

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 атестаційної роботи
бакалавра
(освітньо-кваліфікаційний рівень)

на тему: «Розробка програмних засобів для аналізу відгуків користувачів
тематичних порталів»

Виконав: студент 4 курсу, групи 4ПР1
напряму підготовки (спеціальності)
121 -«Інженерія програмного забезпечення»
(шифр і назва напряму підготовки, спеціальності)

Фролов А.О.
(прізвище та ініціали)

Керівник д.т.н., професор Жарікова М.В.
(прізвище та ініціали)

Рецензент _____
(прізвище та ініціали)

Херсонський національний технічний університет _____
(повне найменування вищого навчального закладу)

Факультет Інформаційних технологій та дизайну _____
Кафедра _____ Програмних засобів і технологій _____

Освітньо-кваліфікаційний рівень _____ бакалавр _____

Галузь знань _____ 12 – Інформаційні технології _____

Спеціальність _____ 121 «Інженерія програмного забезпечення» _____

ЗАТВЕРДЖУЮ
Завідувач кафедри ПЗ і Т
д.т.н. проф. В.Г. Шерстюк _
“ ___ ” _____ 2021 р.

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ) СТУДЕНТУ

_____ Фролову А.О. _____

1. Тема проекту (роботи) _____ Розробка програмних засобів для аналізу відгуків користувачів тематичних порталів _____

керівник проекту (роботи) _____ д.т.н. професор Жарікова М.В. _____,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджена наказом вищого навчального закладу від “ ___ ” _____ № _____

2. Строк подання студентом проекту (роботи) _____ 25.05.2021 _____

3. Вихідні дані до проекту (роботи) _____ ДСТУ з обробки інформації, літературні та періодичні джерела, матеріали практики _____

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) _____ 1. Аналіз предметної області та постановка завдання; 2. Алгоритми машинного навчання та методи дослідження; 3. Проектування та розробка програмного застосунку; 4. Хід дослідження ефективності алгоритмів _____

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

1. Мета, об'єкт, предмет та задачі дослідження;

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 дипломного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Відбір та вивчення літературних джерел	25.09.2021	
2	Аналіз стану вирішення завдання на сучасному етапі	30.09.2021	
3	Побудова концептуальної моделі	3.10.2021	
4	Розробка математичної моделі	10.10.2021	
5	Побудова алгоритму функціонування програмного продукту	15.10.2021	
6	Написання вихідного коду програми	01.11.2021	
7	Налагодження програмного коду	15.11.2021	
8	Оформлення пояснювальної записки	02.12.2021	

Студент _____ Терновий Д.В.

Керівник проекту (роботи) _____ (підпис) _____ (прізвище та ініціали)
Жарікова М.В. _____ (прізвище та ініціали)

РЕФЕРАТ

Об'єкт дослідження – програмні засоби для аналізу відгуків користувачів тематичних порталів.

В першому розділі проаналізовано предметну область за поставлено задачу дослідження.

В другому розділі проаналізовано алгоритми машинного навчання, існуючі на даний час, такі як лінійна регресія, лінійна класифікація, дерева рішень.

В третьому розділі описано процес проектування ті розробки програмних засобів для аналізу відгуків користувачів тематичних порталів.

В червертому розділі розділі описано хід дослідження ефективності алгоритмів машинного навчання для аналізу відгуків користувачів тематичних порталів.

Повний обсяг роботи 71 сторінку. Містить 21 рисунок, 1 таблицю, 1 додаток, 26 літературних джерел.

Ключові слова: ПРЕДМЕТНА ОБЛАСТЬ, ІНФОРМАЦІЙНА БАЗА, РОЗПІЗНАВАННЯ ЦИФР, ДІАГРАМА ВАРІАНТІВ ВИКОРИСТАННЯ, ДІАГРАМА ПОСЛІДОВНОСТІ, ДІАГРАМА ДІЯЛЬНОСТІ, ІНФОРМАЦІЙНА СИСТЕМА, ТАБЛИЦЯ, СУТНІСТЬ, МОДЕЛЬ, ФОРМА, МАШИННЕ НАВЧАННЯ.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ	7
ВСТУП	8
РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАВДАННЯ	10
1.1 Аналіз сучасного стану питання	10
1.2 Поняття машинного навчання	11
1.3 Завдання аналізу даних	14
1.4 Постановка завдання	16
РОЗДІЛ 2 АЛГОРИТМИ МАШИННОГО НАВЧАННЯ ТА МЕТОДИ ДОСЛІДЖЕННЯ	17
2.1 Лінійні моделі регресії та класифікації	17
2.1.1 Модель лінійної регресії	17
2.1.2 Модель лінійної класифікації	21
2.2 Метрики якості	28
2.2.1 Метрики якості в задачах регресії	29
2.2.2 Метрики якості в задачах класифікації	30
2.2.3 Якість оцінок приналежності класу	33
2.1.3 Проблема перенавчання та боротьба з нею	24
2.3 Градієнтний бустінг	35
2.5 Древа рішень	28
РОЗДІЛ 3 ПРОЕКТУВАННЯ ТА РОЗРОБКА ПРОГРАМНИХ ЗАСОБІВ	39
3.1 Вхідні дані	39
3.2 Вибір програмних засобів	39
3.3 Оцінка втрат набору даних	43
3.4 Алгоритм перцептрону	44

3.5 Алгоритм Пегаса	47
3.6 Зміна міток	44
РОЗДІЛ 4 ХІД ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ АНАЛІЗУ ВІДГУКІВ	50
4.1 Збіжність алгоритмів	50
4.2 Реалізація автоматичного аналізатора відгуків	51
4.3 Класифікація та точність	52
4.4 Налаштування параметрів	53
4.5 Точність на тестовому наборі	55
4.6 Технічна характеристика	55
ВИСНОВОК	58
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	59
ДОДАТОК А	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

DILI - Drug-induced liver injury

КПК - Кишеньковий комп'ютер

CAI - computer assisted learning

VR - віртуальна реальність

ACLS - Advanced Cardiovascular Life Support

ЕКГ - електрокардіографія

EMR - Electronic medical record

CPOE - Computerized physician order entry

OPC - Open Platform Communications

РСА – аналіз основних компонентів

NLP - Natural Language Processing

ML - Машинне навчання

ВСТУП

Метою даної кваліфікаційної роботи було створення програмних засобів для застосування методів машинного навчання для аналізу відгуків користувачів тематичних порталів.

Машинне навчання — підрозділ штучного інтелекту, який вивчає методи побудови моделей, які здатні навчатися, та алгоритмів для їх побудови та навчання. Метою машинного навчання є часткова чи повна автоматизація розв'язання складних професійних задач в найрізноманітніших областях людської діяльності.

Машинне навчання знаходиться на стику математичної статистики, методів оптимізації та дискретної математики, але має також і власну специфіку, пов'язану з проблемами обчислювальної ефективності та перенавчання. Багато методів індуктивного навчання розроблялися як альтернатива класичним статистичним підходам. Багато методи тісно пов'язані з витягом інформації, інтелектуальним аналізом даних (Data Mining).

Наразі сфера застосування машинного навчання постійно розширюється. Повсюдна інформатизація призводить до накопичення величезних об'ємів даних в науці, виробництві, бізнесі, транспорті, медицині. Задачі прогнозування, управління та прийняття рішень, які виникають при цьому дуже часто зводяться до навчання по прецедентам. Раніше, коли таких даних не було, ці задачі або ж вирішувалися іншими методами, або ж взагалі залишалися нерозв'язаними.

Задача створення програмних засобів для застосування методів машинного навчання для аналізу відгуків користувачів тематичних порталів є доволі актуальною через недосконалість наукових і математичних методів. Тому виникає необхідність побудови ефективної системи моніторингу відгуків

користувачів, системи їх класифікації і прогнозування.

Робота містить чотири розділи.

В першому розділі проаналізовано предметну область за поставлено задачу дослідження.

В другому розділі проаналізовано алгоритми машинного навчання, існуючі на даний час, такі як лінійна регресія, лінійна класифікація, дерева рішень.

В третьому розділі описано процес проектування ті розробки програмних засобів для аналізу відгуків користувачів тематичних порталів.

В червертому розділі розділі описано хід дослідження ефективності алгоритмів машинного навчання для аналізу відгуків користувачів тематичних порталів.