

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

Пояснювальна записка

до кваліфікаційної роботи

бакалавра

(освітньо-кваліфікаційний рівень)

на тему: «Розробка програмного рішення для прогнозування відтоку
клієнтів банку»

Виконав: здобувач 4 курсу, групи 4ПР5
напряму підготовки (спеціальності)

121 «Інженерія програмного забезпечення»

(шифр і назва напряму підготовки, спеціальності)

Петров Владислав Олександрович

(прізвище та ініціали)

Керівник ст. викладач Комісаров О.С.

(прізвище та ініціали)

Рецензент Воронеко Марія Олександрівна

(прізвище та ініціали)

Хмельницький – 2025

ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Інститут, факультет, відділення Інформаційних технологій та дизайну
Кафедра, циклова комісія Програмних засобів і технологій
Освітньо-кваліфікаційний рівень бакалавр
ОПП Програмування смарт-пристроїв та вбудованих систем
(шифр і назва)
Спеціальність 121 – Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри, голова циклової
комісії програмних засобів і технологій

_____ к.т.н., доцент О.Е. Огнєва
«___» _____ 2024 року

З А В Д А Н Н Я

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ

Петров Владислав Олександрович

(прізвище, ім'я, по батькові)

1. Тема проекту (роботи) «Розробка програмного рішення для прогнозування відтоку клієнтів банку»

керівник проекту (роботи) старший викладач Комісаров Олександр Сергійович

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «20» січня 2025 року №98-с_

2. Строк подання здобувачом проекту(роботи) 10.06.2025

3. Вихідні дані до проекту (роботи) _____

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

Дослідження та аналіз предметної області, проектування програмного рішення, програмна реалізація, тренування моделі ШІ, тестування та впровадження

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

Комп'ютерна презентація

6. Консультанти розділів проекту (роботи)

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

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів атестаційного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Отримання завдання	09.02.2024	Виконано
2	Підбір літератури	10.02.2024 – 25.02.2024	Виконано
3	Аналіз предметної області	28.02.2024 – 10.03.2024	Виконано
4	Розробка концептуальної моделі	12.03.2024 –22.03.2024	Виконано
5	Розробка та проектування системи	25.03.2024 – 30.03.2024	Виконано
6	Розробка та проектування бази даних	02.04.2024 – 10.04.2024	Виконано
7	Розробка інтерфейсу додатка	12.04.2024 – 20.04.2024	Виконано
8	Написання вихідного коду програми	21.04.2024 – 20.05.2024	Виконано
9	Оформлення пояснювальної записки	21.05.2024- 01.06.2024	Виконано
10	Захист кваліфікаційної роботи	20.06.2024	Виконано

Здобувач _____
(підпис)

В.О.Петров _____
(прізвище та ініціали)

Керівник проекту (роботи) _____
(підпис)

О.С.Комісаров _____
(прізвище та ініціали)

РЕФЕРАТ

Бакалаврська кваліфікаційна робота складається зі 110 сторінок, містить 21 рисуноків, 2 додатки, посилається на 30 джерел та має 1 таблицю.

Метою дослідження є створення ефективної моделі прогнозування відтоку клієнтів із подальшим формуванням рекомендацій для підвищення рівня їх утримання. Об'єктом дослідження виступає база даних клієнтів банку, що включає як активних, так і втрачених користувачів. Предметом є процес розробки, навчання та валідації моделі на основі алгоритмів машинного навчання.

У роботі послідовно реалізовано підготовку даних, аналіз та порівняння моделей класифікації, вибір найбільш ефективної, її тренування та оцінку якості за допомогою метрик точності. Також проведено аналіз важливості ознак з метою виявлення ключових факторів відтоку клієнтів. Методологічну основу дослідження становлять статистичний аналіз, алгоритми машинного навчання, крос-валідація та тестування.

У результаті створено програмне забезпечення, що дозволяє з високою точністю передбачати ймовірність відтоку клієнтів банку. Одержано інсайти щодо поведінкових чинників, які спричиняють втрату клієнтів, та розроблено практичні рекомендації для покращення клієнтського утримання. Новизна роботи полягає у використанні сучасних ML-підходів для вирішення задачі прогнозування в контексті банківської сфери.

Ключові слова: програмне рішення, машинне навчання, прогнозування, аналіз даних, відтік клієнтів.

АНОТАЦІЯ

Бакалаврська кваліфікаційна робота має обсяг 142 сторінок, містить 21 рисунок, 2 додатки, посилається на 60 джерел та включає 1 таблицю.

Мета роботи полягає в розробці програмного рішення з використанням штучного інтелекту для прогнозування відтоку клієнтів банку з подальшим аналізом та пошуком стимулюючих пропозицій для їх довготривалого утримання. Об'єктом дослідження виступає база даних клієнтів банку, що містить інформацію як про поточних клієнтів, так і про тих, які вже припинили користування банківськими послугами.

Предметом дослідження є розробка, навчання та валідація моделі прогнозування відтоку клієнтів банку на основі методів машинного навчання. Цей процес включає такі ключові етапи, як підготовка даних, що охоплює очищення та нормалізацію вхідної інформації. Далі відбувається побудова моделі через аналіз алгоритмів машинного навчання, вибір оптимального та його реалізацію. Необхідною складовою є оцінка моделі з використанням метрик для оцінки якості прогнозування. Завершується предмет дослідження аналізом важливості ознак для виявлення ключових факторів, що впливають на відтік клієнтів.

Для проведення дослідження застосовувались наступні методи: статистичний аналіз для дослідження розподілу даних, виявлення кореляцій і аномалій. Використовувалось машинне навчання, зокрема застосування алгоритмів класифікації для побудови прогнозової моделі. Методи оцінки моделі включали крос-валідацію та тестові вибірки для оцінки здатності моделі до узагальнення.

Результатами роботи є розроблене програмне рішення для прогнозування відтоку клієнтів банку. Сформовано та навчено модель з високою точністю прогнозування. Визначено ключові чинники, що впливають на втрату клієнтів. Розроблено рекомендації для банку щодо утримання клієнтів. Новизна роботи

полягає у застосуванні сучасних алгоритмів машинного навчання для вирішення задачі прогнозування та подальшої оптимізації бізнес-процесів.

Ключові слова: програмне рішення, машинне навчання, прогнозування, аналіз даних.

ABSTRACT

The Bachelor's qualification work consists of 110 pages, includes 9 figures, 1 appendix, refers to 7 sources, and contains 1 table.

The objective of the work is to develop a software solution using artificial intelligence for predicting bank customer churn, with subsequent analysis and the search for stimulating offers for their long-term retention.

The object of research is the bank's customer database, which contains information about both current clients and those who have already stopped using banking services. The subject of research is the development, training, and validation of a bank customer churn prediction model based on machine learning methods. This process includes key stages such as data preparation, covering the cleaning and normalization of input information. Model building follows through the analysis of machine learning algorithms, selecting the optimal one, and its implementation. An essential component is model evaluation using metrics to assess the quality of prediction. The subject of research concludes with the analysis of feature importance to identify key factors influencing customer churn.

The following methods were used for the research: statistical analysis to study data distribution and identify correlations and anomalies. Machine learning was utilized, specifically the application of classification algorithms to build a predictive model. Model evaluation methods included cross-validation and test samples to assess the model's generalization capability.

The results of the work include a developed software solution for predicting bank customer churn. A model with high prediction accuracy has been formed and trained. Key factors influencing customer loss have been identified. Recommendations for the bank regarding customer retention have been developed. The novelty of the work lies in the application of modern machine learning algorithms to solve the prediction problem and further optimize business processes.

Keywords: software solution, machine learning, prediction, data analysis.

ЗМІСТ

РЕФЕРАТ.....	5
АНОТАЦІЯ.....	6
ABSTRACT.....	8
ЗМІСТ.....	9
ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ.....	11
ВСТУП.....	12
РОЗДІЛ 1. ДОСЛІДЖЕННЯ ТА АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	14
1.1 Зародження машинного навчання.....	14
1.2. Причини використання.....	15
1.3 Основні типи типів алгоритмів машинного навчання.....	17
1.4 Використання ШІ для потреб бізнесу.....	19
1.5 Використовуємі в рішенні алгоритми.....	20
1.5.1 Логістична регресія.....	20
1.5.2 Градієнтний бустинг.....	24
1.5.3 Регресія Кокса.....	28
РОЗДІЛ 2. ПРОГРАМНА РЕАЛІЗАЦІЯ ОСНОВНИХ АЛГОРИТМІВ КЛАСИФІКАЦІЇ.....	33
2.1 Створення базової моделі для машинного навчання.....	33
2.2 Реалізація логістичної регресії.....	34
2.3 Реалізація градієнтного бустингу.....	37
2.3.1 Внутрішнє дерево — DecisionTreeRegressor.....	37
2.3.2 Реалізація ансамблю — GradientBoosting.....	38
2.4 Реалізація регресії Кокса.....	42
РОЗДІЛ 3. ПРЕПРОЦЕСІНГ ТА ТРЕНУВАННЯ.....	46
3.1 Базовий клас для препроцесінгу.....	46
3.2 Препроцесінг даних для Логістичної регресії.....	48
3.3 Препроцесінг для Випадкового лісу.....	54
3.4 Препроцесінг для регресії Кокса.....	60
3.5 Утиліти для тренування.....	65
3.6 Тренування моделі на базі логістичної регресії.....	68
3.7 Тренування моделі на базі випадкового лісу.....	71
3.8 Тренування моделі на базі регресії Кокса.....	74
РОЗДІЛ 4. АНАЛІЗ МОДЕЛЕЙ.....	78

4.1 Метрики Оцінки Моделей.....	79
4.2 Візуальні Порівняння Продуктивності.....	80
4.3 Специфіка Аналізу Вживаності (Регресія Кокса).....	85
4.4 Висновки щодо Порівняння Моделей.....	86
РОЗДІЛ 5 РОЗРОБКА WEB-ЧАСТИНИ.....	88
5.1 Backend частина.....	88
5.2 Frontend частина.....	93
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	99
ДОДАТОК А.....	101
Лістинг коду Backend частини.....	101
Лістинг коду Frontend-частини.....	113
ДОДАТОК Б.....	139
Вигляд програмного інтерфейсу.....	139

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- ПЗ – Програмне забезпечення;
- ML – Machine learning
- ШІ – Штучний інтелект (моделі на базі машинного навчання);
- МН – Машинне навчання (Machine Learning), використовується як альтернативна українська аббревіатура
- SVM – Support Vector Machines
- PCA – Principal Component Analysis
- t-SNE – t-Distributed Stochastic Neighbor Embedding
- ABC – Abstract Base Class
- API – Application Programming Interface
- SGD – Stochastic Gradient Descent
- KNN – k-Nearest Neighbors
- SMOTE – Synthetic Minority Over-sampling Technique
- RFE – Recursive Feature Elimination
- CSV – Comma Separated Values
- ROC AUC – Area Under the Receiver Operating Characteristic Curve
- TN – True Negative
- FP – False Positive
- FN – False Negative
- TP – True Positive

ВСТУП

Утримання клієнтів вважається життєво важливим для будь-якого бізнесу. Прогнозування відтоку дає змогу компаніям вживати заходів для утримання клієнтів, що, своєю чергою, призводить до збільшення прибутку та зміцнення ринкових позицій. У цій роботі зосереджено увагу на застосуванні сучасних методів машинного навчання для прогнозування відтоку клієнтів банку.

Для вирішення цього завдання проводиться аналіз ефективності різних алгоритмів класифікації, включно з логістичною регресією, деревами рішень, випадковим лісом, градієнтним бустингом та SVM. Метою дослідження є визначення оптимального алгоритму класифікації, здатного з максимальною точністю прогнозувати відтік клієнтів на основі наявних даних про їхню поведінку та взаємодію з банком. Очікується, що результати дослідження дадуть змогу розробити систему раннього попередження, яка дозволить банку вчасно вживати заходів для утримання клієнтів.

Предметом дослідження є розробка ефективної моделі, яка дозволить виявляти клієнтів з високою ймовірністю відтоку. Етапи дослідження включають збір даних з різних джерел, очищення даних від помилок і пропусків, формування навчальних і тестових вибірок, а також відбір та створення релевантних ознак. Далі відбувається вибір та налаштування алгоритмів, що передбачає порівняння ефективності різних алгоритмів машинного навчання та налаштування гіперпараметрів моделі для досягнення оптимальної точності. Якість моделей оцінюється за допомогою метрик точності та шляхом аналізу важливості ознак для розуміння факторів, що впливають на відтік клієнтів. Нарешті, розробляється система раннього попередження, яка автоматично ідентифікує клієнтів з високим ризиком відтоку.

Практична значущість отриманих результатів полягає в тому, що розроблена модель може бути використана для цільової маркетингової пропозиції клієнтам, які мають високу ймовірність відтоку, тим самим

зменшуючи маркетингові витрати та збільшуючи ефективність кампаній. Теоретична значущість полягає в застосуванні методів машинного навчання для вирішення бізнес-завдань та вивченні впливу різних типів даних на точність прогнозування. Результати можуть бути апробовані шляхом публікації в науковому журналі, індексованому в міжнародних наукометричних базах даних.