КИЇВСЬКИЙ СТОЛИЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ БОРИСА ГРІНЧЕНКА
Кафедра комп'ютерних наук

Затверджено на засіданні кафедри
комп'ютерних наук
(протокол № 3 від 03.04.24)

РОБОЧА ПРОГРАМА ІСПИТУ

АНАЛІЗ ВЕЛИКИХ ДАНИХ

Спеціальності *111 Математика*
Освітнього рівня *першого (бакалаврського)*
Освітньої програми 111.00.01 *Математика*

2023-2024 навчальний рік

Опис програми іспиту

Київський столичний університет імені Бориса Грінченка

Кафедра комп'ютерних наук

Програма іспиту з дисципліни

«Аналіз великих даних»

3 курс – освітній рівень – перший (бакалаврський)

Спеціальність **111 Математика**

Освітня програма: **111.00.01 Математика**

Форма проведення: тестування на платформі Moodle в ЕНК дисципліни: https://elearning.kubg.edu.ua/course/view.php?id=24587

Тривалість проведення **1 год. 20 хв.**

Максимальна кількість балів: **40 балів**

Екзамен проводиться в університетській аудиторії у тестовій формі із використанням персональних комп'ютерів, якщо ситуація дозволяє проведення освітнього процесу офлайн. Якщо ж освітній процес проходить дистанційно, то екзамен проводиться онлайн в режимі відеоконференції засобами Google Meet.

Студент дає відповіді на запитання електронного тесту в системі Moodle. Тест містить 40 питань, які передбачають автоматичну (комп'ютерну) перевірку і оцінюються по 1 балу за кожне.

Екзамен проводиться із суворим дотриманням принципів академічної доброчесності, що передбачає недопустимість списування, фальсифікацій та обману. При порушенні студент відсторонюється від подальшого проходження екзаменаційного тесту із підсумковою оцінкою Fx за дисципліну. При виконанні завдань допускається користування довідковою літературою, таблицями значень функції, критеріїв та ін.

Підсумкова оцінка в балах (максимально 100 балів) за дисципліну є сумою результату поточного контролю за семестр (60 балів) та відповіді на екзамені (40 балів).

Перелік тем, які виносяться на іспит:

1. Вступ в аналітику Великих Даних. Основні поняття Big Data.
2. Інтернет Речей та зростання даних. Платформа Kaggle. DrivenData. Визначення великих даних.
3. Приклади великих даних у реальному світі. Відкриті дані. Приватність даних. Структуровані та неструктуровані дані.
4. Хмарні та туманні обчислення. Дані в спокої та дані в русі. Інфраструктура великих даних. Розподілені дані та їх обробка.
5. Життєвий цикл проекту аналітики великих даних
6. Можливості інструментів аналізу даних. Традиційна аналітика великих даних та аналітика нового покоління.
7. Життєвий цикл аналізу даних. Відкриті дані, їх формати та засоби обробки. Веб-скрепінг. Витягування, перетворення та завантаження даних.
8. Основні підходи аналітики великих даних

9. Перевірка статистичних гіпотез. Основні поняття. Параметричні тести. Непараметричні тести. Визначення моделей розподілу емпіричних даних. Приклад ідентифікації функції розподілу однорідної вибірки. Приклад ідентифікації функції розподілу неоднорідної вибірки. Перевірка статистичних гіпотез із використанням Microsoft Excel. Перевірка статистичних гіпотез засобами SPSS.
10. Дисперсійний аналіз. Однофакторний аналіз. Двофакторний аналіз. Приклад виконання дисперсійного аналізу. Приклад виконання рангового однофакторного аналізу
11. Кореляційний аналіз. Поняття кореляційного зв'язку між досліджуваними величинами. Групування даних для кореляційного аналізу. Коефіцієнт кореляції Пірсона. Коефіцієнт кореляції Спірмена. Множинний та частинний коефіцієнти кореляції. Кореляційний аналіз із використанням Microsoft Excel. Можливості SPSS у дослідженні кореляції.
12. Факторний аналіз. Метод головних компонент. Метод головних факторів. Інші методи факторного аналізу. Приклади проведення факторного аналізу.
13. Завдання та методи класифікації даних. Параметричні методи класифікації без навчання. Кластерний аналіз. Класифікація з навчанням. Приклади здійснення класифікації даних
14. Використання програмного пакета Weka для аналітики великих даних
15. Використання програмного пакета Weka: класифікація і кластеризація
16. Використання програмного пакета Weka: метод найближчих сусідів
17. Математичний апарат методу найбліжчих сусідів. Реалізація методу найближчих сусідів в WEKA.
18. Візуалізація даних за допомогою системи Statistica. Основні поняття та принципи роботи з таблицями в системі Statistica. Типи графічних даних в системі Statistica. Використання функції для аналізу великиї даних в системі Statistica
19. Візуалізація даних за допомогою інструменту Gephi. Основне призначення та принцип роботи інструменту візуалізації даних Gephi. Генерація даних. Формати графов Gephi
20. Екосистема Microsoft Power BI. Компоненти Power BI. Задачі, що вирішує Power BI. Знайомство та налаштування Power BI Desktop. Підключення до даних – Power Query, Excel, SQL Server, CSV, папки з даними, інтернет-джерела.
21. Підготовка даних до використання в Power BI. Основи Power Query. Робота з даними – типові операції з очищення та перетворення даних: Об'єднання даних з декількох джерел. Видалення непотрібних стовпців і рядків. Знаходження потрібних даних і застосування фільтрів. Додавання розрахункових та умовних стовпців. Екстракція додаткової інформації із стовпців типу дата. Очищення текстових стовпців. Заміна даних. Розбивка стовпців на кілька і навпаки. Групування даних. Створення зведених стовпців і навпаки. Транспонування таблиці, зміна порядку рядків.

22. Моделювання в Power BI. Створення та управління зв’язками в Power BI. Типи зв’язків.
23. Створення візуальних елементів в Power BI. Таблиця та Матриця. Гістограми, діаграми: звичайні, нормовані, з накопиченням. Лінійчата діаграма та діаграма з областями. Секторні, кільцеві, деревоподібні. Комбіновані діаграми. Водоспад, воронка, точкова та стрічкова діаграми.
24. Налаштування крос-фільтрації, фільтр візуального елементу, фільтр сторінки
25. Кастомні візуалізації в Power BI.
26. Створення інтерактивного звіту в Power BI.. Безпека доступу. Публікація звіту на портал
27. Створення дашборду в Power BI. Налаштування дашборду для використання на мобільному додатку. Варіанти спільної роботи з дашбордами, звітами на порталі. Налаштування автоматичного оновлення даних.

Екзаменатор

Глушак О.М.

Завідувач кафедри

Машкіна І.В.