

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ НАЦІОНАЛЬНИЙ  
ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО”**

Факультет прикладної математики  
Кафедра програмного забезпечення комп'ютерних систем

**КУРСОВИЙ ПРОЕКТ  
ТЕХНІЧНЕ ЗАВДАННЯ**  
з дисципліни “Бази даних”  
спеціальність 121 – Програмна інженерія  
на тему: Моніторингова система ресурсів з продажами смартфонів

**Студента**

групи КП-82

Суходольський Є.В.

\_\_\_\_\_  
(підпис)

**Викладач**

к.т.н, доцент кафедри

СПіСКС

Петрашенко А.В.

\_\_\_\_\_  
(підпис)

## ЗМІСТ

	стор.
1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ.....	3
2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ .....	3
3. МЕТА РОЗРОБКИ .....	3
4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ .....	3
5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД .....	6
6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА .....	6
7. ВИБІР ЗАСОБІВ РОЗРОБКИ .....	6
8. ЕТАПИ РОЗРОБКИ .....	7

## **1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ**

Найменування: моніторингова система онлайн-магазинів з продажами смартфонів: збір даних про ціни в залежності від кількості ядер процесора, тактової частоти процесора, обсягу оперативної пам'яті, розміру постійної пам'яті, діагоналі екрану, роздільної здатності відео, кількості мега пікселів основної камери, ємності акумулятору та ціни.

Галузь застосування: підвищення швидкості пошуку актуальних пропозицій щодо продажу смартфонів. Збір, аналіз та фільтрація інформації з пропозицій щодо продажу смартфонів. Визначення найбільш вигідних для покупки смартфонів за результатами пошуку користувачів, прогноз тенденцій ринку електроніки.

## **2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ**

Дата початку проекту – 11 лютого 2021 року (дата видачі завдання курсового проекту).

Дата закінчення проекту - 20 травня 2021 року ( захист курсового проекту).

## **3. МЕТА РОЗРОБКИ**

Метою розробки даного курсового проекту є набуття виконавцем (студентом) практичних навичок розробки сучасного програмного забезпечення, що взаємодіє з постреляційними базами даних, а також здобуття навичок оформлення відповідного текстового, програмного та ілюстративного матеріалу у формі проектної документації. У результаті виконання курсового проекту студент повинен вміти розробляти програмне забезпечення для постреляційних баз даних, володіти основами використання СУБД, а також інструментальними засобами аналізу великих обсягів даних.

## **4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**

- Підсистема попередньої обробки даних містить у собі:
  - Засоби генерації даних: має бути реалізовано ПЗ для збору інформації стосовно пропозицій щодо продажу смартфонів в онлайн-магазинах України, що розміщені на відповідних веб-сайтах: [foxtrot.com.ua](http://foxtrot.com.ua), [moyo.ua](http://moyo.ua), [rozetka.ua](http://rozetka.ua), [citrus.ua](http://citrus.ua), [comfy.ua](http://comfy.ua).

- Засоби фільтрації та валідації даних:
  - Доповнення ПЗ з попереднього пункту функціоналом фільтрації та перевірки даних на коректність, відсіювання дублікацій.
- База даних:
  - MongoDB
- Засоби реплікації даних:
  - оскільки для використання у даній розробці була обрана нереляційна база даних MongoDB, то використовуватимемо реплісети (Replica Sets).
- Засоби масштабування:
  - шардинг (sharding), який використовує MongoDB для обробки великої кількості даних.
- Засоби аналізу даних:
  - NumPy – бібліотека для роботи із великими масивами даних.
  - Matplotlib – бібліотека для візуалізації даних у вигляді 2D і 3D графіків.
  - Pandas – бібліотека, яка використовуватиметься як надбудова до NumPy для структуризації роботи із масивами даних.
- Задачі аналізу даних:
  - Структурувати усі отримані дані з різних веб-ресурсів про смартфони: кількість ядер, частота процесора, об'єм оперативної та постійної пам'яті, діагональ екрану, роздільна здатність відео, кількість мега пікселів основної камери, ємність акумулятору та ціна.
  - Здійснити валідацію даних про смартфони, отриманих після роботи ПЗ щодо збору інформації з інтернет-магазинів, та прибрати зайву інформацію із усіх масивів даних, а саме:
    - Атрибути, що стосуються кількості ядер, частоти процесора, об'єму оперативної пам'яті, діагоналі екрану, кількості мега пікселів, ціни та ємності акумулятору повинні містити лише

числові дані без вказання одиниць виміру кожної з характеристик;

- Атрибути, що стосуються роздільної здатності відео повинні містити строкові дані у форматі «число пікселів по горизонталі х число пікселів по вертикалі, формат дисплею смартфону(наприклад, Full HD)».
- Об'єднати інформацію про один і той же смартфон, отриману з різних веб-ресурсів, для відсіювання дублікацій.
- Заповнити пусті місця(атрибути кожного смартфона) у базі даних для встановлення більш чітких причинно-наслідкових зв'язків між характеристиками смартфонів та їх ціною.
- Провести кореляцію по всіх даних смартфона (кількість ядер, частота процесора, об'єм оперативної та постійної пам'яті, діагональ екрану, роздільна здатність відео, кількість мега пікселів основної камери, ємність акумулятору, ціна) для визначення кореляційних характеристик (місць з найбільшою кореляцією).
- У місцях, де найбільша кореляція, здійснити більш детальний аналіз для встановлення залежності між ціною та характеристиками смартфона.
- Сформувати статистичну оцінку параметрів зв'язку (характеристики смартфона у місцях з найбільшою кореляцією та ціна смартфона) у вигляді графіків.
- Знайти зв'язок між ціною смартфона та вказаними вище характеристиками.
- Засоби резервування та відновлення даних:
  - Передбачені при використанні Replica Sets у MongoDB.

## **5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД**

В даному курсовому проєкті використовується MongoDB. Ця база даних є об'єктно орієнтованою та дозволяє зберігати великі масиви неструктурованих даних. На відміну від SQL баз даних ми можемо зберігати дані у “сирому” об'єктному вигляді, який використовується програмою та є більш близьким за структурою до моделі даних, яку буде використовувати ПЗ написане з використанням мови програмування Python. Це пришвидшить збір, збереження та отримання даних програмним забезпеченням. Оскільки MongoDB є представником NoSQL баз даних, вона не потребує жорсткої схеми даних, що дозволяє пришвидшити процес розробки та зробити його більш гнучким. Окрім цього дана СУБД підтримує горизонтальне масштабування за допомогою шардингу з метою зменшення навантаження на кожен окремий вузол шляхом розподілення навантаження між ними всіма.

## **6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА**

Інтерфейс користувача буде консольним. Інтерфейсу користувача буде використовуватися для запуску на виконання ПЗ, його налаштування та передачі параметрів для збору та аналізу даних, генерації звітної інформації (графіків та діаграм) у вигляді збережених на диск файлів - зображень. Звітна інформація стосується візуалізації роботи засобів аналізу даних.

## **7. ВИБІР ЗАСОБІВ РОЗРОБКИ**

Мова програмування – Python 3.8. Дана мова програмування відносно проста та за думкою фахівців, які займаються розробкою ПЗ у сфері науки про дані, є найкращим рішенням для даного роду задач. Окрім цього для цієї мови існує велика кількість бібліотек для аналізу даних. Зокрема:

- numpy — математична бібліотека з підтриманням багатомірних масивів;
- matplotlib — бібліотека для графічного представлення даних, а саме для побудови графіків;
- pandas — бібліотека для обробки та аналізу даних, використовується для первинної обробки даних;
- sklearn — бібліотека алгоритмів машинного навчання, використовується для класифікації досліджених даних;

## 8. ЕТАПИ РОЗРОБКИ

№	Назва етапів розроблення	Термін виконання
1	Затвердження теми курсової роботи. Опрацювання відповідної літератури. Розроблення та узгодження технічного завдання.	21.02.2021
2	Аналіз постановки задачі	15.03.2021
3	Розробка засобів генерації даних.	10.04.2021
4	Додавання засобів фільтрації та валідації даних.	15.04.2021
5	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи.	20.04.2021
6	Додавання засобів аналізу даних (реалізацію алгоритмів буде запозичено у великих бібліотеках аналізу даних).	25.04.2021
7	Додавання засобів резервування та відновлення даних (призначені для оперативного та пакетного збереження фрагментів та усієї бази даних з можливістю її відновлення з урахуванням необхідності підключення додаткового комп'ютера як елемента горизонтального масштабування).	30.04.2021
8	Тестування програми	05.05.2021
9	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	10.05.2021
10	Захист курсової роботи	20.05.2021