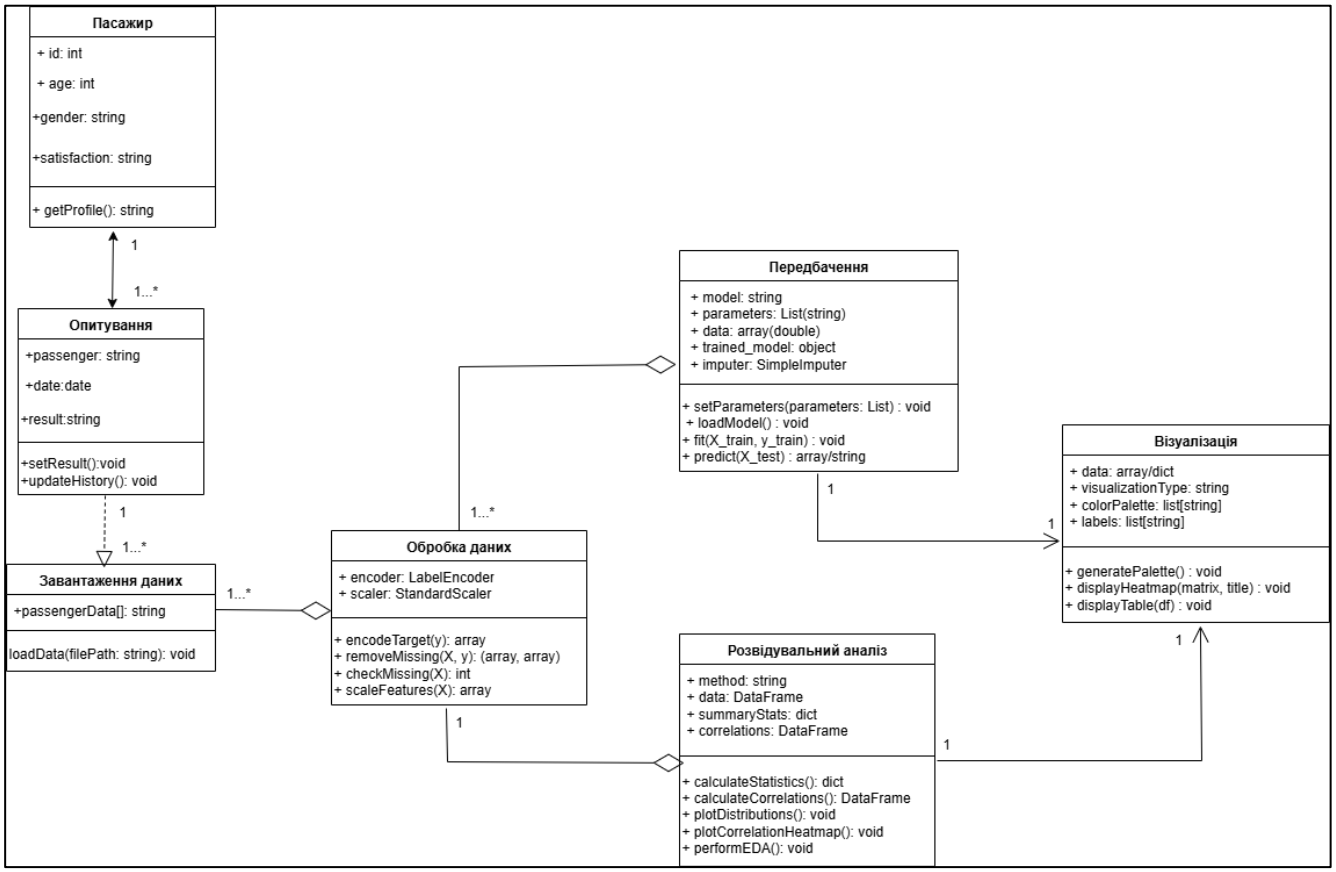


Рисунок 2.16 – Діаграма класів інформаційної технології

Діаграма класів відображає основні сутності технології аналізу та передбачення задоволеності пасажирів та логічні зв’язки між ними. Клас «Пасажир» представляє респондента й пов’язаний асоціацією з класом «Опитування»; кратність зв’язку становить (1..*) , оскільки один пасажир може пройти одне або декілька опитувань. Клас «Опитування» агрегує множину елементів класу «Дані» — агрегування позначається тим, що опитування складається з багатьох записів даних (1..*), але самі дані можуть існувати окремо й використовуватися на наступних етапах обробки. Клас «Дані» з’єднаний асоціацією з класом «Завантаження даних», у якому одна операція завантаження (1) може породжувати множину записів (1..*).

Далі клас «Дані» передається до класу «Обробка даних», який також має асоціацію з кратністю (1..*) , оскільки обробка застосовується до великої кількості записів. Результатом обробки є підготовлений масив даних, який переходить до класу «Розвідувальний аналіз», що отримує рівно один набір оброблених даних.

Клас «Передбачення» використовує дані після етапів EDA та обробки, формуючи модель і прогнозні значення, а клас «Візуалізація» асоційований із «Передбаченням» і «Розвідувальним аналізом», оскільки будує графіки як на аналітичному етапі, так і на етапі роботи моделей.

Уся структура побудована так, що дані проходять послідовні етапи — від завантаження до обробки, далі до аналітичного опрацювання, потім до етапу побудови моделей і завершуються формуванням візуалізацій. Зв’язки між класами чітко відображають ролі кожної сутності, характер їх взаємодії та кратності зв’язків, що забезпечує логічність і цілісність роботи всієї технології.

Діаграма об’єктів є одним із статичних структурних інструментів UML і використовується для демонстрації конкретних екземплярів класів та їхніх зв’язків у певний момент часу. На відміну від діаграми класів, яка описує загальну структуру технології та можливі взаємозв’язки між сутностями, діаграма об’єктів показує реальний стан технології на конкретному прикладі: які об’єкти існують, які значення мають їхні атрибути та як вони пов’язані між собою в рамках певного сценарію роботи [24].

У дипломній роботі діаграма об’єктів використовується для того, щоб продемонструвати практичну реалізацію описаної архітектури технології. Вона наочно ілюструє, як саме взаємодіють окремі елементи технології під час виконання завдань аналізу даних, обробки, передбачення або візуалізації. Це дозволяє підтвердити коректність побудованої моделі, показати узгодженість між логічним проектуванням і фактичними процесами, а також надати читачеві чітке розуміння того, яким чином технологія працює в реальній ситуації (рис. 2.17).

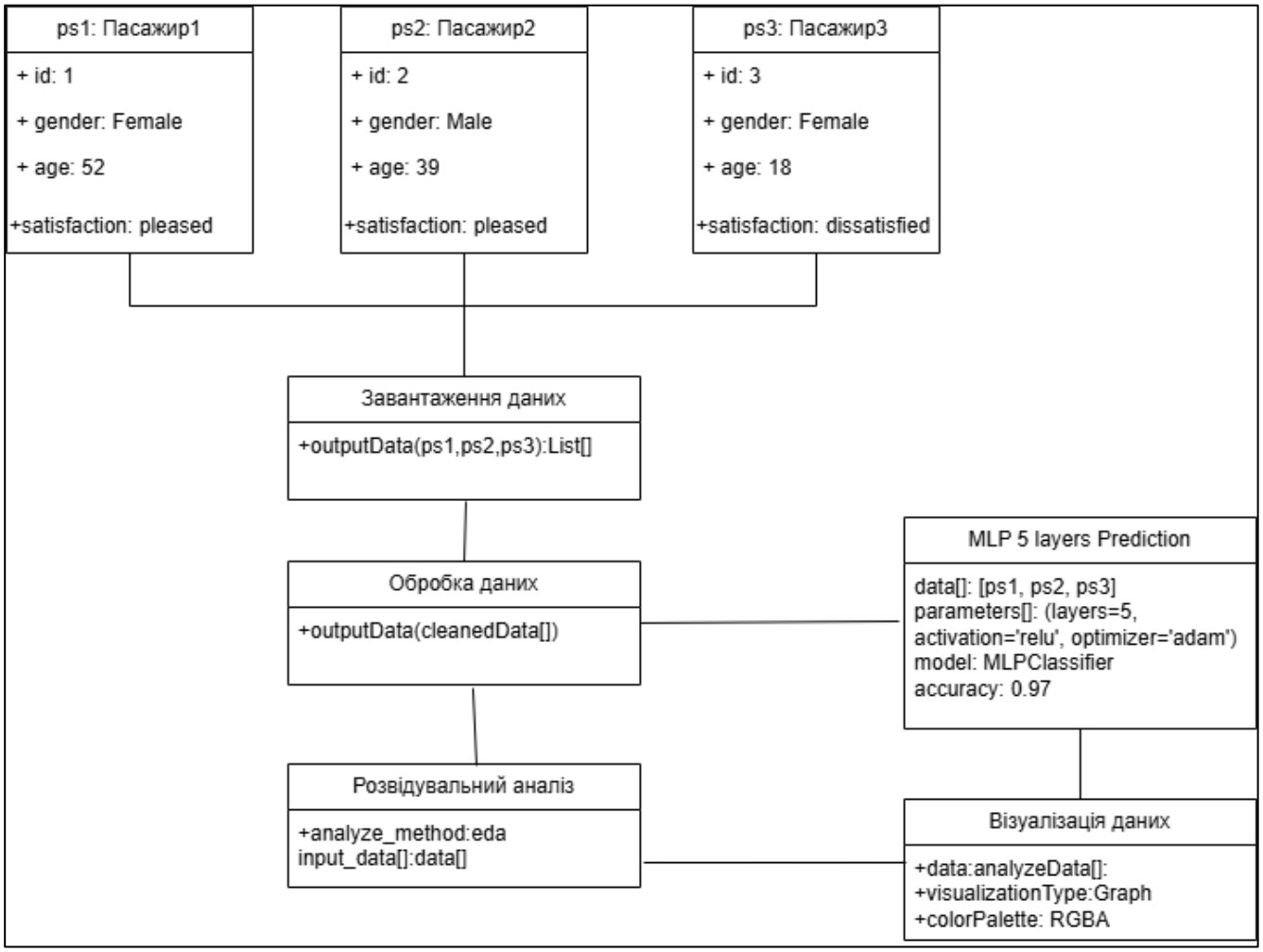


Рисунок 2.17 – Діаграма об’єктів

Діаграма об’єктів складається з конкретних екземплярів класів технології та зв’язків між ними, що показують реальний стан даних і взаємодій у певний момент часу. На відміну від діаграми класів, яка демонструє загальну логічну структуру,

діаграма об’єктів відображає, як саме створені об’єкти фактично використовуються в процесі роботи технології.

На представленій діаграмі показані об’єкти трьох пасажирів із зазначенням їхніх індивідуальних характеристик: ідентифікатора, статі, віку та рівня задоволеності. Ці об’єкти формують вхідні дані, які передаються до компонента «Завантаження даних». Далі відображені об’єкти, що представляють ключові етапи обробки: модуль завантаження створює первинний набір даних, модуль обробки здійснює очищення та підготовку, а об’єкт розвідувального аналізу проводить дослідження підготовлених даних.

Особливе місце займає об’єкт передбачення, який у цьому випадку представляє екземпляр найточнішої моделі — п’ятишарової нейронної мережі. Він містить дані для моделювання, налаштування параметрів та результат роботи моделі. Завершує схему об’єкт «Візуалізація даних», відповідальний за графічне подання аналізу та результатів передбачення.

Таким чином, діаграма об’єктів демонструє цілісний приклад проходження даних через основні модулі технології: від отримання інформації про пасажирів до побудови прогнозу й візуального представлення результатів.

2.3 Висновки

У даному розділі проведено повний цикл підготовки, аналізу та моделювання даних для розроблення інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями. На основі набору даних «Airline Passenger Satisfaction» виконано імпорт, очищення, кодування змінних і масштабування ознак, що забезпечило коректність подальшого навчання моделей. Проведений розвідувальний аналіз дав змогу виявити ключові закономірності — вплив віку, статі, класу обслуговування, типу подорожі та відстані перельоту на рівень задоволеності пасажирів.

Розроблена архітектура технології та блок-схема алгоритму демонструють послідовність виконання етапів: від завантаження та обробки

даних до формування передбачення за допомогою моделей машинного навчання та нейронних мереж.

У межах розділу побудовано UML-діаграму класів, яка відображає основні сутності технології, їх структуру та логічні взаємозв’язки. Класи «Пасажир», «Опитування», «Дані», «Завантаження даних», «Обробка даних», «Розвідувальний аналіз», «Передбачення» та «Візуалізація» структуровані за принципом послідовного проходження даних усіма етапами. Діаграма класів дає змогу чітко побачити атрибути, методи та кратності зв’язків між компонентами, що забезпечує цілісність архітектури та логічність реалізації технології.

Додатково побудовано UML-діаграму об’єктів, яка демонструє конкретний стан технології в момент виконання аналізу. На ній наведені екземпляри трьох пасажирів, а також об’єкти модулів завантаження, обробки, розвідувального аналізу, передбачення та візуалізації. Діаграма наочно показує, як саме реальні дані проходять через усі етапи технології та як формується результат передбачення на основі найточнішої моделі — п’ятишарової нейронної мережі. Це підтверджує узгодженість між логічним проектуванням та фактичною реалізацією механізмів аналізу.

3. РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПЕРЕДБАЧЕННЯ РІВНІВ ЗАДОВОЛЕНОСТІ ПАСАЖИРІВ АВІАКОМПАНІЯМИ

3.1 Розробка інформаційної технології

Для початку було створено нейронну мережу з трьома шарами за допомогою бібліотеки Keras. Модель містить два приховані шари з 64 та 32 нейронами відповідно, у яких використовується активаційна функція ReLU, та вихідний шар з одним нейроном і функцією активації sigmoid, що дозволяє розв'язувати задачу бінарної класифікації. Мережу було скомпільовано з функцією втрат binary_crossentropy та оптимізатором adam. Навчання тривало 10 епох та розподілом даних на тренувальний і валідаційний набори (80/20) (рис. 3.1).

```
# NN with 3 layers
model = Sequential()
model.add(Dense(64, input_dim=X_train.shape[1], activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

history = model.fit(X_train, y_train, epochs=10, batch_size=32, validation_split=0.2)
loss_before, keras_accuracy_before = model.evaluate(X_test, y_test)
print("Keras Accuracy (NN with 3 layers:", keras_accuracy_before)
```

Рисунок 3.1 – Нейронна мережа з трьома шарами

Після навчання моделі були отримані передбачення на тренувальному та тестовому наборах даних із порогом 0.5 для класифікації. Далі обчислено точність класифікації (accuracy) на обох наборах для оцінки якості навченої нейронної мережі (рис. 3.2).

```
567/567 [==============================] - 1s 1ms/step
243/243 [==============================] - 0s 1ms/step
Train Accuracy (NN with 3 layers): 0.9554157700160018
Test Accuracy (NN with 3 layers): 0.9427284427284427
```

Рисунок 3.2 – Результат передбачення для тренувального та тестового набору

Також для тестового набору було побудовано матрицю конфузії, що наочно демонструє якість класифікації моделі за кількістю правильних та помилкових передбачень для кожного класу. Матриця візуалізована за допомогою теплової карти з підписами класів (рис. 3.3).

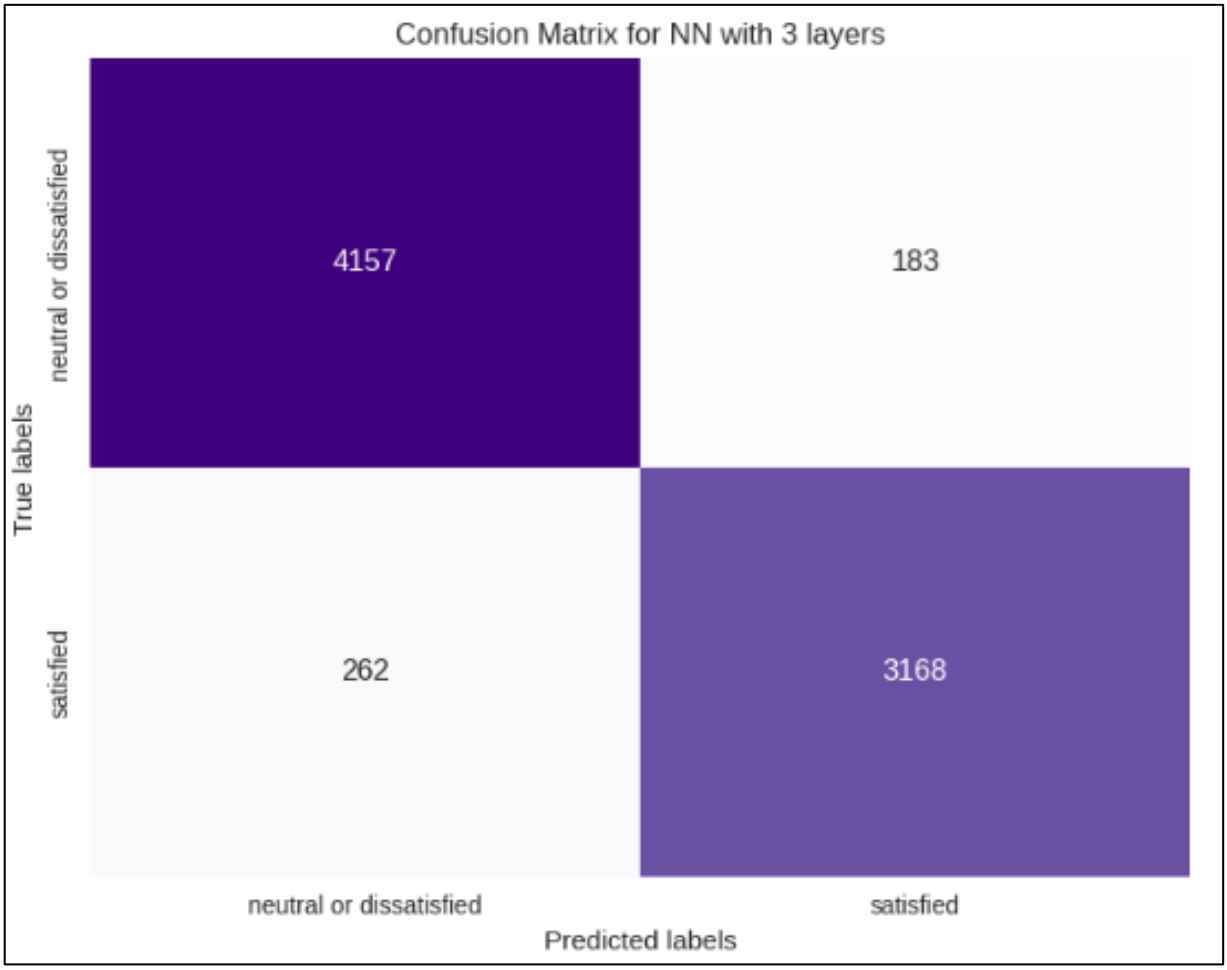
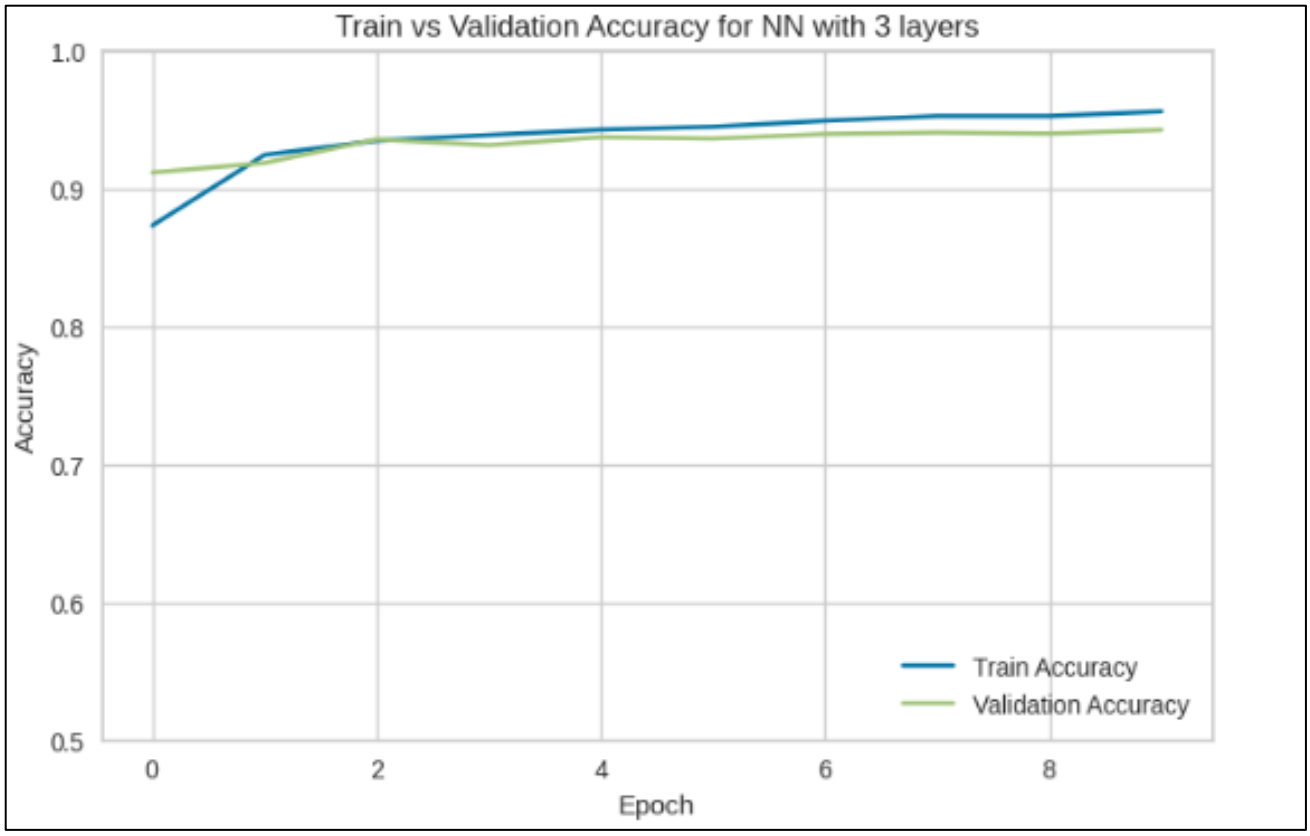


Рисунок 3.3 – Матриця конфузії для тестового набору

Матриця неточностей для тестового набору показує, що модель нейронної мережі з трьома шарами правильно класифікувала 4157 випадків як «нейтральний або незадоволений» і 3168 як «задоволений». Водночас було 183 помилкових класифікацій задоволених як незадоволених, і 262 — навпаки.

Загальна точність моделі на тестових даних становить приблизно 94,27%.

У ході дослідження було побудовано графік зміни точності моделі на тренувальній та валідаційній вибірках упродовж епох навчання (рис. 3.4).



Рисунки 3.4 – Графік точності навчання та валідації

З графіку видно, що обидві криві зростають і стабілізуються на високому рівні, що свідчить про добре узгодження моделі без ознак перенавчання.

Також було розраховано класифікаційний звіт, який містить основні метрики якості моделі — точність (precision), повноту (recall), F1-міру та кількість прикладів (support) для кожного класу: «neutral or dissatisfied» та «satisfied» (рис. 3.5).

	precision	recall	f1-score	support
neutral or dissatisfied	0.94	0.96	0.95	4340
satisfied	0.95	0.92	0.93	3430
accuracy			0.94	7770
macro avg	0.94	0.94	0.94	7770
weighted avg	0.94	0.94	0.94	7770

Рисунок 3.5 – Класифікаційний звіт

Далі була побудована 4-шарова нейронна мережа, яка розширює архітектуру попередньої 3-шарової моделі. У цій моделі збільшено кількість нейронів у першому шарі до 96, додано шар Dropout з коефіцієнтом 0.2 для зменшення перенавчання, а також введено додатковий прихований шар з 64 нейронами, що робить мережу глибшою і здатнішою захоплювати складніші закономірності. Крім того, змінено параметри навчання — збільшено кількість епох до 20 і розмір батчу до 48. Використано колбек ReduceLROnPlateau, який автоматично знижує швидкість навчання при відсутності покращення валідаційної втрати, що покращує стабільність навчання (рис. 3.6).

```
# NN with 4 layers
model2 = Sequential()
model2.add(Dense(96, input_dim=X_train.shape[1], activation='relu'))
model2.add(Dropout(0.2))
model2.add(Dense(64, activation='relu'))
model2.add(Dense(32, activation='relu'))
model2.add(Dense(1, activation='sigmoid'))

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=3, verbose=1)

model2.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

history2 = model2.fit(
    X_train, y_train,
    epochs=20,
    batch_size=48,
    validation_split=0.2,
    callbacks=[reduce_lr],
    verbose=1
)

loss2, accuracy2 = model2.evaluate(X_test, y_test)
print("Keras Accuracy (NN with 4 layers):", accuracy2)
```

Рисунок 3.6 – 4-шарова нейронна мережа

Для 4-шарової нейронної мережі точність класифікації на навчальній вибірці склала приблизно 96.3%, що свідчить про ефективне навчання моделі. Точність на тестових даних становила близько 95.2%, що підтверджує хорошу здатність моделі до узагальнення і високу якість передбачень на нових, раніше невідомих даних. На рисунку 3.7 зображено матрицю конфузії для тестового набору даних.

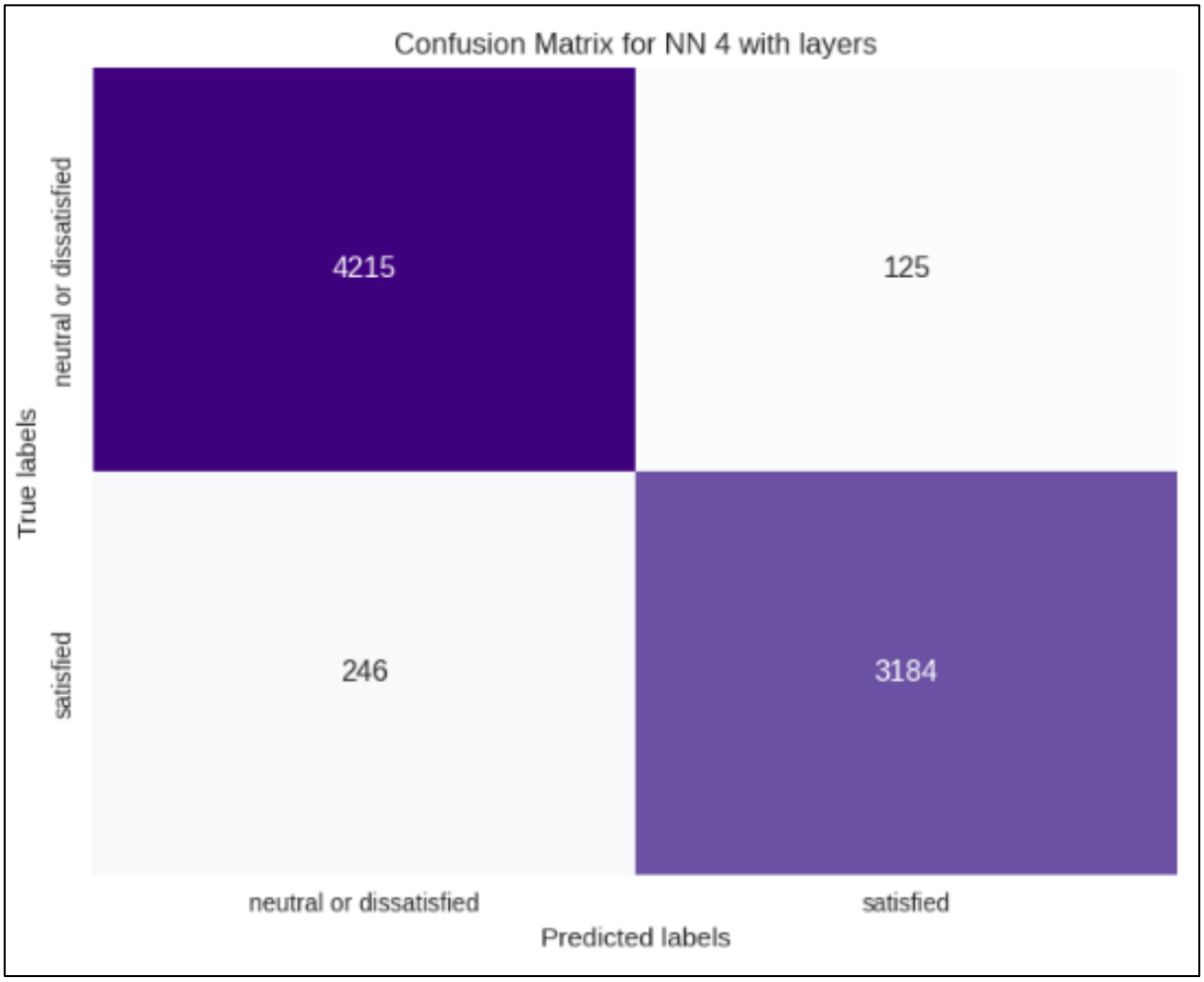


Рисунок 3.7 – Матриця конфузії для тестового набору

Далі було побудовано криву навчання для 4-шарової нейронної мережі, яка показує зміну точності на тренувальній та валідаційній вибірках протягом епох. На графіку видно, як з кожною епохою модель покращує свої результати, і це дозволяє оцінити динаміку навчання, стабільність та наявність перенавчання. Крива демонструє, що модель навчається впевнено, а валідаційна точність залишається на високому рівні, наближеному до тренувальної (рис. 3.8).

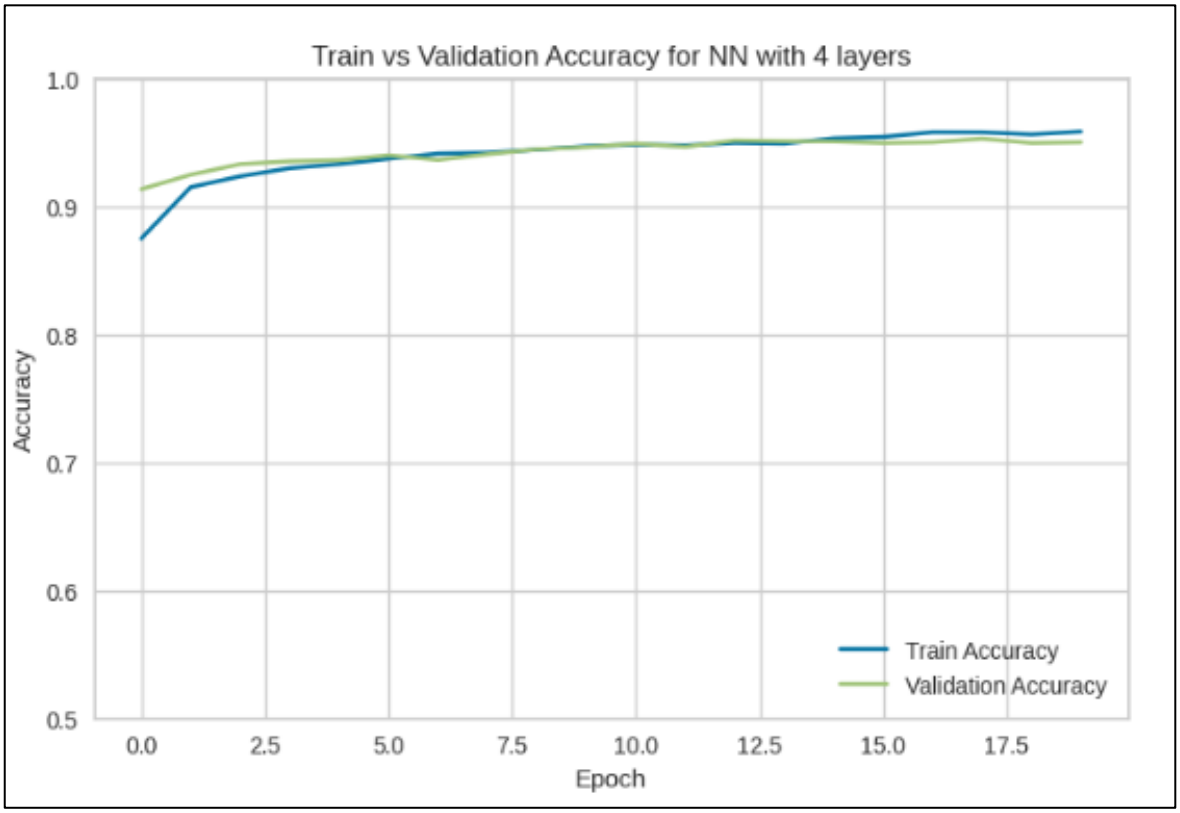


Рисунок 3.8 – Крива навчання

На рисунку 3.9 з класифікаційним звітом видно, що модель забезпечує високу якість класифікації для обох класів. Значення precision, recall та f1-score перебувають у межах 0.93–0.97, що свідчить про збалансовану роботу. Загальна точність становить 95%, отже модель добре узагальнює і не має суттєвого перекосу в один із класів.

	precision	recall	f1-score	support
neutral or dissatisfied	0.94	0.97	0.96	4340
satisfied	0.96	0.93	0.94	3430
accuracy			0.95	7770
macro avg	0.95	0.95	0.95	7770
weighted avg	0.95	0.95	0.95	7770

Рисунок 3.9 – Класифікаційний звіт 4-шарової нейронної мережі

У наступному етапі було побудовано 5-шарову нейронну мережу, яка має ще глибшу архітектуру порівняно з попередньою моделлю. У цій мережі додано додатковий прихований шар на 16 нейронів, а також збільшено кількість нейронів у першому шарі до 128. Змінено параметри навчання: кількість епох зросла до 30, розмір батчу — до 64, що дозволяє моделі краще адаптуватися до даних. Шар Dropout залишено з більшим коефіцієнтом (0.3), що посилює регуляризацію. Дана модель показала точність на навчальній вибірці — 96.6%, а на тестовій — 95.35%. У порівнянні з 4-шаровою моделлю (95.2% на тесті), покращення є незначним, але помітним. Це свідчить про те, що додаткова глибина моделі не призвела до перенавчання і дозволила дещо підвищити якість класифікації.

На рисунку 3.10 зображено матрицю конфузії для тестового набору.

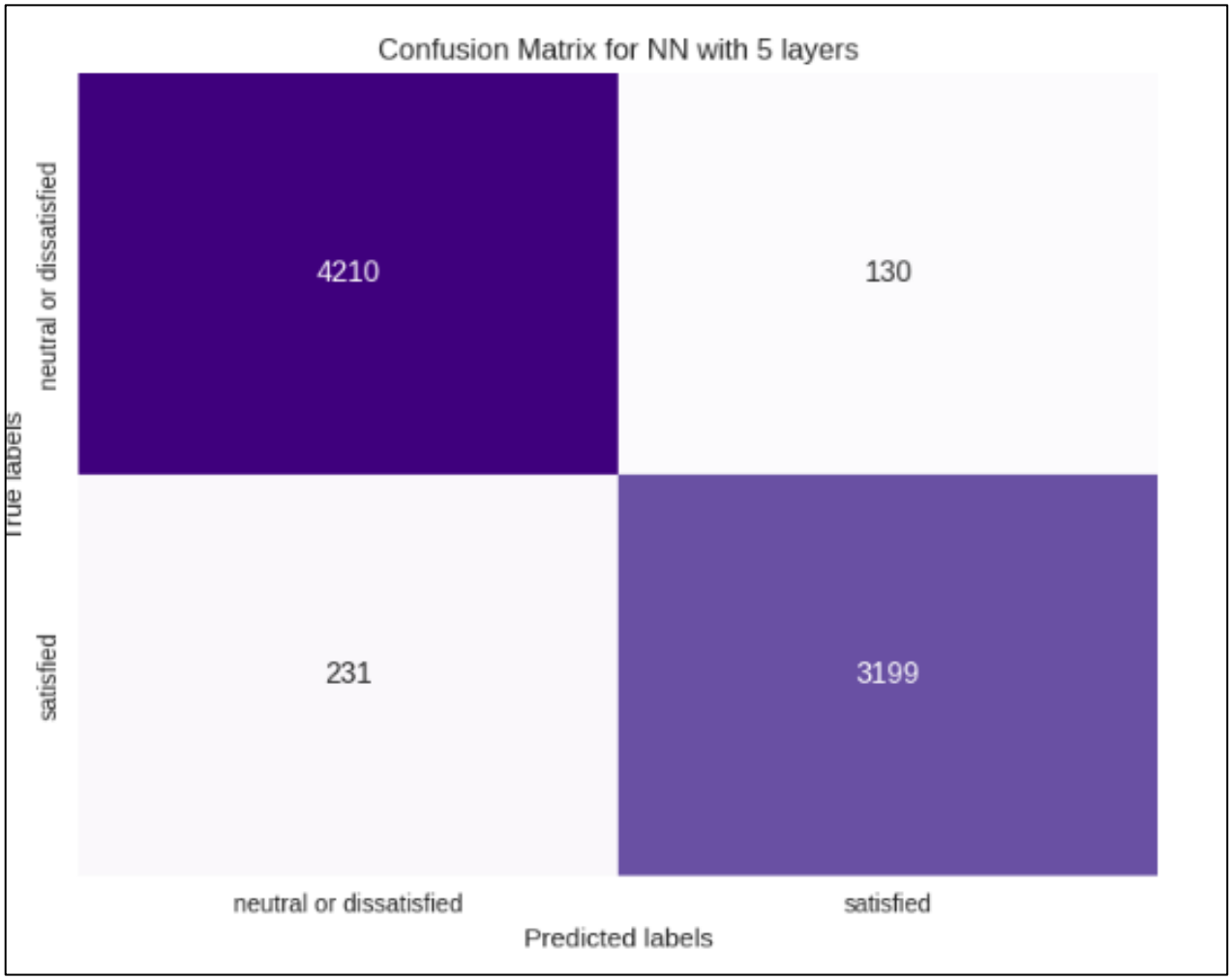


Рисунок 3.10 – Матриця конфузії для тестового набору

Матриця конфузії для 5-шарової моделі свідчить про її високу ефективність: з 4340 фактично позитивних прикладів правильно класифіковано 4210, а з 3430 негативних — 3199. Кількість помилок залишилася низькою: 130 хибнонегативних і 231 хибнопозитивний випадок. Це підтверджує, що модель добре справляється з обома класами, забезпечуючи збалансовану класифікацію та загальну точність 95.35% на тестових даних.

На рисунку 3.11 зображено класифікаційний звіт для даної моделі.

	precision	recall	f1-score	support
neutral or dissatisfied	0.95	0.97	0.96	4340
satisfied	0.96	0.93	0.95	3430
accuracy			0.95	7770
macro avg	0.95	0.95	0.95	7770
weighted avg	0.95	0.95	0.95	7770

Рисунок 3.11 – Класифікаційний звіт 5-шарової моделі

Класифікаційний звіт для 5-шарової моделі показує високі значення precision, recall та f1-score для обох класів (від 0.93 до 0.97). Загальна точність становить 95%, що свідчить про збалансовану та надійну роботу моделі при класифікації обох типів пасажирів.

На рисунку 3.12 зображено криву навчання для 5-шарової нейронної мережі.

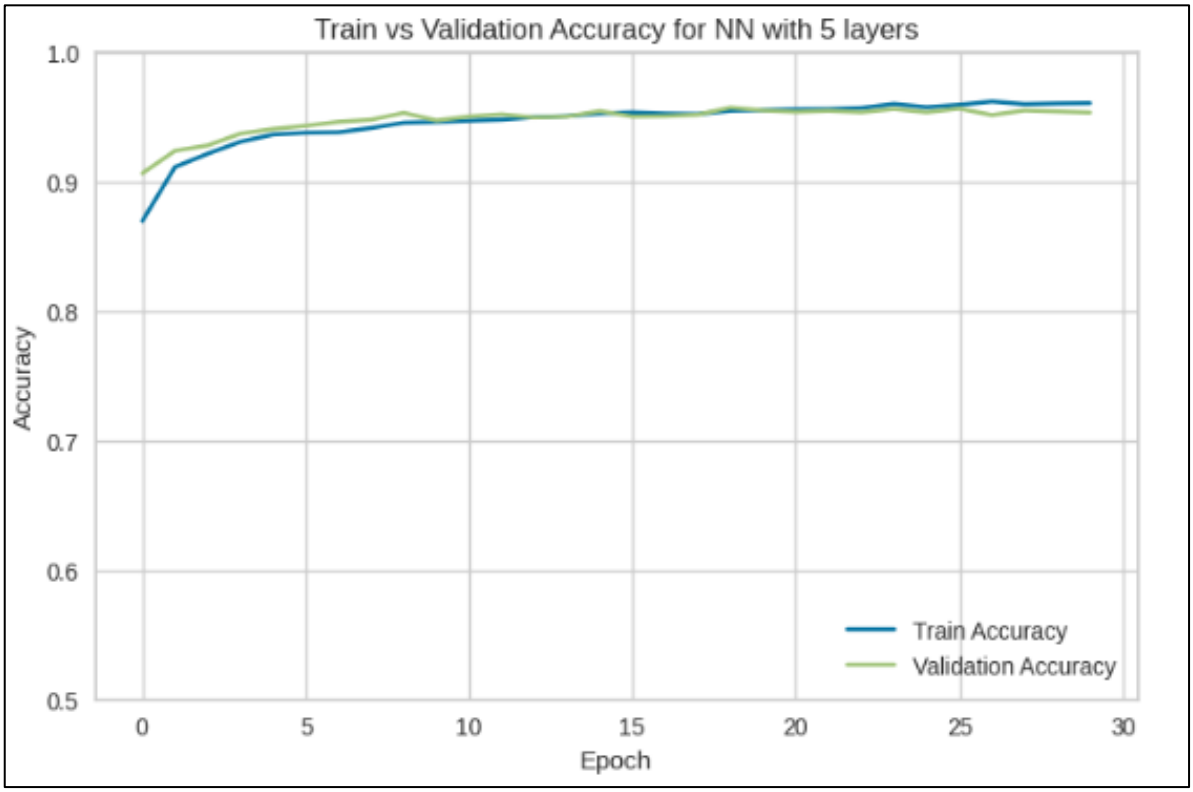


Рисунок 3.12 – Крива навчання для 5-шарової нейронної мережі

Крива навчання для 5-шарової моделі показує стабільне зростання точності на тренувальній вибірці та високі значення на валідації без різкого розриву між ними. Це свідчить про відсутність перенавчання та добре узгодження моделі з даними, що підтверджує її надійність і якість навчання.

Також для порівняння була побудована модель MLP з бібліотеки sklearn з двома прихованими шарами (32 і 16 нейронів), регуляризацією (alpha=0.01) та ранньою зупинкою для запобігання перенавчанню. Модель навчалась до 300 ітерацій.

Точність на навчальній вибірці становила 95.5%, а на тестовій — 94.0%, що трохи поступається глибшим нейронним мережам, але все одно демонструє гідний рівень класифікації.

На рисунку 3.13 зображено класифікаційний звіт для MLP-моделі sklearn.

	precision	recall	f1-score	support
0	0.94	0.95	0.95	4340
1	0.94	0.93	0.93	3430
accuracy			0.94	7770
macro avg	0.94	0.94	0.94	7770
weighted avg	0.94	0.94	0.94	7770

Рисунок 3.13 – Класифікаційний звіт для MLP-моделі sklearn

Класифікаційний звіт моделі MLP sklearn показує рівномірні та високі значення precision, recall і f1-score близько 0.94 для обох класів. Загальна точність становить 94%, що свідчить про стабільну та збалансовану роботу моделі при класифікації обох категорій.

Для магістерської роботи було обрано додатково три моделі машинного навчання: XGBoost, Gaussian Process Classifier (GPC) та Ridge Classifier.

Використання моделі XGBoost наведено на рисунку 3.14, де показано передбачення рівнів задоволеності пасажирів та відповідну матрицю конфузії для тестового набору даних.

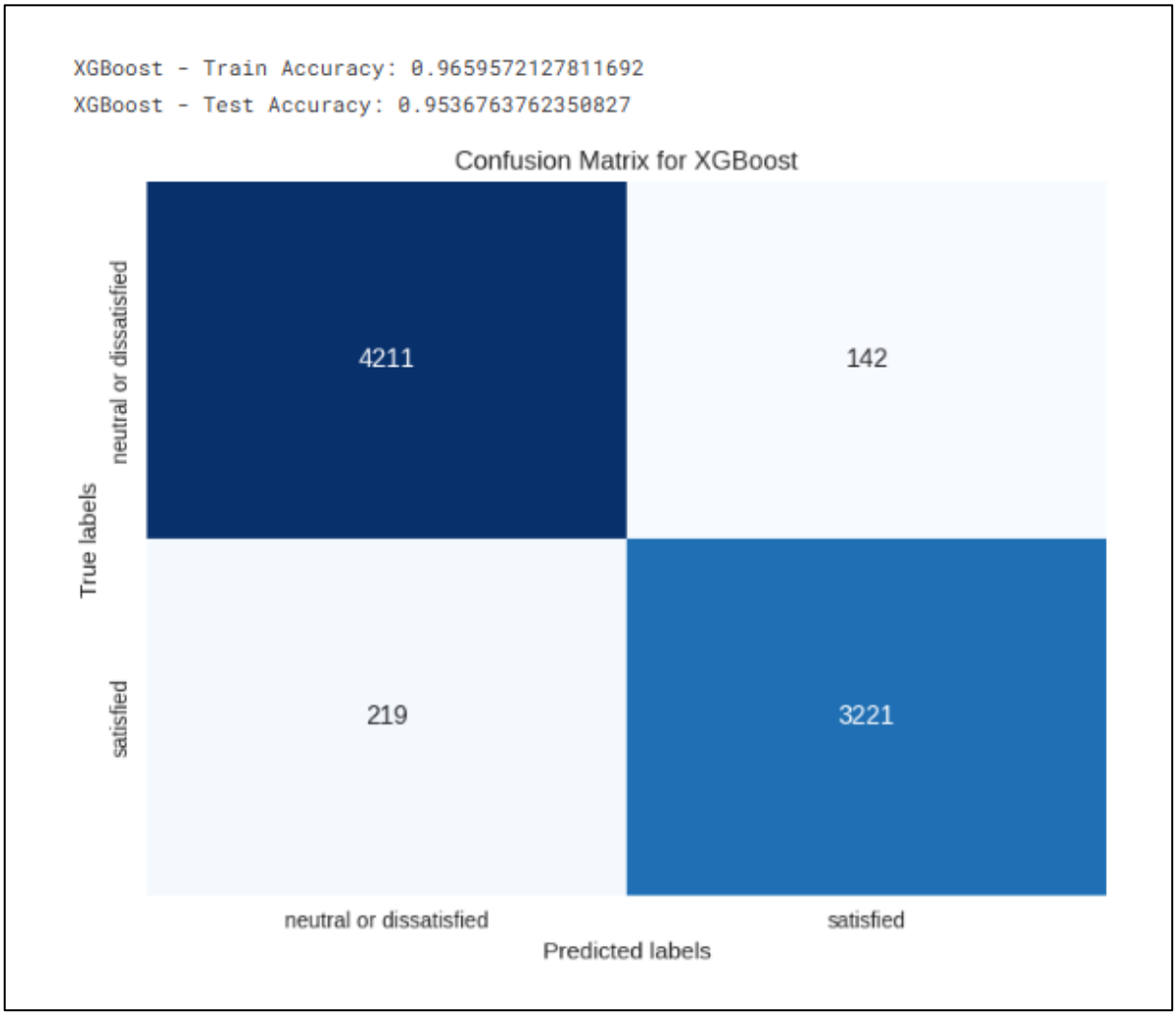


Рисунок 3.14 – Матриця конфузії для моделі XGBoost

Модель XGBoost продемонструвала високу точність: Train Accuracy = 0.966, Test Accuracy = 0.954, що свідчить про ефективне навчання та здатність моделі узагальнювати на нових даних. Матриця конфузії показує, що більшість зразків класифіковані правильно (4211 + 3221), а помилки зустрічаються рідко: 142 пасажири, які були незадоволені, передбачені як задоволені, та 219 задоволених — як незадоволені, що підтверджує надійність моделі для даного набору даних.

На рисунку 3.15 зображено звіт класифікації моделі XGBoost.

	precision	recall	f1-score	support
neutral or dissatisfied	0.95	0.97	0.96	4353
satisfied	0.96	0.94	0.95	3440
accuracy			0.95	7793
macro avg	0.95	0.95	0.95	7793
weighted avg	0.95	0.95	0.95	7793

Рисунок 3.15 – Звіт класифікації моделі XGBoost

Модель показала високі показники якості класифікації: precision, recall та F1-score для обох класів знаходяться близько 0.95–0.96, що свідчить про правильну роботу моделі як для задоволених, так і для незадоволених пасажирів. Загальна точність (accuracy = 0.95) підтверджує надійність моделі на тестових даних.

Gaussian Process Classifier (GPC) — ймовірнісна модель, яка використовує гаусові процеси для розділення класів і оцінки невизначеності прогнозів. GPC особливо ефективна при нелінійних або складно роздільних даних.

Використання моделі GPC наведено на рисунку 3.16, де показано передбачення рівнів задоволеності пасажирів та відповідну матрицю плутанини для тестового набору даних.

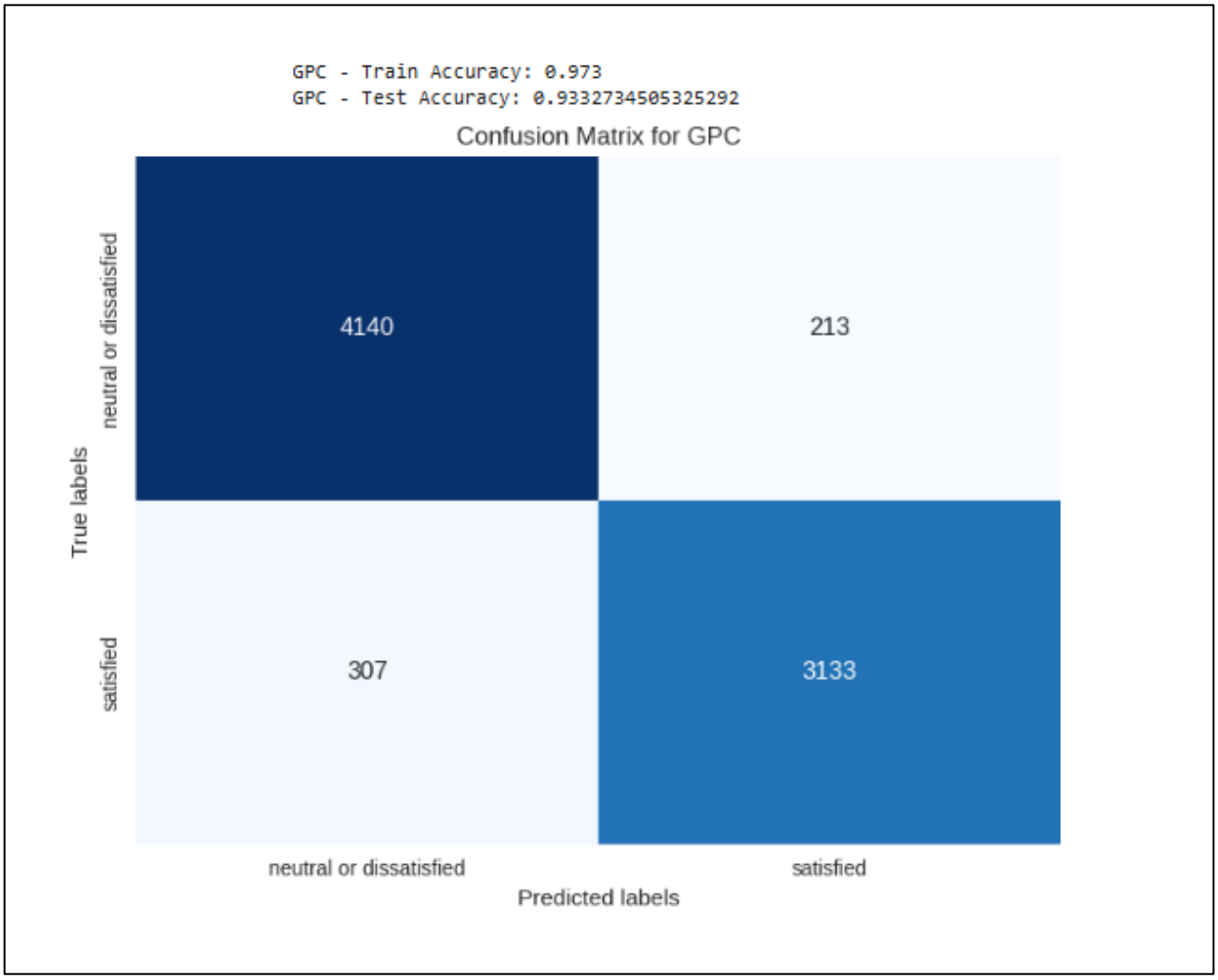


Рисунок 3.16 – Матриця конфузії для моделі GPC

Модель GPC показала Train Accuracy = 0.973 та Test Accuracy = 0.933, що свідчить про високу здатність моделі до узагальнення. Матриця плутанини показує, що більшість зразків класифіковані правильно (4140 + 3133), проте помилки все ще трапляються: 213 пасажирів незадоволені були передбачені як задоволені, а 307 задоволених — як незадоволені, що вказує на невелику кількість неправильних передбачень та високу точність моделі.

На рисунку 3.17 зображено звіт класифікації моделі GPC.

Classification Report (Test Data):

	precision	recall	f1-score	support
neutral or dissatisfied	0.93	0.95	0.94	4353
satisfied	0.94	0.91	0.92	3440
accuracy			0.93	7793
macro avg	0.93	0.93	0.93	7793
weighted avg	0.93	0.93	0.93	7793

Рисунок 3.17 – Звіт класифікації моделі GPC

Модель GPC продемонструвала високі та збалансовані показники класифікації, з precision, recall та F1-score близько 0.95 для обох класів, що свідчить про здатність моделі правильно класифікувати як задоволених, так і незадоволених пасажирів. Загальна точність (accuracy = 0.95) підтверджує високу ефективність моделі на тестових даних.

– Ridge Classifier — лінійний метод з регуляризацією, який зменшує ризик перенавчання та забезпечує стабільні результати при великій кількості ознак [12].

Використання моделі Ridge Classifier наведено на рисунку 3.18, де показано передбачення рівнів задоволеності пасажирів та відповідну матрицю конфузії для тестового набору даних.

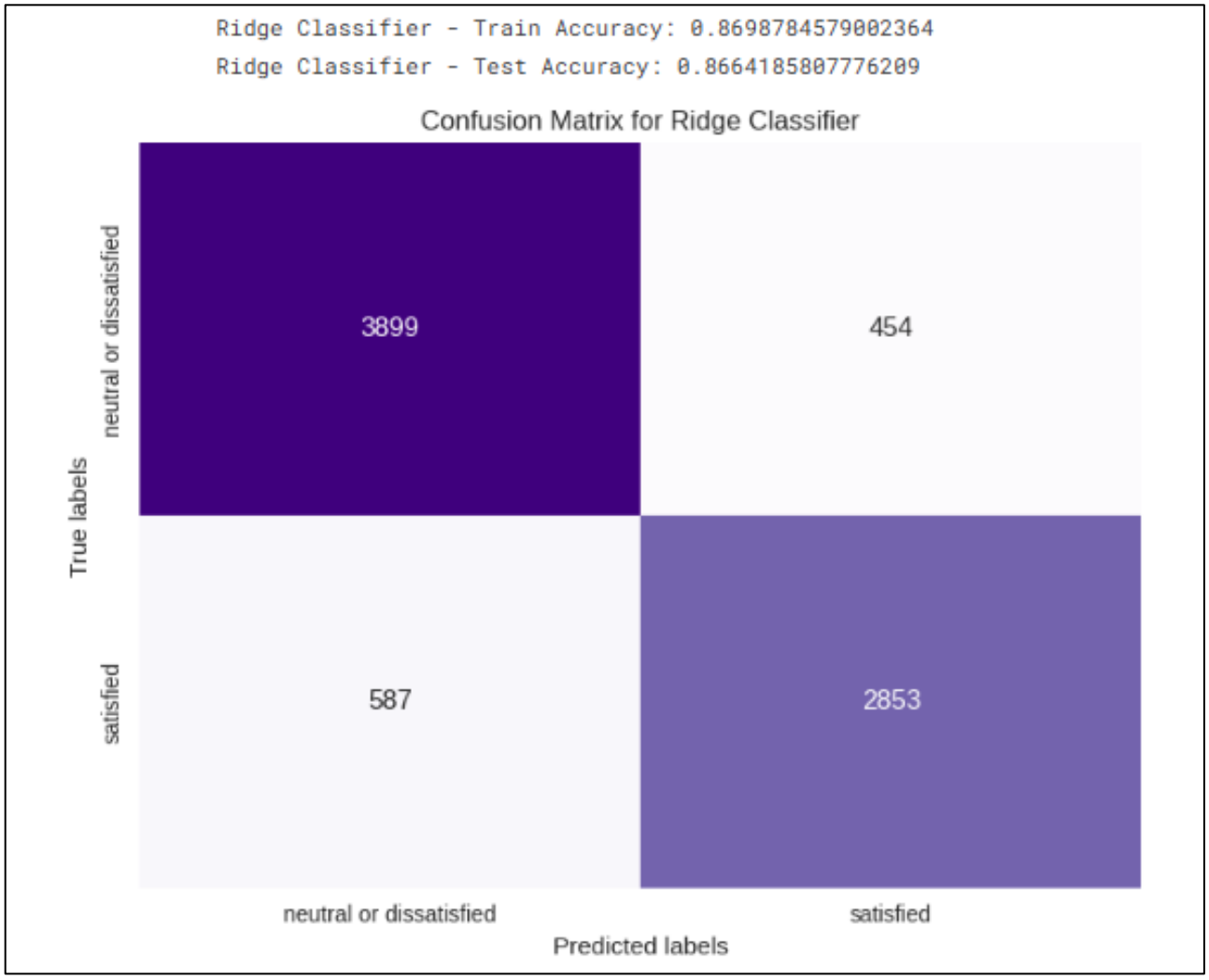


Рисунок 3.18 – Матриця конфузії для моделі Ridge Classifier

Результати Ridge Classifier показують збалансовану точність на тренувальних і тестових даних, з невеликою кількістю помилок у класифікації. Модель продемонструвала Train Accuracy = 0.870 та Test Accuracy = 0.866, а матриця плутанини показує, що більшість зразків класифіковані правильно (3899 + 2853), проте 454 незадоволених пасажирів були передбачені як задоволені, а 578 задоволених — як незадоволені, що вказує на характерні, але незначні помилки класифікації.

На рисунку 3.19 зображено звіт класифікації моделі Ridge Classifier.

	precision	recall	f1-score	support
neutral or dissatisfied	0.87	0.90	0.88	4353
satisfied	0.86	0.83	0.85	3440
accuracy			0.87	7793
macro avg	0.87	0.86	0.86	7793
weighted avg	0.87	0.87	0.87	7793

Рисунок 3.19 – Звіт класифікації моделі Ridge Classifier

Модель Ridge Classifier продемонструвала збалансовані показники класифікації: precision, recall та F1-score для обох класів знаходяться близько 0.83–0.90, що свідчить про задовільну здатність моделі правильно класифікувати як задоволених, так і незадоволених пасажирів. Загальна точність (accuracy = 0.87) підтверджує помірну ефективність моделі на тестових даних.

Використання цих моделей дозволяє розширити підхід бакалаврської роботи, оцінити альтернативні алгоритми та порівняти їх точність з нейронними мережами.

На рисунку 3.20 зображено порівняння точності моделей

	Model	Train Precision	Test Precision
0	NN with 3 layers	0.967	0.952
1	NN with 4 layers	0.966	0.946
2	NN with 5 layers	0.986	0.974
3	MLP_sklearn	0.954	0.937
4	XGBoost	0.966	0.954
5	GPC	0.973	0.933
6	Ridge Classifier	0.870	0.866

Рисунок 3.20 – Загальні результати

За наведеними результатами найкращою моделлю є 5-шарова нейронна мережа (NN with 5 layers). Вона демонструє найвищі показники точності як на тренувальній вибірці (0.986), так і на тестовій (0.974), що свідчить про її кращу здатність до узагальнення та ефективніше навчання порівняно з іншими моделями.

3.2 Оцінка впливу ознак на результат моделі

Для глибшого розуміння, які ознаки мають найбільший вплив на передбачення 5-шарової нейронної мережі, було застосовано метод SHAP (SHapley Additive exPlanations). Цей метод дозволяє інтерпретувати вклад кожної вхідної змінної в кінцеве рішення моделі [25].

Для аналізу взяли випадкові 100 зразків з тренувальних даних (X_train) як фон. Потім створили пояснювач DeepExplainer для найкращої моделі — нейромережі з 5 шарів (model1) — і порахували SHAP-значення для перших 100 зразків тестових даних (X_test).

На рисунку 3.21 зображено графік впливу ознак.

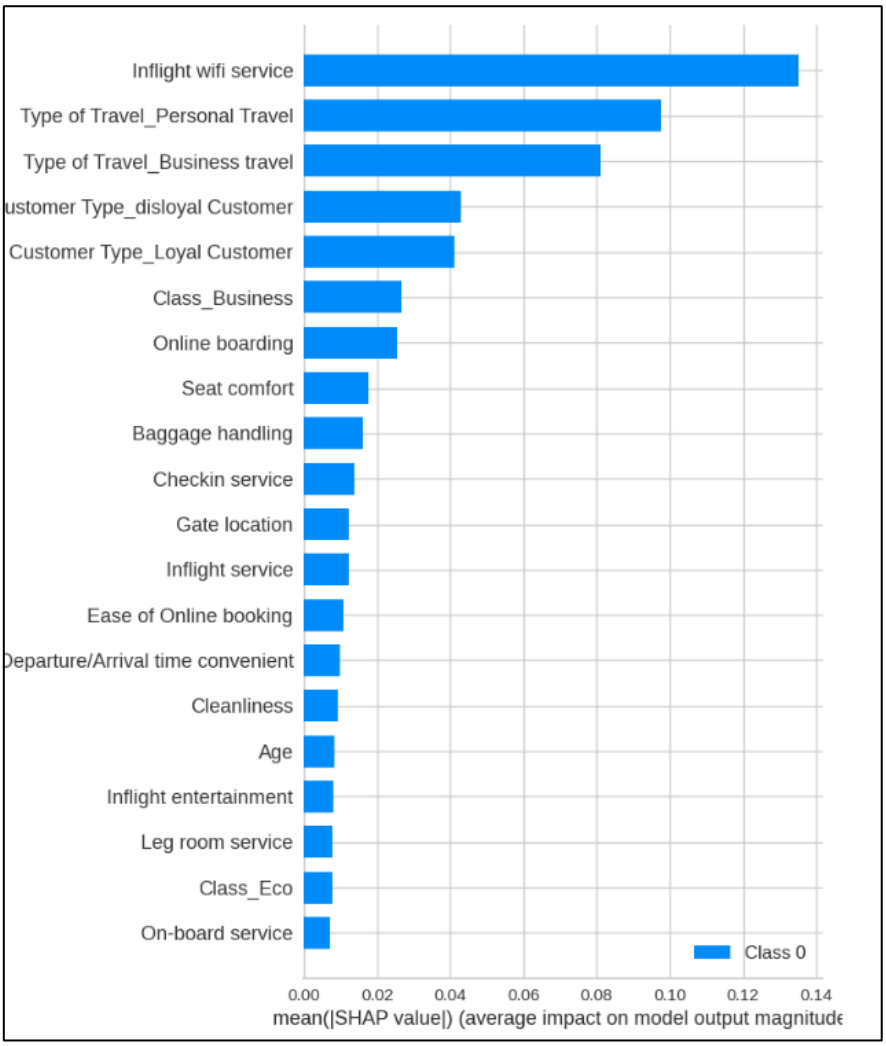


Рисунок 3.21 – Графік впливу ознак

Найбільший вплив на модель має ознака Inflight wifi service — вона найбільш суттєво впливає на передбачення. Далі за значущістю йдуть Type of Travel (Personal Travel, Business travel) та Customer Type (disloyal і loyal customers) [26]. Інші ознаки мають поступово менший вплив, при цьому всі вони враховуються моделлю.

3.3 Висновки

У даному розділі розроблено та реалізовано послідовний алгоритм для побудови моделей передбачення задоволеності пасажирів авіакомпаній. Створено кілька нейронних мереж різної глибини (3, 4 та 5 шарів), модель MLP зі sklearn, а також додатково застосовано моделі XGBoost, GPC та Ridge Classifier. Найкращі результати серед нейронних мереж показала 5-шарова модель, яка продемонструвала найвищу точність на тренувальних і тестових даних, підтверджуючи ефективність навчання та здатність до узагальнення. Моделі XGBoost і GPC показали значно вищу точність порівняно з лінійною моделлю Ridge Classifier, яка продемонструвала трохи нижчі результати.

Аналіз кривих навчання нейронних мереж засвідчив відсутність перенавчання, а матриці плутанини та класифікаційні звіти для всіх моделей підтвердили високу якість класифікації як для задоволених, так і для незадоволених пасажирів.

За допомогою методу SHAP визначено, що найбільший вплив на передбачення має ознака Inflight wifi service, а також значущі ознаки Type of Travel і Customer Type, що допомагає краще інтерпретувати роботу моделей. Загалом, 5-шарова нейронна мережа та моделі XGBoost і GPC є оптимальним вибором для задачі класифікації задоволеності пасажирів, забезпечуючи високу точність та надійність передбачень.

4 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу так і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями [27]. Результати оцінювання зведемо до таблиці 4.1.

Таблиця 4.1 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	5	4	5
2. Ринкові переваги (наявність аналогів)	4	3	3
3. Ринкові переваги (ціна продукту)	4	4	3
4. Ринкові переваги (технічні властивості)	4	4	4
5. Ринкові переваги (експлуатаційні витрати)	4	4	4
6. Ринкові перспективи (розмір ринку)	3	3	3
7. Ринкові перспективи (конкуренція)	3	2	2
8. Практична здійсненність (наявність фахівців)	4	4	4
9. Практична здійсненність (наявність фінансів)	2	2	2
10. Практична здійсненність (необхідність нових матеріалів)	3	4	4
11. Практична здійсненність (термін реалізації)	4	4	4
12. Практична здійсненність (розробка документів)	3	2	3
Сума балів	43	40	41
Середньоарифметична сума балів $СБ_с$	41,3		

За результатами розрахунків, наведених в таблиці 4.1, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в [27].

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» становить 41,3 бала, що, відповідно до [27], свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

4.2 Розрахунок узагальненого коефіцієнта якості розробки

Узагальнений коефіцієнт якості ($В_н$) для нового технічного рішення розрахуємо за формулою [28]:

$$B_н = \sum_{i=1}^{k} \alpha_i \cdot \beta_i , \qquad (4.1)$$

де k – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

α_i – коефіцієнт, який враховує питому вагу *i-го* технічного показника в загальній якості розробки. Коефіцієнт α_i визначається експертним шляхом і при цьому має виконуватись умова $\sum_{i=1}^{k} \alpha_i = 1$;

β_i – відносне значення *i-го* технічного показника якості нової розробки.

Результати порівняння зведемо до таблиці 4.2.

Таблиця 4.2 – Порівняння основних параметрів розробки та аналога.

Показники (параметри)	Одиниця вимірю-вання	Аналог	Проектований продукт	Відношення параметрів нової розробки до аналога	Питома вага показника
1. Кількість використаних моделей машинного навчання	од.	2	3	1,5	0,15
2. Попередня обробка та очистка даних	од.	1	2	2	0,2
3. Точність передбачення	%	93	97	1,04	0,3
4. Кількість графіків розвідувального аналізу	од.	3	10	3,33	0,2
5. Кількість використаних багатошарових нейронних мереж	од.	1	4	4	0,15

Узагальнений коефіцієнт якості ($B_н$) для нового технічного рішення складе:

$$B_н = \sum_{i=1}^{k} \alpha_i \cdot \beta_i = 1{,}5 \cdot 0{,}15 + 2 \cdot 0{,}2 + 1{,}04 \cdot 0{,}3 + 3{,}33 \cdot 0{,}2 + 4 \cdot 0{,}15 = 2{,}20.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,20 рази.

4.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов’язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

4.3.1 Витрати на оплату праці

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ($З_о$) розраховуємо у відповідності до посадових окладів працівників, за формулою [27]:

$$З_о = \sum_{i=1}^{k} \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.2)$$

де k – кількість посад дослідників залучених до процесу досліджень;

M_{ni} – місячний посадовий оклад конкретного дослідника, грн;

t_i – число днів роботи конкретного дослідника, дн.;

T_p – середнє число робочих днів в місяці, T_p=20 дні.

$$З_о = 25200{,}00 \cdot 10 / 20 = 12600{,}00 \ (грн).$$

Проведені розрахунки зведемо до таблиці 4.3.

Таблиця 4.3 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту (проєктний менеджер)	25200,00	1260,00	10	12600,00
Консультант (менеджер сфери ground handling)	23500,00	1175,00	5	5875,00
Інженер-програміст	22000,00	1100,00	30	33000,00
Фахівець з аналітично-математичних досліджень	24000,00	1200,00	10	12000,00
Консультант-аналітик цифрових обчислюваних систем	29000,00	1450,00	5	7250,00
Всього				70725,00

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ($З_р$) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» розраховуємо за формулою:

$$З_р = \sum_{i=1}^{n} C_i \cdot t_i, \quad (4.3)$$

де C_i – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{М_М \cdot K_i \cdot K_c}{T_р \cdot t_{зм}}, \quad (4.4)$$

де $М_М$ – розмір мінімальної місячної заробітної плати, приймемо $Мм$=8000,00 грн;

K_i – коефіцієнт міжкваліфікаційного співвідношення (табл. Б.2, додаток Б) [27];

K_c – мінімальний коефіцієнт співвідношень місячних тарифних ставок;

$T_р$ – середнє число робочих днів в місяці, приблизно $T_р$ = 20 дн;

$t_{зм}$ – тривалість зміни, год.

C_1 = 8000,00 · 1,10 · 1,15 / (20 · 8) = 63,25 *(грн)*.
$З_{р1}$ = 63,25 · 8,00 = 506,00 *(грн)*.

Проведені розрахунки зведемо до таблиці 4.4.

Таблиця 4.4 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
Встановлення допоміжного офісного обладнання	8,00	2	1,10	63,25	506,00
Монтаж робочого місця розробника системи прогнозування	12,00	2	1,10	63,25	759,00
Інсталяція програмного забезпечення	5,00	5	1,70	97,75	488,75
Встановлення цифрових обчислювальних систем	3,00	4	1,50	86,25	258,75
Відлагодження інтерполяційних модулів	7,00	5	1,70	97,75	684,25
Тренування цифрової експериментальної моделі	4,50	4	1,50	86,25	388,13
Формування бази даних прогнозного аналізу	16,00	3	1,35	77,63	1242,00
Інші допоміжні роботи	10,00	3	1,35	77,63	776,25
Монтаж серверного обладнання	3,30	4	1,50	86,25	284,63
Всього					5387,75

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 … 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$З_{дод} = (З_о + З_р) \cdot \frac{Н_{дод}}{100\%}, \quad (4.5)$$

де $Н_{дод}$ – норма нарахування додаткової заробітної плати. Приймемо 10%.

$З_{дод}$ = (70725,00 + 5387,75) · 10 / 100% = 7611,28 (грн).

4.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_н = (З_о + З_р + З_{дод}) \cdot \frac{Н_{зп}}{100\%} \quad (4.6)$$

де $Н_{зп}$ – норма нарахування на заробітну плату. Приймаємо 22%.

Зн = (70725,00 + 5387,75 + 7611,28) · 22 / 100% = 18419,29 (грн).

4.3.3 Сировина та матеріали

Витрати на матеріали (*М*), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$М = \sum_{j=1}^{n} Н_j \cdot Ц_j \cdot К_j - \sum_{j=1}^{n} В_j \cdot Ц_{вj}, \quad (4.7)$$

де $Н_j$ – норма витрат матеріалу *j*-го найменування, кг;

n – кількість видів матеріалів;

$Ц_j$ – вартість матеріалу *j*-го найменування, грн/кг;

$К_j$ – коефіцієнт транспортних витрат, ($К_j$ = *1,1 ... 1,15*);

$В_j$ – маса відходів *j*-го найменування, кг;

$Ц_{вj}$ – вартість відходів *j*-го найменування, грн/кг.

М1 = 3,000 · 186,00 · 1,05 - 0 · 0 = 585,90 (грн).

Проведені розрахунки зведемо до таблиці 4.5.

Таблиця 4.5 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 од, грн	Норма витрат, од.	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний ECONOMIC (A4-500)	186,00	3,000	0	0	585,90
Папір для заміток ECONOMIC (A5)-60	96,00	4,000	0	0	403,20
Начиння канцелярське DATUM FX	180,00	3,000	0	0	567,00
Органайзер офісний DATUM Office	264,00	3,000	0	0	831,60
Картридж для принтера HP-210A	1250,00	2,000	0	0	2625,00
Диск оптичний VEKO-10 (CD-R)	29,00	4,000	0	0	121,80
Диск оптичний VEKO-W (CD-RW)	32,00	4,000	0	0	134,40
FLASH-пам'ять Kingstar (32 ГБ) Class 10	199,00	1,000	0	0	208,95
FLASH-пам'ять Kingstar (64 ГБ) Class 10 A	299,00	1,000	0	0	313,95
Всього					5791,80

4.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі *($К_в$)*, які використовують при проведенні НДР на тему «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями», розраховуємо, згідно з їхньою номенклатурою, за формулою:

$$К_в = \sum_{j=1}^{n} Н_j \cdot Ц_j \cdot К_j \qquad (4.8)$$

де $Н_j$ – кількість комплектуючих j-го виду, шт.;

$Ц_j$ – покупна ціна комплектуючих j-го виду, грн;

K_j – коефіцієнт транспортних витрат, (K_j = *1,1 … 1,15*).

$K_в = 1 \cdot 4640{,}00 \cdot 1{,}05 = 4872{,}00$ *(грн).*

Проведені розрахунки зведемо до таблиці 4.6.

Таблиця 4.6 – Витрати на комплектуючі

Найменування комплектуючих	Кількість, шт.	Ціна за штуку, грн	Сума, грн
Router TP-Link230	1	4640,00	4872,00
База даних (тестувальна) DataBase 64-A10	1	2850,00	2992,50
Пам'ять (SSD диск) Samsung 870 QVO 1TB 2.5" V-NAND 4bit MLC (QLC) SATA III (MZ-77Q1T0BW)	2	3200,00	6720,00
Всього			14584,50

4.3.5 Спецустаткування для наукових (експериментальних) робіт

Балансову вартість спецустаткування розраховуємо за формулою:

$$B_{спец} = \sum_{i=1}^{k} Ц_i \cdot C_{пр.i} \cdot K_i \ , \qquad (4.9)$$

де $Ц_i$ – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{пр.i}$ –кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, (K_i = *1,10…1,12);*

k – кількість найменувань устаткування.

$B_{спец} = 48499{,}00 \cdot 1 \cdot 1{,}05 = 50923{,}95$ *(грн).*

Отримані результати зведемо до таблиці 4.7.

Таблиця 4.7 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Серверне обладнання на основі ПК ZEVS PC 13430U i5 9400F + GTX 1060 3GB	1	48499,00	50923,95
Маршрутизатор XIAOMI MI WIFI ROUTER 4C GLOBAL (DVB4231GL)	1	1099,00	1153,95
Всього			52077,90

4.3.6 Програмне забезпечення для наукових (експериментальних) робіт

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{прг} = \sum_{i=1}^{k} Ц_{iпрг} \cdot C_{прг.i} \cdot K_i \ , \qquad (4.10)$$

де $Ц_{iпрг}$ – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{прг.i}$ – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ($K_i = 1,10…1,12$);

k – кількість найменувань програмних засобів.

$B_{прг} = 9860,00 \cdot 1 \cdot 1,05 = 10353,00$ (грн).

Отримані результати зведемо до таблиці 4.8.

Таблиця 4.8 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Середовище мови програмування Python	1	9860,00	10353,00
Середовище розробки програмного забезпечення PyCharm	1	7580,00	7959,00
Програмне забезпечення розробки Kaggle	1	1059,00	1111,95
Високошвидкісний доступ до мережі (міс)	2	410,00	861,00
Навчальна база даних нейромережі карток опитування	1	12399,00	13018,95
Всього			33303,90

4.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (4.11)$$

де $Ц_б$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$ – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$A_{обл} = (38599{,}00 \cdot 2) / (3 \cdot 12) = 2144{,}39$ (грн).

Проведені розрахунки зведемо до таблиці 4.9.

Таблиця 4.9 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Програмно-аналітичний комплекс Компютер ARTLINE X39 v67 (X39v67) Intel Core i5-11400F / RAM 16ГБ / SSD 1ТБ / nVidia GeForce RTX 3060 12ГБ	38599,00	3	2	2144,39
Персональний комп'ютер Ноутбук HP Pavilion 15-e001	52399,00	3	2	2911,06
Спеціалізоване робоче місце розробника інформаційної технології	8400,00	5	2	280,00
Пристрій виводу текстової інформації Принтер HP Laser	11299,00	4	2	470,79
Оргтехніка Samsung Office	11399,00	5	2	379,97
Приміщення лабораторії розробки та дослідження	505000,00	35	2	2404,76
ОС Windows 11	8628,00	3	2	479,33
Прикладний пакет Microsoft Office 2021 Professional Plus	7320,00	3	2	406,67
Мережеве обладнання передачі цифрових даних	6700,00	4	2	279,17
Всього				9756,13

4.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^{n} \frac{W_{yi} \cdot t_i \cdot Ц_e \cdot K_{вni}}{\eta_i}, \qquad (4.12)$$

де W_{yi} – встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i – тривалість роботи обладнання на етапі дослідження, год;

$Ц_е$ – вартість 1 кВт-години електроенергії, грн; приймемо $Ц_е$ = 12,56 грн;

$К_{вп і}$ – коефіцієнт, що враховує використання потужності, $К_{вп і} < 1$;

η_i – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$В_е = 0,25 \cdot 240,0 \cdot 12,56 \cdot 0,95 / 0,97 = 753,60$ *(грн).*

Проведені розрахунки зведемо до таблиці 4.10.

Таблиця 4.10 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Програмно-аналітичний комплекс Компютер ARTLINE X39 v67 (X39v67) Intel Core i5-11400F / RAM 16ГБ / SSD 1ТБ / nVidia GeForce RTX 3060 12ГБ	0,25	240,0	753,60
Персональний комп'ютер Ноутбук HP Pavilion 15-e001	0,08	240,0	241,15
Спеціалізоване робоче місце розробника інформаційної технології	0,07	240,0	211,01
Пристрій виводу текстової інформації Принтер HP Laser	0,12	2,4	3,62
Оргтехніка Samsung Office	0,52	2,1	13,72
Мережеве обладнання передачі цифрових даних	0,10	240,0	301,44
Серверне обладнання на основі ПК ZEVS PC 13430U i5 9400F + GTX 1060 3GB	0,32	240,0	964,61
Всього			2489,14

4.3.9 Службові відрядження

Витрати за статтею «Службові відрядження» розраховуємо як 20…25% від суми основної заробітної плати дослідників та робітників за формулою:

$$В_{св} = (З_о + З_р) \cdot \frac{Н_{св}}{100\%}, \quad (4.13)$$

де $Н_{св}$ – норма нарахування за статтею «Службові відрядження», приймемо $Н_{св}$ = 20%.

$В_{св} = (70725,00 + 5387,75) \cdot 20 / 100\% = 15222,55$ *(грн).*

4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати розраховуємо як 30…45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{сп} = (З_о + З_р) \cdot \frac{Н_{сп}}{100\%}, \quad (4.14)$$

де $Н_{сп}$ – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», приймемо $Н_{сп}$= 30%.

$B_{сп}$ = (70725,00 + 5387,75) · 30 / 100% = 22833,83 (грн).

4.3.11 Інші витрати

Витрати за статтею «Інші витрати» розраховуємо як 50…100% від суми основної заробітної плати дослідників та робітників за формулою:

$$І_в = (З_о + З_р) \cdot \frac{Н_{ів}}{100\%}, \quad (4.15)$$

де *$Н_{ів}$* – норма нарахування за статтею «Інші витрати», приймемо $Н_{ів}$ = 50%.

$І_в$ = (70725,00 + 5387,75) · 50 / 100% = 38056,38 (грн).

4.3.12 Накладні (загальновиробничі) витрати

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100…150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (З_о + З_р) \cdot \frac{Н_{нзв}}{100\%}, \quad (4.16)$$

де *$Н_{нзв}$* – норма нарахування за статтею «Накладні (загальновиробничі) витрати», приймемо $Н_{нзв}$ = 100%.

$B_{нзв}$ = (70725,00 + 5387,75) · 100 / 100% = 76112,75 (грн).

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = З_о + З_р + З_{дод} + З_н + М + К_в + В_{спец} + В_{прг} + А_{обл} + В_е + В_{св} + В_{сп} + І_в + В_{нзв}. \quad (4.17)$$

$B_{заг}$= 70725,00 +5387,75 +7611,28 +18419,29 +5791,80 +14584,50 +52077,90 + 33303,90 + 9756,13 +2489,14 +15222,55 +22833,83 +38056,38 +76112,75 = 372372,18 (грн).

Загальні витрати *ЗВ* на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{B_{заг}}{\eta}, \quad (4.18)$$

де η - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, приймемо η=0,9.

ЗВ = 372372,18 / 0,9 = 413746,87 (грн).

4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

Результати дослідження проведені за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

В цьому випадку основу майбутнього економічного ефекту будуть формувати:

ΔN – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	500	900	1200	500

N – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, приймемо 6000 осіб;

$Ц_б$ – вартість послуги у році до впровадження інформаційної системи, приймемо 8199,00 (грн);

$\pm\Delta Ц_о$ – зміна вартості послуги від впровадження результатів, приймемо 663,40 (грн).

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta П_i$ для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [27]:

$$\Delta П_i = (\pm\Delta Ц_о \cdot N + Ц_о \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot (1 - \frac{\vartheta}{100}), \quad (4.19)$$

де λ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2025 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність інноваційного продукту). Приймемо $\rho = 38\%$;

ϑ – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2025 році $\vartheta = 18\%$;

Збільшення чистого прибутку 1-го року:

$\Delta П_1 = (663,40 \cdot 6000,00 + 8862,40 \cdot 500) \cdot 0,83 \cdot 0,38 \cdot (1\text{-}0,18/100\%) = 2175470,24$ (грн).

Збільшення чистого прибутку 2-го року:

$\Delta П_2 = (663,40 \cdot 6000,00 + 8862,40 \cdot 1400) \cdot 0,83 \cdot 0,38 \cdot (1\text{-}0,18/100\%) = 4238327,85$ (грн).

Збільшення чистого прибутку 3-го року:

$\Delta П_3 = (663,40 \cdot 6000,00 + 8862,40 \cdot 2600) \cdot 0,83 \cdot 0,38 \cdot (1\text{-}0,18/100\%) = 6988804,67$ (грн).

Збільшення чистого прибутку 4-го року:

$\Delta П_4$ =(663,40·6000,00+8862,40·3100)·0,83·0,38·(1-0,18/100%)=8134836,67 (грн).

Приведена вартість збільшення всіх чистих прибутків *ПП*, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^{T} \frac{\Delta П_i}{(1+\tau)^t}, \quad (4.20)$$

де $\Delta П_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, τ=0,1;

t – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$ПП$ =2175470,24/(1+0,1)^1+4238327,85/(1+0,1)^2+6988804,67/(1+0,1)^3+

+8134836,67/(1+0,1)^4=1977700,22+3502750,29+5250792,39+5556202,90=

=16287445,80 (грн).

Величина початкових інвестицій PV, які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{інв} \cdot ЗВ, \quad (4.21)$$

де $k_{інв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо $k_{інв}$=2;

ЗВ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 413746,87 (грн).

$PV = k_{інв} \cdot ЗВ$= 2 · 413746,87 = 827493,74 (грн).

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV \quad (4.22)$$

де $ПП$ – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 16287445,80 (грн).

PV – теперішня вартість початкових інвестицій, 827493,74 (грн).

$E_{абс} = ПП - PV$ = 16287445,80 - 827493,74 = 15459952,06 (грн).

Внутрішня економічна дохідність інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.23)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, 15459952,06 грн;

PV – теперішня вартість початкових інвестицій, 827493,74 (грн);

$T_ж$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримування позитивних результатів від її впровадження, 4 роки.

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1 = (1+15459952,06/827493,74)^{1/4} = 1,11.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$:

$$\tau_{мін} = d + f, \quad (4.24)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2025 році в Україні d =0,1;

f – показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,25.

$\tau_{мін}$= 0,1+0,25 = 0,35 < 1,11 свідчить про те, що внутрішня економічна дохідність інвестицій $E_в$, вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» доцільно.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_в}, \tag{4.25}$$

де $E_в$ – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 1,11 = 0,90 \text{ р.}$$

$T_{ок}$ < 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

4. 5 Висновки

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями» становить 41,3 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,20 рази.

Також термін окупності становить 0,90 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями».

ВИСНОВКИ

Магістерська кваліфікаційна робота присвячена розробці інформаційній технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями.

У першому розділі проведено детальний аналіз предметної області та сучасних інформаційних технологій, виявлено, що автоматизація збору та обробки великих обсягів даних, а також застосування сучасних методів машинного навчання дозволяють виявляти закономірності й тенденції, що сприяють оптимізації сервісу та підвищенню лояльності пасажирів. Аналіз актуальності показав, що українська авіаційна галузь, незважаючи на складні умови воєнного часу та закриття повітряного простору, продовжує функціонувати та інтегруватися у європейський ринок, що підкреслює необхідність впровадження сучасних технологій передбачення задоволеності пасажирів. На основі проведеного огляду готових рішень та існуючих методів нейронних мереж обґрунтовано вибір інструментів (Python, Kaggle, багатошарові нейронні мережі) і сучасних алгоритмів машинного навчання (XGBoost, Gaussian Process Classifier, Ridge Classifier) для розробки моделі передбачення.

У другому розділі виконано підготовку, обробку та аналіз даних із датасету «Airline Passenger Satisfaction». Проведено очистку даних, кодування змінних та масштабування ознак, що забезпечило коректність навчання моделей. Розвідувальний аналіз виявив ключові фактори впливу на задоволеність пасажирів, зокрема вік, стать, клас обслуговування, тип подорожі та відстань перельоту. Розроблена архітектура інформаційної технології та UML-діаграми класів і об’єктів продемонстрували логіку послідовного проходження даних через усі етапи технології – від завантаження та обробки до передбачення та візуалізації результатів. Це підтвердило узгодженість між логічним проектуванням та практичною реалізацією моделі, зокрема п’ятишарової нейронної мережі.

У третьому розділі реалізовано та протестовано послідовний алгоритм побудови моделей передбачення задоволеності пасажирів. Створено та порівняно кілька нейронних мереж різної глибини (3, 4 та 5 шарів), а також моделі XGBoost,

Gaussian Process Classifier та Ridge Classifier. Найвищу точність показала 5-шарова нейронна мережа з точністю 0,97 на тестовому наборі даних, що підтвердило ефективність навчання та здатність моделі до узагальнення. Варто відзначити, що отримана точність перевищує результати передбачення, досягнуті у бакалаврській дипломній роботі, що свідчить про покращення методології та ефективності застосованих алгоритмів. Метод SHAP дозволив визначити найбільш значущі фактори, які впливають на прогноз задоволеності — Inflight wifi service, Type of Travel та Customer Type, що сприяє кращій інтерпретації моделей.

Четвертий розділ присвячено економічній оцінці розробки. Рівень комерційного потенціалу інформаційної технології становить 41,3 бала, що свідчить про її високий комерційний потенціал. Узагальнений коефіцієнт якості розробки показав, що науково-технічне рішення переважає існуючі аналоги приблизно у 2,2 рази. Термін окупності складає 0,9 року, що підтверджує доцільність впровадження розробки та її комерційну привабливість.

Отже, проведене дослідження підтверджує ефективність та практичну значимість розробленої інформаційної технології аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями. Результати роботи створюють надійну основу для оптимізації сервісу авіаперевізників, підвищення лояльності пасажирів і комерційної реалізації технології.

За результатами даного дослідження була опубліковано тези на «LV Всеукраїнську науково-технічну конференцію підрозділів Вінницького національного технічного університету (ВНТКП ВНТУ)».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Судець А.О., Крижановський Є.М., Штельмах І.М., «Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями», «LV Всеукраїнська науково-технічна конференція підрозділів Вінницького національного технічного університету (ВНТКП ВНТУ)» (Вінниця, 2025-2026 рр.). URL: https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view/26305/21672. (дата звернення: 12.11.2025).

2. McKinsey & Company. The future of airline business models and competitive advantage. URL: https://www.mckinsey.com. (дата звернення: 13.11.2025).

3. B. H. Hayadi, J. M. Kim, K. Hulliyah, et al., "Predicting Airline Passenger Satisfaction with Classification Algorithms," International Journal of Informatics and Information Systems, vol. 4, no. 1, pp. 82-94, 2021. URL: https://doi.org/10.47738/ijiis.v4i1.80. (дата звернення: 13.11.2025).

4. Як виживає українська цивільна авіація під час війни. URL: https://thepage.ua/ua/economy/yak-vizhivaye-ukrayinska-civilna-aviaciya-pid-chas-povnomasshtabnoyi-vijni. (дата звернення: 14.11.2025).

5. Чи можливо відновити цивільні польоти в Україні під час війни. URL: https://tsn.ua/ukrayina/chy-mozlyvo-vidnovyty-tsyvilni-polyoty-v-ukrayini-pid-chas-viyny-ekspert-vidpoviv-2891582.html. (дата звернення: 15.11.2025).

6. Офіційний сайт Міжнародної асоціації повітряного транспорту (IATA). URL: https://www.iata.org. (дата звернення: 15.11.2025).

7. Kaggle. URL: https://www.kaggle.com/. (дата звернення: 16.11.2025).

8. Що таке Kaggle та чи варто витрачати на нього час? URL: https://senior.ua/articles/scho-take-kaggle-ta-chi-varto-vitrachati-na-nogo-chas. (дата звернення: 16.11.2025).

9. Python Introduction. URL: https://www.w3schools.com/python/python_intro.asp. (дата звернення: 16.11.2025).

10. Multilayer Perceptrons in Machine Learning: A Comprehensive Guide. URL: https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning. (дата звернення: 17.11.2025).

11. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – URL: https://docs.vntu.edu.ua/card.php?id=8163. (дата звернення: 17.11.2025).

12. A Comprehensive Guide to the Gaussian Process Classifier in Python. URL: https://www.dataspoof.info/post/gaussian-process-classifier-in-python/. (дата звернення: 18.11.2025).

13. Ridge Classifier. URL: https://www.geeksforgeeks.org/python/ridge-classifier/. (дата звернення: 18.11.2025).

14. Swayam Patil, Kaggle Notebook «ANN From Scratch using Numpy | 81% | EDA» Updated 1 year ago. URL: https://www.kaggle.com/code/swish9/ann-from-scratch-using-numpy-81-eda#notebook-container. (дата звернення: 19.11.2025).

15. Seyedali Rafazi, Kaggle Notebook «93% Accuracy with Neural Network» Updated 8 months ago. URL: https://www.kaggle.com/code/alirafazi/93-accuracy-with-neural-network#Step-8-%7C-Build-Model. (дата звернення: 20.11.2025).

16. Sajjad Hadi, Kaggle Notebook «Airline Satisfaction Prediction w/ PyTorch» Updated 8 months ago. URL: https://www.kaggle.com/code/sajjadhadi/airline-satisfaction-prediction-w-pytorch#Model. (дата звернення: 21.11.2025).

17. Niladri, Kaggle Notebook «Airline Satisfaction» Updated 1 year ago. URL: https://www.kaggle.com/code/niladri54/airline-satisfaction. (дата звернення: 22.11.2025).

18. Eyad Wael, Kaggle Notebook «Airline satisfaction» Updated 2 years ago. URL: https://www.kaggle.com/code/eyadwael/airline-satisfaction#Modeling. (дата звернення: 23.11.2025).

19. What is exploratory data analysis (EDA). URL: https://www.ibm.com/think/topics/exploratory-data-analysis. (дата звернення: 23.11.2025).

20. Блок-схеми: визначення, призначення і приклади використання в практиці. URL: https://liderua.com/blok-shemy-vyznachennya-pryznachennya-i-pryklady-vykorystannya-v-praktytsi/. (дата звернення: 26.11.2025).

21. What is Deployment Diagram? URL: https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-deployment-diagram/. (дата звернення: 26.11.2025).

22. Use Case Diagram – Unified Modeling Language (UML). URL: https://www.geeksforgeeks.org/system-design/use-case-diagram/. (дата звернення: 26.11.2025).

23. UML Class Diagram. URL: https://www.geeksforgeeks.org/system-design/unified-modeling-language-uml-class-diagrams/. (дата звернення: 27.11.2025).

24. What is Object Diagram? URL: https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-object-diagram/. (дата звернення: 28.11.2025).

25. Welcome to the SHAP documentation. URL: https://shap.readthedocs.io/en/latest/index.html. (дата звернення: 29.11.2025).

26. Anna Sudets, Kaggle Notebook «Satisfaction Prediction» Updated 1 day ago. URL: https://www.kaggle.com/code/annasudets/satisfaction-prediction. (дата звернення: 03.12.2025).

27. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

28. Кавецький В. В. Економічне обґрунтування інноваційних рішень: практикум / В. В. Кавецький, В. О. Козловський, І. В. Причепа – Вінниця : ВНТУ, 2016. – 113 с.

Додаток А

Технічне завдання

Міністерство освіти і науки України
Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ
Завідувач кафедри САІТ
____ д.т.н., проф. Віталій МОКІН
«__» _____2025 року

ТЕХНІЧНЕ ЗАВДАННЯ
на магістерську кваліфікаційну роботу
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПЕРЕДБАЧЕННЯ РІВНІВ ЗАДОВОЛЕНОСТІ ПАСАЖИРІВ АВІАКОМПАНІЯМИ
08-34.МКР.013.00.000 ТЗ

Керівник: к.т.н., асистент
_____ Ігор ШТЕЛЬМАХ
«__» _____ 2025 р.
Розробила студентка гр. 2ІСТ-24м
_____Анна СУДЕЦЬ
«__» _____ 2025 р.

Вінниця 2025

1. Підстава для проведення робіт.
Підставою для виконання роботи є наказ №__ по ВНТУ від «__» _____2025р., та індивідуальне завдання на МКР, затверджене протоколом №__ засідання кафедри САІТ від «__» _____ 2025р.
2. Джерела розробки:
1) B. H. Hayadi, J. M. Kim, K. Hulliyah, et al., "Predicting Airline Passenger Satisfaction with Classification Algorithms," International Journal of Informatics and Information Systems, vol. 4, no. 1, pp. 82-94, 2021. URL: https://doi.org/10.47738/ijiis.v4i1.80;
2) Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс] / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу: https://docs.vntu.edu.ua/card.php?id=8163.
3. Мета і призначення роботи:
Метою дослідження є підвищення точності передбачення рівнів задоволеності пасажирів авіакомпаніями.
4. Вихідні дані для проведення робіт:
Kaggle Dataset «Airline Passenger Satisfaction» https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction.
5. Методи дослідження:
У даній роботі використано методи машинного навчання, а саме алгоритми XGBoost, Gaussian Process Classifier, Ridge Classifier та багатошарові нейронні мережі різної глибини. Для інтерпретації результатів застосовано метод SHAP.
6. Етапи роботи і терміни їх виконання:
а) Характеристика об’єкту досліджень _____ – _____
б) Вибір оптимальних інформаційних технологій _____ – _____
в) Розвідувальний аналіз даних _____ – _____
г) Розроблення інформаційної технології _____ – _____
д) Економічна частина _____ – _____
е) Оформлення матеріалів до захисту МКР _____ – _____
7. Очікувані результати та порядок реалізації:
Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями.
8. Вимоги до розробленої документації:
Текстова та ілюстративна частини роботи оформлені у відповідності до вимог «Методичних вказівок до виконання магістерських кваліфікаційних робіт для студентів спеціальності 126 «Інформаційні системи та технології» (освітня програма «Інформаційні технології аналізу даних та зображень»).
9. Порядок приймання роботи:
Публічний захист «___» _____ 2025 р.
Початок розробки «___» _____ 2025 р.
Граничні терміни виконання МКР «___» _____ 2025 р.

Розробила студентка групи 2ІСТ-24м _____ Анна СУДЕЦЬ

Додаток Б

ПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Назва роботи: « Інформаційна технологія аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ, ФІІТА, гр. 2ІСТ-24м

Коефіцієнт подібності текстових запозичень, виявлених у роботі системою StrikePlagiarism 2,15 %

Висновок щодо перевірки кваліфікаційної роботи (відмітити потрібне):

■ Запозичення, виявлені у роботі, є законними і не містять ознак плагіату, фабрикації, фальсифікації. Роботу прийняти до захисту

□ У роботі не виявлено ознак плагіату, фабрикації, фальсифікації, але надмірна кількість текстових з апозичень та/або наявність типових розрахунків не дозволяють прийняти рішення про оригінальність та самостійність її виконання. Роботу направити на доопрацювання.

□ У роботі виявлено ознаки плагіату та/або текстових маніпуляцій як спроб укриття плагіату, фабрикації, фальсифікації, що суперечить вимогам законодавства та нормам академічної доброчесності. Робота до захисту не приймається.

Експертна комісія:

Віталій МОКІН, зав. каф. САІТ _____
(підпис)

Сергій ЖУКОВ, доц. каф. САІТ _____
(підпис)

Особа, відповідальна за перевірку _____ Сергій ЖУКОВ
(підпис)

З висновком експертної комісії ознайомлена

Керівник _____ Ігор ШТЕЛЬМАХ, к.т.н., асистент каф. САІТ
(підпис)

Здобувач _____ Анна СУДЕЦЬ
(підпис)

Додаток В
Лістинг програми

```
# Ignore warnings
import warnings
from numba.core.errors import NumbaDeprecationWarning

warnings.filterwarnings("ignore", category=NumbaDeprecationWarning)
warnings.filterwarnings("ignore", message="unable to load libtensorflow_io_plugins.so")
warnings.filterwarnings("ignore", message="file system plugins are not loaded")

# Work with Data - the main Python libraries
import os
import numpy as np
import pandas as pd

# Data Profiling (optional)
import ydata profiling as pp    # pandas profiling deprecated

# Visualization
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
import seaborn as sns
import plotly.express as px

# Preprocessing
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.model selection import (
        train test split, KFold, ShuffleSplit,
        GridSearchCV, RandomizedSearchCV,
        cross val score, learning curve
)

# Modeling - sklearn classifiers and regressors
from sklearn.neural_network import MLPClassifier, MLPRegressor
from sklearn.ensemble import (
        RandomForestClassifier, RandomForestRegressor,
        ExtraTreesClassifier, AdaBoostClassifier,
        GradientBoostingClassifier
)
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression

# Metrics
from sklearn.metrics import (
        accuracy_score, confusion_matrix,
        classification report, r2 score,
        mean absolute error, mean squared error as mse
)

# Visualization Tools
from yellowbrick.classifier import ConfusionMatrix

# Saving models
```

```
import joblib

# Deep Learning with Keras
import keras
from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.callbacks import ReduceLROnPlateau
from keras import optimizers
from keras.wrappers.scikit_learn import KerasRegressor
directory = "/kaggle/input/airline-passenger-satisfaction/"
feature_tables = ['train.csv', 'test.csv']

df_train_path = directory + feature_tables[0]
df_test_path = directory + feature_tables[1]

print(f'Reading csv from {df_train_path}...')
train = pd.read_csv(df_train_path)
print('...Complete')

print(f'Reading csv from {df_test_path}...')
test = pd.read_csv(df_test_path)
print('...Complete')
train.head(10)
train.info()
train.describe()
train.dtypes
corr = train.corr(numeric_only=True).round(2)
plt.figure(figsize=(25, 20))
sns.heatmap(corr, annot=True, cmap='YlOrBr')
plt.show()
plt.figure(figsize=(20, 10))
gs = gridspec.GridSpec(2, 2, height_ratios=[1, 1])

plt.subplot(gs[0, 0])
plt.gca().set_title('Variable Gender')
sns.countplot(x='Gender', palette='Set2', data=train)

plt.subplot(gs[0, 1])
plt.gca().set_title('Variable Customer Type')
sns.countplot(x='Customer Type', palette='Set2', data=train)

plt.subplot(gs[1, :])
plt.gca().set_title('Variable Age')
sns.countplot(x='Age', palette='Set2', data=train)

plt.show()
plt.figure(figsize=(20, 15))
gs = gridspec.GridSpec(3, 2, height_ratios=[1, 1, 1])

plt.subplot(gs[0, 0])
plt.gca().set_title('Variable Type of Travel')
sns.countplot(x='Type of Travel', palette='Set2', data=train)

plt.subplot(gs[0, 1])
plt.gca().set_title('Variable Class')
sns.countplot(x='Class', palette='Set2', data=train)

plt.subplot(gs[1, :])
plt.gca().set_title('Variable Flight Distance')
```

```
sns.countplot(x='Flight Distance', palette='Set2', data=train)

plt.show()

min_flight_distance = np.min(train['Flight Distance'])
mean_flight_distance = np.mean(train['Flight Distance'])
max_flight_distance = np.max(train['Flight Distance'])

print(f"Min Flight Distance: {min_flight_distance}")
print(f"Mean Flight Distance: {mean_flight_distance:.2f}")
print(f"Max Flight Distance: {max_flight_distance}")
train['Departure Delay in Minutes'].fillna(0, inplace=True)
train['Arrival Delay in Minutes'].fillna(0, inplace=True)

train['Departure Delay in Minutes'] = train['Departure Delay in Minutes'].astype(int)
train['Arrival Delay in Minutes'] = train['Arrival Delay in Minutes'].astype(int)

plt.figure(figsize=(20, 15))
gs = gridspec.GridSpec(2, 1, height_ratios=[1, 1])

plt.subplot(gs[0, :])
plt.gca().set_title('Variable Departure Delay in Minutes')
sns.countplot(x='Departure Delay in Minutes', palette='Set2', data=train)
plt.gca().set_xlim([0, 60])

plt.subplot(gs[1, :])
plt.gca().set_title('Variable Arrival Delay in Minutes')
sns.countplot(x='Arrival Delay in Minutes', palette='Set2', data=train)
plt.gca().set_xlim([0, 60])

plt.show()
le = LabelEncoder()
y_train = le.fit_transform(y_train)
y_test = le.transform(y_test)
X_train_df = pd.DataFrame(X_train)
X_test_df = pd.DataFrame(X_test)

mask_train = ~X_train_df.isnull().any(axis=1)
X_train = X_train_df[mask_train].values
y_train = y_train[mask_train]

mask_test = ~X_test_df.isnull().any(axis=1)
X_test = X_test_df[mask_test].values
y_test = y_test[mask_test]
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
model = Sequential()
model.add(Dense(64, input_dim=X_train.shape[1], activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

history = model.fit(X_train, y_train, epochs=10, batch_size=32, validation_split=0.2)
loss_before, keras_accuracy_before = model.evaluate(X_test, y_test)
print("Keras Accuracy (NN with 3 layers:", keras_accuracy_before)
y_train_pred_simple = (model.predict(X_train) > 0.5).astype(int)
y_test_pred_simple = (model.predict(X_test) > 0.5).astype(int)
```

```
train_acc_simple = accuracy_score(y_train, y_train_pred_simple)
test acc simple = accuracy score(y test, y test pred simple)

print(f"Train Accuracy (NN with 3 layers): {train_acc_simple}")
print(f"Test Accuracy (NN with 3 layers): {test acc simple}")
cm_matrix = confusion_matrix(y_test, y_test_pred_simple)
plt.figure(figsize=(8, 6))
sns.heatmap(cm matrix, annot=True, cmap='Purples', fmt='g', cbar=False)
tick labels = ['neutral or dissatisfied', 'satisfied']
plt.xticks(ticks=[0.5, 1.5], labels=tick_labels)
plt.yticks(ticks=[0.5, 1.5], labels=tick_labels)
plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.title('Confusion Matrix for NN with 3 layers')
plt.show()
plt.figure(figsize=(8, 5))
plt.plot(history.history['accuracy'], label='Train Accuracy')
plt.plot(history.history['val accuracy'], label='Validation Accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.title('Train vs Validation Accuracy for NN with 3 layers')
plt.legend(loc='lower right')
plt.grid(True)
plt.show()
print(classification_report(y_test, y_test_pred_simple, target_names=['neutral or dissatisfied', 'satisfied']))
model1 = Sequential()
model1.add(Dense(128, input dim=X train.shape[1], activation='relu'))
model1.add(Dropout(0.3))
model1.add(Dense(64, activation='relu'))
model1.add(Dense(32, activation='relu'))
model1.add(Dense(16, activation='relu'))
model1.add(Dense(1, activation='sigmoid'))

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=5, verbose=1)

model1.compile(loss='binary crossentropy', optimizer='adam', metrics=['accuracy'])

history1 = model1.fit(
        X_train, y_train,
        epochs=30,
        batch_size=64,
        validation_split=0.2,
        callbacks=[reduce lr],
        verbose=1
)

loss2, accuracy1 = model1.evaluate(X_test, y_test)
print("Keras Accuracy (NN with 5 layers):", accuracy1)
y train pred dropout = (model1.predict(X train) > 0.5).astype(int)
y test pred dropout = (model1.predict(X test) > 0.5).astype(int)

train acc dropout = accuracy score(y train, y train pred dropout)
test acc dropout = accuracy score(y test, y test pred dropout)

print(f"Train Accuracy (NN with 5 layers): {train acc dropout}")
print(f"Test Accuracy (NN with 5 layers): {test_acc_dropout}")
cm_matrix = confusion_matrix(y_test, y_test_pred_dropout)
```

```
plt.figure(figsize=(8, 6))
sns.heatmap(cm_matrix, annot=True, cmap='Purples', fmt='g', cbar=False)

tick_labels = ['neutral or dissatisfied', 'satisfied']
plt.xticks(ticks=[0.5, 1.5], labels=tick_labels)
plt.yticks(ticks=[0.5, 1.5], labels=tick_labels)

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.title('Confusion Matrix for NN with 5 layers')
plt.show()
plt.figure(figsize=(8, 5))
plt.plot(history1.history['accuracy'], label='Train Accuracy')
plt.plot(history1.history['val_accuracy'], label='Validation Accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.title('Train vs Validation Accuracy for NN with 5 layers')
plt.legend(loc='lower right')
plt.grid(True)
plt.show()
print(classification_report(y_test, y_test_pred_dropout, target_names=['neutral or dissatisfied', 'satisfied']))
model2 = Sequential()
model2.add(Dense(96, input_dim=X_train.shape[1], activation='relu'))
model2.add(Dropout(0.2))
model2.add(Dense(64, activation='relu'))
model2.add(Dense(32, activation='relu'))
model2.add(Dense(1, activation='sigmoid'))

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=3, verbose=1)

model2.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

history2 = model2.fit(
    X_train, y_train,
    epochs=20,
    batch_size=48,
    validation_split=0.2,
    callbacks=[reduce_lr],
    verbose=1
)

loss2, accuracy2 = model2.evaluate(X_test, y_test)
print("Keras Accuracy (NN with 4 layers):", accuracy2)
y_train_pred_medium = (model2.predict(X_train) > 0.5).astype(int)
y_test_pred_medium = (model2.predict(X_test) > 0.5).astype(int)

train_acc_medium = accuracy_score(y_train, y_train_pred_medium)
test_acc_medium = accuracy_score(y_test, y_test_pred_medium)

print(f"Train Accuracy (NN with 4 layers): {train_acc_medium}")
print(f"Test Accuracy (NN with 4 layers): {test_acc_medium}")
cm_matrix = confusion_matrix(y_test, y_test_pred_medium)
```

Додаток Г

(обов’язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПЕРЕДБАЧЕННЯ РІВНІВ ЗАДОВОЛЕНОСТІ ПАСАЖИРІВ АВІАКОМПАНІЯМИ

Нормоконтроль: к.т.н., доцент

_____ Сергій ЖУКОВ

«___» _____ 2025 р.

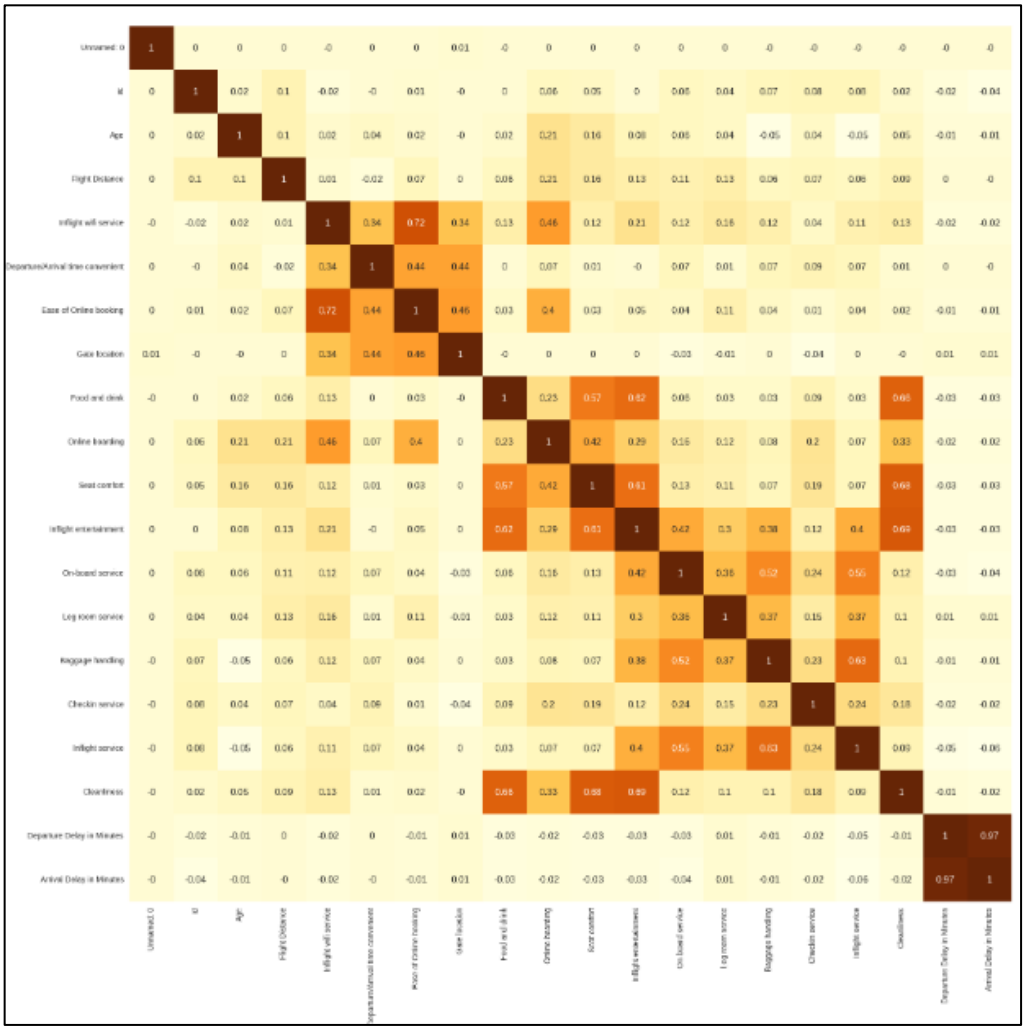


Рисунок Г.1 – Кореляційна матриця

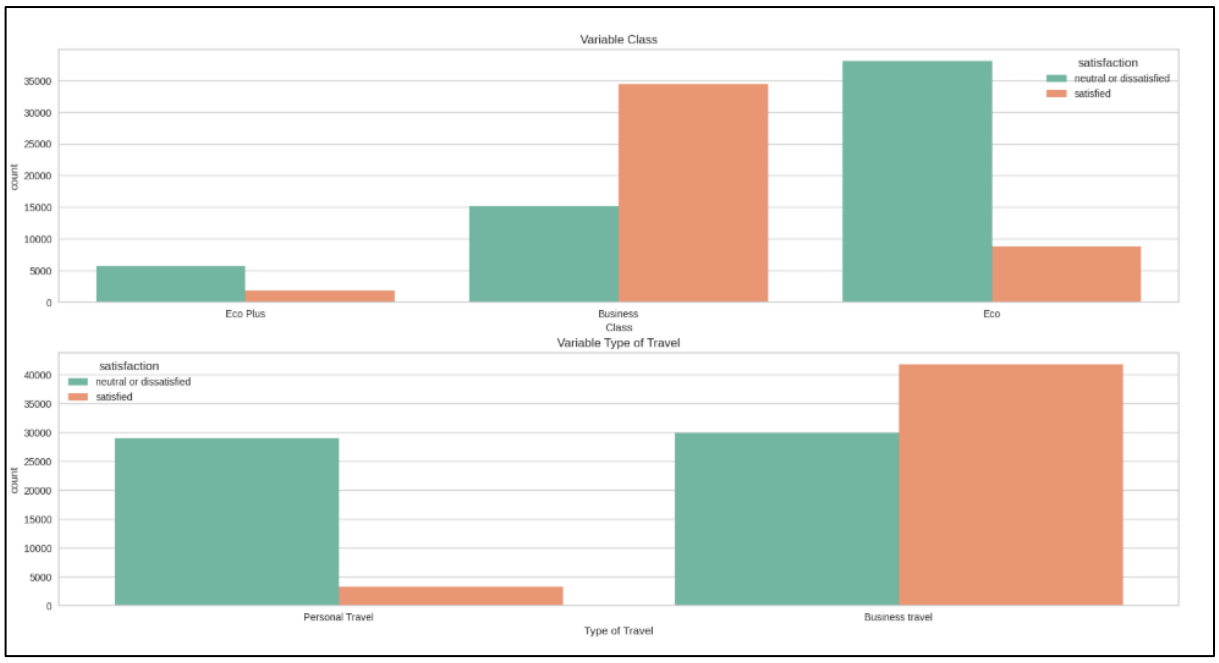


Рисунок Г.2 – Вплив класу та типу подорожі на задоволеність

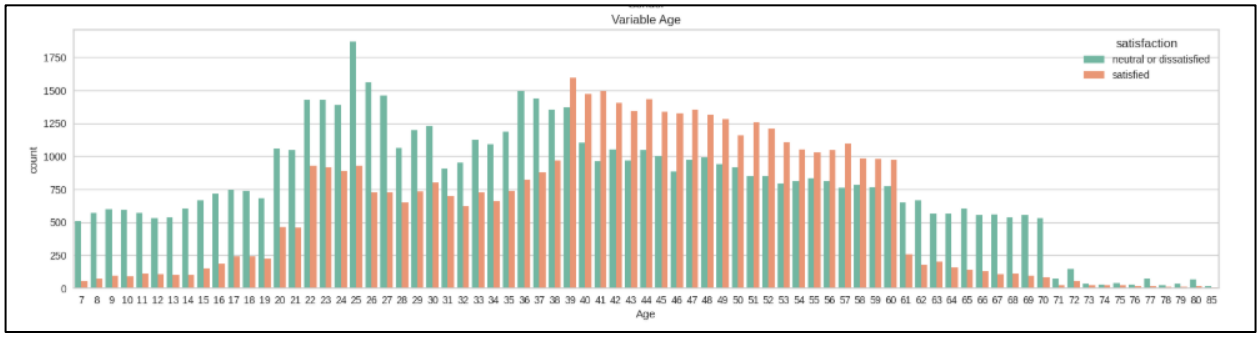


Рисунок Г.3 – Вплив віку на рівень задоволеності пасажирів

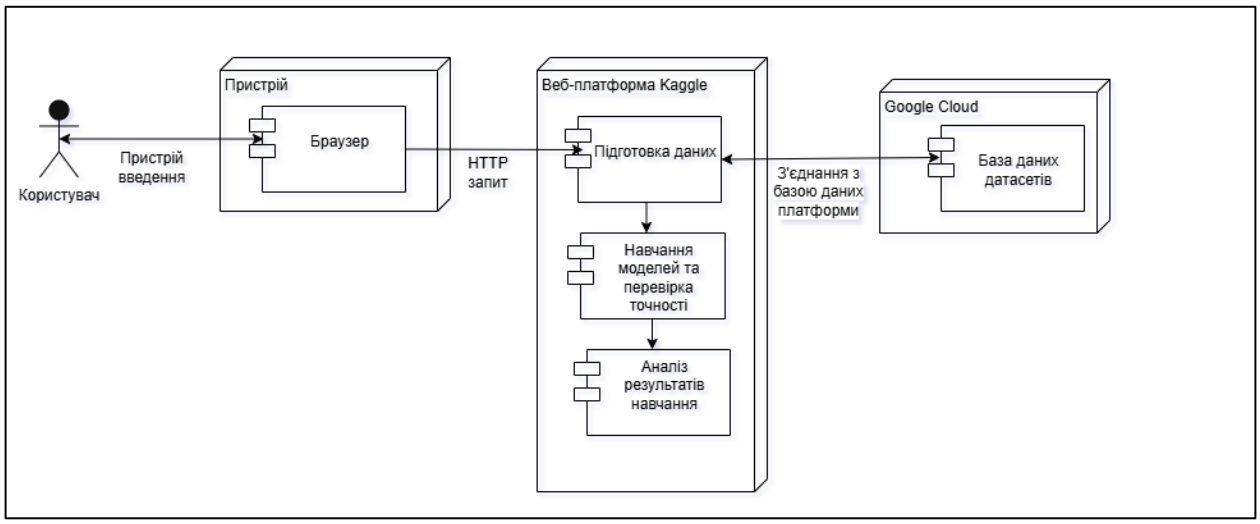


Рисунок Г.4 – Діаграма розгортання

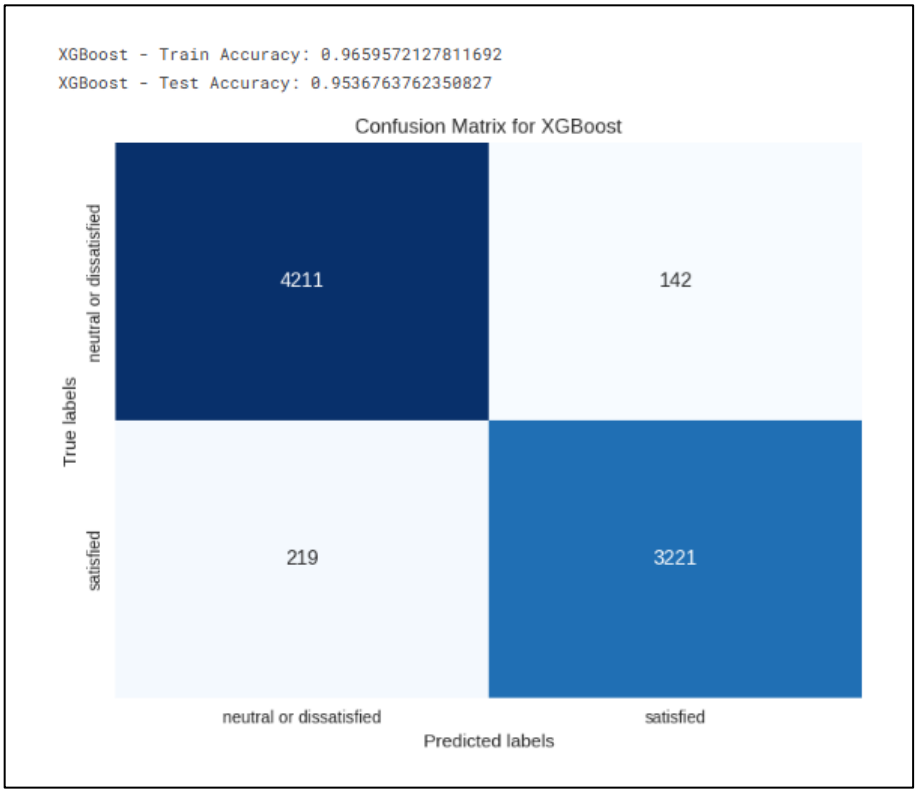


Рисунок Г.5 – Матриця плутанини моделі XGBoost

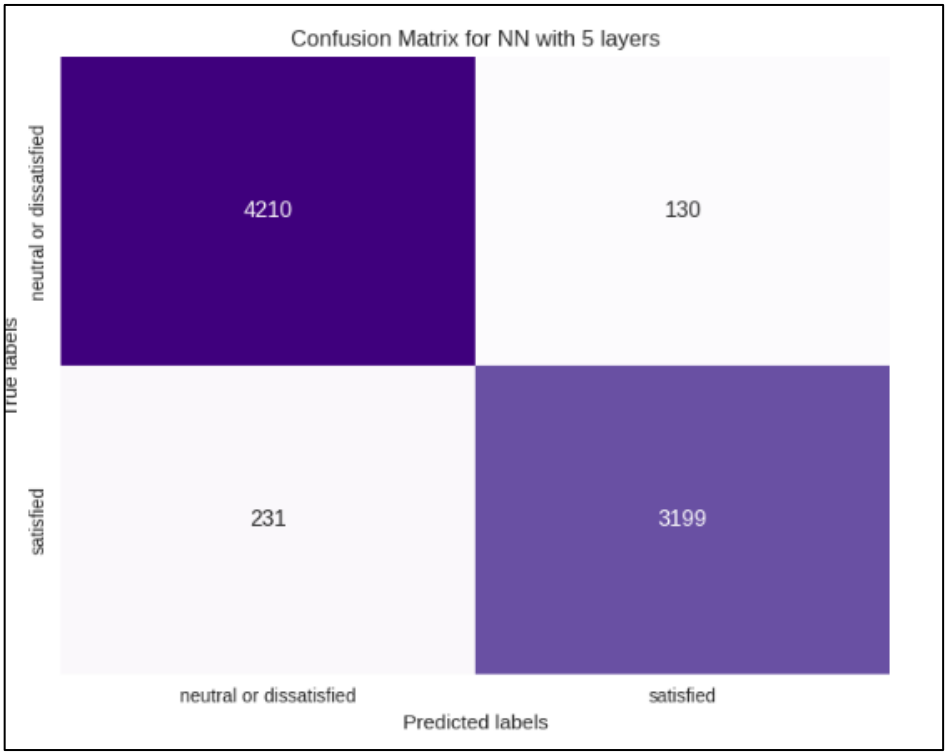


Рисунок Г.6 – Матриця плутанини 5-шарової нейронної мережі