

ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
(повне найменування вищого навчального закладу)
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ
(повне найменування інституту, назва факультету (відділення))
КАФЕДРА ПРОГРАМНИХ ЗАСОБІВ ТА ТЕХНОЛОГІЙ
(повна назва кафедри (предметної, циклової комісії))

Пояснювальна записка

до кваліфікаційної роботи
бакалавра
(освітній рівень)

на тему: «Розробка програмного додатку для дослідження рівня
задоволеності клієнтів»

Виконав: здобувачка освіти 4 року навчання,
групи 4ПРЗ спеціальності
121 - «Інженерія програмного забезпечення»
(шифр і назва спеціальності)

Ястребова Оксана Юріївна
(прізвище та ініціали)

Керівник д.т.н., проф. Шерстюк В.Г.
(прізвище та ініціали)

Рецензент к.т.н. доцент Вишемирська С.В.
(прізвище та ініціали)

Хмельницький - 2024

Херсонський національний технічний університет

(повне найменування вищого навчального закладу)

Факультет, відділення	<u>Інформаційних технологій та дизайну</u>
Кафедра	<u>Програмних засобів і технологій</u>
Освітній рівень	<u>бакалавр</u>
Спеціальність	<u>121 – Інженерія програмного забезпечення</u>

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри
Програмних засобів та технологій
к.т.н. доц. О.Є. Огнєва
 “ ” 2024 р.

З А В Д А Н Н Я

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ

Ястребовій Оксані Юріївні

(прізвище, ім'я, по батькові)

Тема роботи «Розробка програмного додатку для дослідження рівня задоволеності клієнтів»

керівник роботи д.т.н. професор Шерстюк В.Г.,
 (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджена наказом вищого навчального закладу від 09.02.2024 р. №219-С

2. Строк подання здобувачем роботи 15.06.2024 р.

3. Вихідні дані до роботи літературні та періодичні джерела, статистичні дані задоволеності пасажирів компанії-перевізника

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):

1) огляд методів прогнозування задоволеності пасажирів компанії-перевізника за допомогою методів машинного навчання;

2) дослідження основних алгоритмів машинного навчання;

3) оцінювання якості роботи класифікаторів та аналіз результатів прогнозування.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Структура набору даних.

2. Діаграми розподілу значень.

3. Кореляційна матриця ознак.

4. Статистичні характеристики ознак набору даних.

5. Основні алгоритми машинного навчання.

6. Аналіз результатів моделювання.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів виконання роботи	Термін виконання етапів роботи	Примітки
1.	Отримання завдання	09.02.2024	Виконано
2.	Підбір літератури	09.02.2024-21.02.2024	Виконано
3.	Аналіз предметної області	22.02.2024-28.02.2024	Виконано
4.	Розробка та обґрунтування завдання	01.03.2024-14.03.2024	Виконано
5.	Розробка концептуальної моделі	15.03.2024-21.03.2024	Виконано
6.	Розробка алгоритму	22.03.2024-04.04.2024	Виконано
7.	Проектування програмної системи	05.04.2024-11.04.2024	Виконано
8.	Розробка інтерфейсу програмної системи	12.04.2024-18.03.2024	Виконано
9.	Тестування програмної системи	19.04.2024-25.04.2024	Виконано
10.	Оформлення пояснювальної записки	26.04.2024-09.05.2024	Виконано
11.	Захист кваліфікаційної роботи	17.06.2024	Виконано

Здобувачка вищої освіти _____ Ястребова О.Ю.
(підпис) (прізвище та ініціали)Керівник роботи _____ Шерстюк В.Г.
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Кваліфікаційна робота бакалавра: 106 сторінок, 55 рисунків, 1 таблиця, 1 додаток, 40 джерел.

Мета роботи – підвищення рівня задоволеності пасажирів компанії-перевізника за рахунок використання методів та алгоритмів машинного навчання.

Об'єкт дослідження – пасажирів компаній-перевізників, представлені статистичними даними.

Предмет дослідження – методи обробки статистичних даних та моделі машинного навчання на основі статистичних даних для прогнозування задоволеності пасажирів компанією-перевізником.

Методи дослідження – математичні та статистичні методи машинного навчання, об'єктно-орієнтоване програмування.

Новизна роботи полягає в удосконаленні методів прогнозування задоволеності пасажирів компанії-перевізника.

Практична цінність результатів роботи полягає в тому, що методи та моделі, запропоновані при її виконанні, дозволяють компаніям-перевізникам передбачити задоволеність різних категорій пасажирів якістю наданих послуг, що дає можливість вдосконалити та оптимізувати послуги, а також вжити маркетингові заходи для втримання окремих груп пасажирів.

Ключові слова: *МАШИННЕ НАВЧАННЯ, ЛОГІСТИЧНА РЕГРЕСІЯ, ДИСКРИМІНАЦІЙНИЙ АНАЛІЗ, МЕТОД ОПОРНИХ ВЕКТОРІВ, ДЕРЕВА РІШЕНЬ, МЕТОД НАЙБЛИЖЧИХ СУСІДІВ, НАЇВНИЙ БАССІВ КЛАСИФІКАТОР, АНСАМБЛЕВІ МЕТОДИ МАШИННОГО НАВЧАННЯ, ШТУЧНІ НЕЙРОННІ МЕРЕЖІ, ACCURACY, PRECISION.*

АНОТАЦІЯ

Кваліфікаційна робота бакалавра складається зі вступу, трьох розділів, висновку, списку використаних джерел та додатку.

Роботу присвячено розробці програмної системи прогнозування задоволеності пасажирів компанії-перевізника на основі моделей машинного навчання та на прикладі гіпотетичної авіакомпанії, що здійснює авіарейси, для якої отримано відкритий реальний набір даних, накопичений в процесі дослідження діяльності компанії, який може бути використаний для аналізу.

У роботі розглянуто процес аналітичного прогнозування задоволеності пасажирів авіакомпанією на основі методів та алгоритмів машинного навчання, особливу увагу приділено підготовці даних та створенню моделей прогнозування.

Наведено огляд сучасних моделей машинного навчання та методику їх побудови, процес підготовки та аналіз даних щодо задоволеності пасажирів компанією, створення моделей на основі цих даних.

Запропоновано ефективну модель машинного навчання для побудови на її основі системи прийняття рішень.

Розроблене програмне рішення може бути використане компаніями-перевізниками для передбачення задоволеності пасажирів якістю наданих послуг, а також для аналізу сприйняття якості послуг різними категоріями пасажирів.

ABSTRACT

The bachelor's thesis consists of an introduction, three chapters, a conclusion, a list of sources and an appendix.

The work is devoted to the development of a software system for predicting the satisfaction of passengers of the carrier company based on machine learning models and on the example of a hypothetical airline operating flights, for which an open real data set accumulated in the process of researching the company's activities was obtained, which can be used for analysis.

The work examines the process of analytical forecasting of passenger satisfaction with the airline based on machine learning methods and algorithms, special attention is paid to data preparation and creation of forecasting models.

An overview of modern machine learning models and the methodology of their construction, the process of preparation and analysis of data on passenger satisfaction with the company, and the creation of models based on these data are given.

An effective model of machine learning is proposed for building a decision-making system based on it.

The developed software solution can be used by carrier companies to predict passenger satisfaction with the quality of services provided, as well as to analyze the perception of service quality by different categories of passengers.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	9
ВСТУП	10
РОЗДІЛ 1. ОГЛЯД МЕТОДІВ ПРОГНОЗУВАННЯ ЗАДОВОЛЕНOSTІ ПАСАЖИРІВ КОМПАНІЇ-ПЕРЕВІЗНИКА ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ	13
1.1. Застосування методів машинного навчання щодо прогнозування задоволеності споживачів.....	13
1.2. Огляд та підготовка даних задоволеності пасажирів авіакомпанією.....	14
1.2.1. Аналітичний цикл обробки даних щодо задоволеності пасажирів компанії.....	15
1.2.2. Основні характеристики даних задоволеності пасажирів авіакомпанією	20
1.2.3. Підготовка даних щодо задоволеності пасажирів авіакомпанією до моделювання	40
1.3. Постановка задачі дослідження.....	42
1.4. Висновки до розділу 1	42
РОЗДІЛ 2. ДОСЛІДЖЕННЯ ОСНОВНИХ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ.....	44
2.1. Постановка задачі класифікації	44
2.1.1. Логістична регресія.....	47
2.1.2. Дискримінантний аналіз.....	48
2.1.3. Метод опорних векторів	51
2.1.4. Дерева рішень.....	52

2.1.5. Метод найближчих сусідів	55
2.1.6. Наївний баєсів класифікатор.....	57
2.1.7. Ансамблеві методи.....	58
2.1.8. Штучні нейронні мережі	62
2.2. Висновки до розділу 2	64
РОЗДІЛ 3. ОЦІНЮВАННЯ ЯКОСТІ РОБОТИ КЛАСИФІКАТОРІВ ТА	
АНАЛІЗ РЕЗУЛЬТАТІВ ПРОГНОЗУВАННЯ	66
3.1. Способи оцінки якості моделі.....	66
3.2. Аналіз результатів моделювання	74
3.2.1. Модель Random Forest Classifier.....	77
3.2.2. Модель LGBMClassifier.....	82
3.2.3. Модель XGBClassifier	87
3.3. Висновки до розділу 3	91
ВИСНОВКИ.....	92
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	94
ДОДАТОК А. КОД ПРОГРАМИ.....	99

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ANNs - Artificial Neural Networks;

CNN - Convolutional Neural Networks;

CRM - Customer Relationship Management;

DT - Decision Tree;

GANs - Generative Adversarial Networks;

GNNs - Graph Neural Networks;

KNN - K-Nearest Neighbor;

LDA - Linear Discriminant Analysis;

LR - Linear Regression;

MLP - Multi-Layer Perceptrons;

QDA - Quadratic Discriminant Analysis;

RDA - Regularized Discriminant Analysis;

RNN - Recurrent Neural Networks;

SLP - Single-Layer Perceptrons;

SVM - Support Vector Machines.

ВСТУП

В сучасному світі, комерційні, та некомерційні організації, під час своєї діяльності збирають великі обсяги даних, кількість яких з кожним роком збільшується. Разом з цим, збільшується і потреба в аналізі цих даних, задля того, щоб вони приносили якомога більшу користь, з використанням результатів цього аналізу для прийняття рішень. Отримання нової інформації з наявних даних є задачею аналізу даних. У цій роботі увага зосереджена на аналітичному прогнозуванні, яке є важливим різновидом аналізу даних.

Аналітичне прогнозування являє собою процес створення та використання моделей, які передбачають події на основі наявних даних. Модель застосовується для прогнозування з метою надання інформаційної підтримки користувачу при прийнятті рішень. У повсякденному житті під прогнозом ми розуміємо передбачення того, що може або не може статися у майбутньому. У контексті аналітичного прогнозування термін "прогноз" означає присвоєння значення будь-якій невідомій змінній, як, наприклад, прогноз ціни товару, або вірогідності настання тієї чи іншої події. Прогностична модель навчається задля того, щоб створювати прогнози на підставі набору статистичних даних, а для навчання такої моделі, ми використовуємо методи машинного навчання.

Машинне навчання у практичному контексті являє собою автоматизований процес, який виконує функцію отримання шаблонів із даних. Для створення моделей в аналітичному прогнозуванні при виконанні задачі класифікації, використовуються методи машинного навчання з учителем, завдяки яким створені моделі, на основі зібраних даних, автоматично визначають взаємозв'язки між набором описових ознак та цільовою ознакою, та можуть бути використані для прогнозування.

Використання моделей машинного навчання є поширеною практикою у прогнозуванні поведінки споживачів, адже одержані завдяки розробленим інформаційно-аналітичним системам дані є основою для прийняття рішень з

вдосконалення та оптимізації діяльності організації задля досягнення поставлених цілей.

Метою дослідження є підвищення рівня задоволеності пасажирів компанії-перевізника за рахунок використання методів та алгоритмів машинного навчання.

Об'єктом дослідження є пасажирів компаній-перевізників, представлені статистичними даними.

Предметом дослідження є методи обробки статистичних даних та модель машинного навчання на основі статистичних даних для прогнозування задоволеності пасажирів компанією-перевізником.

Для вирішення поставлених задач використано математичні та статистичні методи машинного навчання, об'єктно-орієнтоване програмування.

Новизна результатів кваліфікаційної роботи полягає в удосконаленні методів прогнозування задоволеності пасажирів компанією-перевізником.

Практичне значення результатів кваліфікаційної роботи полягає в тому, що методи та моделі запропоновані в дослідженні, дозволяють компаніям передбачити задоволеність різних категорій пасажирів якістю наданих послуг, що дає можливість вдосконалити та оптимізувати послуги, а також вжити необхідних маркетингових заходів для утримання окремих груп пасажирів, наприклад, за допомогою спеціальних пропозицій та акцій.

Тематика першого розділу присвячена огляду питань застосування методів машинного навчання щодо прогнозування задоволеності споживачів, огляду та аналізу даних задоволеності пасажирів авіакомпанією, а також підготовці їх до моделювання.

У другому розділі розглянуті основні методи та алгоритми машинного навчання які можуть бути застосовані до виконання задачі класифікації.

У третьому розділі проведено моделювання з метою прогнозування задоволеності пасажирів авіакомпанією на основі статистичних даних, аналіз та вдосконалення моделей машинного навчання.