

КЛАСИЧНИЙ ПРИВАТНИЙ УНІВЕРСИТЕТ

КАФЕДРА ЕКОНОМІКИ ТА СТАТИСТИКИ

ЗАТВЕРДЖУЮ

Ректор Класичного приватного
університету

Віктор ОГАРЕНКО

(підпис)

« 28 »

2025 р.



**ПРОГРАМА
АТЕСТАЦІЙНОГО ЕКЗАМЕНУ**

РІВЕНЬ ВИЩОЇ ОСВІТИ	ПЕРШИЙ (БАКАЛАВРСЬКИЙ)
СТУПІНЬ ВИЩОЇ ОСВІТИ	БАКАЛАВР
ГАЛУЗЬ ЗНАНЬ	11 МАТЕМАТИКА ТА СТАТИСТИКА
СПЕЦІАЛЬНІСТЬ	112 СТАТИСТИКА
ОСВІТНЯ ПРОГРАМА	КОМП'ЮТЕРНА СТАТИСТИКА ТА АНАЛІЗ ДАНИХ

Запоріжжя
2025

Програма атестаційного екзамену для здобувачів першого (бакалаврського) рівня вищої освіти галузі знань 11 Математика та статистика за спеціальністю 112 Статистика, освітня програма: Комп'ютерна статистика та аналіз даних / Укладачі: Л.Н. Сергєєва, д.е.н, професор, Ю.В. Онішкевич, к.е.н, Т.Ю. Огаренко, к.е.н. – Запоріжжя : КПУ, 2025. – 13 с.

Програма атестаційного екзамену схвалено на засіданні кафедри економіки та статистики

Протокол № 1 від «27» серпня 2025 року

Завідувач кафедри економіки та статистики
д.е.н., професор  А.Г. Семенов

Директор Інституту економіки та права
д.ю.н., професор  Ю.В. Абакумова

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Програма атестаційного екзамену розроблена згідно освітньо-професійної програми Комп'ютерна статистика та аналіз даних та навчального плану підготовки студентів ступеня вищої освіти бакалавр спеціальності 112 Статистика.

Мета атестаційного екзамену для здобувачів денної та заочної форм навчання за спеціальністю 112 Статистика, освітня програма Комп'ютерна статистика та аналіз даних – визначення рівня теоретичних знань та практичних навичок студентів згідно з вимогами кваліфікаційної характеристики.

Завдання атестаційного екзамену – виявити здатність студентів самостійно обирати і адекватно застосовувати методи аналізу даних, практично застосовувати означені методи із застосуванням комп'ютерної техніки.

Атестаційний екзамен включає завдання з обов'язкових дисциплін:

1. Математична статистика
2. Статистичне моделювання і прогнозування
3. Методи аналізу даних
4. Статистичний аналіз великих даних

Підготовку до екзамену з кожної дисципліни слід починати з повторення програми курсу, в якій визначається обсяг необхідних знань. Повторення та доопрацювання курсу доцільно проводити за підручниками, конспектами лекцій та додатковою літературою.

Атестаційний екзамен є складовою частиною завершального етапу підготовки бакалаврів зі спеціальності 112 Статистика.

Атестаційний екзамен зі спеціальності 112 Статистика, за умов всебічного аналізу його результатів, дозволяє найбільш вичерпно з'ясувати позитивний досвід та недоліки в організації, змісті й методиці викладання фахових дисциплін, а також самостійної роботи здобувачів.

Все це дає можливість визначити конкретні заходи щодо удосконалення викладання фахових дисциплін, накреслити шляхи поліпшення взаємозв'язку та спадкоємності у їх викладанні.

Складання комплексного кваліфікаційного екзамену проводиться **онлайн** (із застосуванням платформи ZOOM) на відкритому засіданні екзаменаційної комісії.

Комплексний кваліфікаційний екзамен проводиться в усній формі за екзаменаційним білетом, який містить три теоретичних запитання з дисциплін Математична статистика, Статистичне моделювання і прогнозування, Методи аналізу даних, Статистичний аналіз великих даних та практичного завдання з однієї з цих дисциплін.

Вимоги до знань і вмінь студентів

У результаті вивчення дисциплін студенти мають **знати**:

- основи вибіркового методу – поняття генеральної сукупності, вибірки, варіаційного ряду, емпіричних характеристик розподілу;
- теорію статистичного оцінювання – властивості точкових оцінок та основні методи їх побудови;
- принципи інтервального оцінювання – поняття довірчої ймовірності та довірчого інтервалу, методи побудови довірчих інтервалів для параметрів нормального розподілу;
- теорію перевірки статистичних гіпотез – види гіпотез, типи помилок, рівень значущості та потужність критерію;
- параметричні критерії – методи застосування критеріїв Стюдента та Фішера для порівняння середніх і дисперсій;
- непараметричні критерії – критерії згоди для перевірки гіпотези про закон розподілу;
- основи кореляційного аналізу – вимірювання сили та напрямку зв'язку між ознаками, коефіцієнт кореляції та його властивості;
- принципи регресійного аналізу – метод найменших квадратів для оцінки параметрів лінійної регресії, поняття детермінації, залишки та їх аналіз;
- основні етапи та принципи побудови статистичних моделей, включаючи логіко-статистичні передумови забезпечення їхньої адекватності;

- методологічні принципи багатофакторного прогнозування та критерії вибору методів прогнозу;
- правила формування інформаційної бази моделі та використання показників для опису об'єкта;
- принципи оцінювання параметрів лінійної та нелінійної парної регресії;
- теорію класичної множинної регресії та властивості її оцінок;
- методи виявлення та оцінки впливу порушень припущень методу найменших квадратів;
- особливості побудови регресії на змішаних множинах чинників та багатофакторних індексних моделей;
- методи багатофакторного ранжирування та побудови рейтингових оцінок;
- основи моделювання взаємозв'язку в комбінаційних групуваннях та концепцію моделі стандартизованих групувань;
- теоретичні засади моделі головних компонент та її застосування для зменшення розмірності;
- принципи ідентифікації та вимірювання головних компонент;
- методи моделювання та прогнозування тенденцій розвитку та повних циклів;
- концепції моделей адаптивного прогнозування та інтегрованої моделі авторегресії;
- сутність моделі об'єкто-періодів та рекурентних моделей;
- класифікацію і особливості аналізу різних типів даних;
- етапи статистичного аналізу даних;
- теорію перевірки статистичних гіпотез;
- сутність дискримінантного, кластерного та факторного аналізу;
- особливості методів дожиття;
- методи інтелектуального аналізу даних, зокрема пошуку асоціативних правил;
- сутність, класифікацію та джерела великих даних;
- особливості зберігання та аналізу великих даних у порівнянні з традиційними системами;
- компоненти екосистеми Hadoop, а також потокові інтерфейси та інтерфейси каналів для вхідних даних до MapReduce;
- фреймворк MapReduce Daemon та модель програмування MapReduce;
- шаблони архітектури даних NoSQL;
- архітектуру Hive та Pig, команди оболонки Grunt, модель даних;
- архітектурні особливості Spark, компоненти програмного стеку та їхні функції;
- поняття потоку даних, моделі, архітектури, управління, джерел, запитів та проблем обробки потоків;
- методи вибірки, фільтрації та підрахунку окремих елементів у поточних обчисленнях;
- методи пошуку частих наборів елементів у потоці, обробки великих наборів даних та аналізу правил асоціації;
- методи моделювання бази даних як графу та представлення графів за допомогою трійок;
- поняття про графи, організацію графових мереж, вибір графів для аналітики; випадки використання графової аналітики.

вміти:

- здійснювати первинну обробку даних для систематизації статистичних даних;
- обчислювати вибіркові характеристики статистичних даних;
- застосовувати метод моментів та метод максимальної правдоподібності для знаходження ефективних, незсунених та спроможних оцінок параметрів розподілів;
- обчислювати та інтерпретувати довірчі інтервали для математичного сподівання та дисперсії нормально розподіленої сукупності;
- правильно формулювати нульову та альтернативну гіпотези, вибирати відповідний критерій, визначати критичну область та приймати статистичне рішення;
- застосовувати параметричні критерії для порівняння середніх і дисперсій;
- використовувати критерії згоди для перевірки гіпотези про закон розподілу;

- обчислювати коефіцієнт кореляції, оцінювати його значущість та інтерпретувати силу й напрямки лінійного зв'язку між змінними;
- оцінювати параметри парної лінійної регресії методом найменших квадратів та оцінювати якість побудованої моделі;
- формувати адекватну інформаційну базу для побудови статистичної моделі;
- використовувати методи багатофакторного ранжирування для побудови інтегральних рейтингових оцінок;
- оцінювати параметри парних та класичних множинних регресійних моделей та перевіряти їхню статистичну значущість;
- проводити діагностику регресійних моделей: виявляти та застосовувати методи корекції при мультиколінеарності та порушеннях гомоскедастичності/автокореляції;
- будувати моделі на основі змішаних множин чинників та багатофакторні індексні моделі;
- застосовувати моделі стандартизованих групувань для аналізу взаємозв'язку в комбінаційних групуваннях;
- ідентифікувати та моделювати тенденції розвитку та циклічні компоненти у часових рядах;
- будувати моделі адаптивного прогнозування та інтегровані моделі авторегресії;
- застосовувати рекурентні моделі та моделі об'єкто-періодів для аналізу динамічних систем;
- застосовувати модель головних компонент для зменшення розмірності та синтезу узагальнених факторів;
- ідентифікувати та вимірювати головні компоненти, інтерпретуючи їхній економічний зміст;
- обирати методи аналізу даних в залежності від їх типу;
- обирати та використовувати критерій для перевірки статистичної гіпотези;
- проводити дискримінантний, кластерний та факторний аналіз із застосуванням програмних засобів та надавати інтерпретацію отриманих результатів;
- будувати таблиці життя та криві Каплана-Мейєра, застосовувати лог-ранк тест для порівняння дожиття в двох групах;
- знаходити асоціативні правила і здійснювати їх інтерпретацію;
- створювати програми MapReduce для обчислень, таких як підрахунок, підсумовування, та алгоритмів реляційних алгебраїчних операцій, проекцій, об'єднань, перетинів, природних з'єднань, групування, операцій агрегації та множення матриць;
- застосовувати бази даних MongoDB та команди запитів та бази даних Cassandra, здійснювати їх інтеграцію з Hadoop;
- застосовувати HiveQL для запитів, сортування, агрегації, запитів скриптів, MapReduce Joins та підзапитів;
- завантажувати Spark, виконувати програмування за допомогою Spark;
- здійснювати аналіз поточкових даних з використанням платформи аналітики в реальному часі Apache SparkStreaming;
- використовувати Apache Spark GraphX, платформи графової аналітики великих даних.

РОЗДІЛ І. МАТЕМАТИЧНА СТАТИСТИКА

Предмет, задачі математичної статистики. Генеральна сукупність і вибірка. Предмет і завдання математичної статистики. Взаємозв'язок з теорією ймовірностей. Генеральна сукупність та її параметри. Вибірка, вимога репрезентативності. Види вибірок. Поняття статистичної оцінки параметра.

Первинна обробка даних: варіаційні ряди, графічне зображення. Побудова варіаційних рядів: дискретний та інтервальний ряди. Частота та відносна частота (частість) варіанти. Графічне зображення: полігон частот для дискретного ряду та гістограма для інтервального ряду. Емпірична функція розподілу та її властивості. Кумулята, її використання для знаходження квантилів.

Емпіричні числові характеристики: середні, дисперсія, моменти, показники форми. Вибіркове середнє, мода, медіана. Міри розсіювання: вибіркова дисперсія, середнє квадратичне відхилення, розмах варіації. Скоригована (незміщена) вибіркова дисперсія. Емпіричні моменти. Вибірковий коефіцієнт асиметрії та вибірковий коефіцієнт ексцесу.

Точкові статистичні оцінки та їх властивості. Методи оцінювання. Концепція точкової оцінки параметра. Властивості оцінок. Нерівність Рао-Крамера. Методи побудови оцінок: Метод моментів та метод максимальної правдоподібності.

Інтервальне оцінювання параметрів: довірча ймовірність, довірчі інтервали для середньої та дисперсії нормального розподілу. Інтервальна оцінка та її переваги. Довірча ймовірність та рівень значущості. Довірчий інтервал. Побудова довірчих інтервалів для середньої нормального розподілу: використання Z-розподілу та t-розподілу Стюдента. Довірчий інтервал для дисперсії.

Основні поняття теорії гіпотез. Нульова та альтернативна гіпотези, помилки I та II роду, рівень значущості та потужність критерію. Визначення статистичної гіпотези. Нульова та альтернативна гіпотези. Помилки прийняття рішень: Помилка I роду та помилка II роду. Рівень значущості. Критична область та область прийняття гіпотези. Потужність критерію.

Параметричні критерії. Перевірка гіпотез про середні та дисперсії двох сукупностей. Параметричні критерії та умови їх застосування. Загальний алгоритм перевірки гіпотези. Перевірка гіпотези про рівність середніх двох незалежних сукупностей. Перевірка гіпотези про рівність дисперсій двох сукупностей.

Критерії згоди. Перевірка гіпотези про закон розподілу. Призначення критеріїв згоди. Критерій χ^2 Пірсона: область застосування, формула емпіричного значення критерію, визначення числа ступенів свободи. Непараметричні критерії згоди. Критерій Колмогорова-Смирнова.

Кореляційний аналіз. Кореляційний зв'язок та його відмінність від функціонального. Коваріація. Коефіцієнт лінійної кореляції Пірсона, інтерпретація знаку та сили зв'язку. Перевірка статистичної значущості коефіцієнта кореляції. Непараметрична кореляція: Коефіцієнти рангової кореляції Спірмена та Кендалла.

Регресійний аналіз. Регресійний аналіз та регресійна модель. Модель парної лінійної регресії. Метод найменших квадратів – принцип роботи, оцінка параметрів. Коефіцієнт детермінації. Інтерпретація якості моделі, частка поясненої варіації. Залишкова дисперсія.

РОЗДІЛ ІІ. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ

Методологічні засади статистичного моделювання. Методологія та етапи статистичного моделювання. Основні принципи побудови статистичних моделей. Поняття статистичної моделі та її відмінність від детермінованої. Класифікація моделей за призначенням і способом побудови. Адекватність моделі та критерії її оцінювання.

Методологічні принципи багатофакторного прогнозування. Сутність та завдання багатофакторного прогнозування. Основні методологічні принципи: науковість, системність, безперервність, комплексність. Класифікація прогнозів за часовим горизонтом та об'єктом. Поняття прогнозової траєкторії та прогнозного інтервалу.

Формування інформаційної бази моделі і опис об'єкта моделювання. Вимоги до інформаційної бази статистичної моделі. Проблема мультиколінеарності та методи її виявлення і усунення. Критерії відбору факторів (чинників) до моделі. Специфікація моделі та вибір форми зв'язку. Етапи опису об'єкта та його декомпозиція.

Багатофакторне ранжирування, рейтингові оцінки Поняття багатофакторного ранжирування та його економічне значення. Методи інтегральної оцінки. Визначення вагових коефіцієнтів чинників. Рейтингові оцінки та їхнє використання в управлінні. Побудова рейтингових моделей на основі багатовимірних даних.

Лінійна та нелінійна парна регресія. Модель парної регресії. Лінійна парна регресія та оцінка її параметрів методом найменших квадратів. Коефіцієнт детермінації. Нелінійні форми зв'язку та методи їх лінеаризації.

Класична множинна регресія. Логіко-статистичні передумови забезпечення адекватності регресійних моделей. Класична модель множинної регресії. Вимоги теореми Гаусса-Маркова до залишків. Логіко-статистичні передумови адекватності: нормальність розподілу залишків, гомоскедастичність (сталість дисперсії залишків), відсутність автокореляції залишків. Критерії перевірки адекватності.

Регресія на змішаних множинах чинників. Змішані множини чинників як поєднання кількісних та якісних змінних. Використання фіктивних змінних для введення якісних ознак у регресійну модель. Інтерпретація коефіцієнтів при фіктивних змінних. Взаємодія чинників та її моделювання.

Багатофакторні індексні моделі. Індексний метод у статистичному моделюванні. Багатофакторні індексні моделі та їхнє використання для аналізу динаміки та структурних зрушень. Поняття ланцюгових та базисних індексів. Декомпозиція загального індексу на вплив окремих чинників.

Моделювання взаємозв'язку в комбінаційних групуваннях і модель стандартизованих групувань. Комбінаційні групування та їхнє застосування для вивчення складних взаємозв'язків. Модель стандартизованих групувань та її переваги перед простими групуваннями. Метод послідовних групувань для багатофакторного аналізу.

Моделювання і прогнозування тенденцій розвитку. Тенденція (тренд) як основна складова часового ряду. Методи вирівнювання часових рядів: ковзна середня, експоненційне згладжування. Аналітичне вирівнювання – лінійний, параболічний, експоненційний тренди. Прогнозування на основі тренду та оцінка похибки прогнозу.

Моделювання повних циклів і процесів оновлення. Циклічні та сезонні коливання у часових рядах. Мультиплікативна та адитивна моделі часового ряду. Індеси сезонності та їх обчислення. Процеси оновлення та моделі життєвого циклу продукту або процесу.

Моделі адаптивного прогнозування та інтегрована модель авторегресії. Адаптивне прогнозування, принцип коригування прогнозів на основі похибок. Експоненційне згладжування як основа адаптивних моделей. Моделі Холта та Вінтерса. Інтегрована модель авторегресії (ARIMA).

Модель об'єкто-періодів. Модель об'єкто-періодів як засіб аналізу панельних даних. Панельні дані: поєднання часового ряду та просторових об'єктів. Фіксовані ефекти та випадкові ефекти в панельній регресії. Переваги цієї моделі для врахування індивідуальних відмінностей об'єктів.

Рекурентні моделі. Рекурентні (авторегресійні) моделі, залежність поточного значення від попередніх значень того ж ряду. Лагові змінні та їхнє використання. Перевірка на стаціонарність часового ряду, критерій Дікі-Фуллера. Моделі ковзного середнього.

Методологічні принципи багатофакторного прогнозування. Етапи багатофакторного прогнозування. Оцінка точності та надійності прогнозів, середня абсолютна похибка, середня квадратична похибка. Порівняння прогностичних моделей та вибір оптимальної. Валідація та бек-тестування прогностичної моделі.

Модель головних компонент. Аналіз головних компонент як метод зниження розмірності даних. Перетворення набору корельованих змінних у набір некорельованих головних компонент. Модель головних компонент та її застосування у випадках сильної

мультиколінеарності.

Ідентифікація та вимірювання головних компонент. Критерії відбору головних компонент. Власні числа (дисперсія) та власні вектори (навантаження) компонент. Інтерпретація головних компонент. Обчислення внеску дисперсії кожної компоненти у загальну варіацію.

РОЗДІЛ III. МЕТОДИ АНАЛІЗУ ДАНИХ

Типи даних. Сутність поняття дані та джерела їх отримання. Класифікація типів даних. Поняття і структура статистичних показників. Абсолютні і відносні величини. Структурні зрушення в даних. Методи збору даних. Етапи аналізу даних.

Статистичні методи аналізу даних. Послідовність застосування методів статистичного аналізу даних: описова статистика, аналітична статистика, статистичне моделювання. Перевірка даних на нормальність розподілу. Представлення результатів описової статистики в залежності від виду розподілу. Параметричний та непараметричний кореляційний аналіз. Підготовка даних до аналізу, візуалізація даних.

Перевірка статистичних гіпотез. Схема тестування статистичних гіпотез. Параметричні та непараметричні критерії в тестуванні статистичних гіпотез. Характеристика та аналіз статистичних помилок. Інструменти тестування статистичних гіпотез. Тестування гіпотез щодо номінальних даних – хі-квадрат критерій.

Огляд методів статистичного моделювання. Вибір методу моделювання в залежності від завдання дослідження. Дисперсійний аналіз для перевірки однорідності груп. Завдання класифікації – дискримінантний аналіз та логістична регресія. Поділ сукупності на однорідні групи – кластерний аналіз. Зменшення розмірності даних – факторний аналіз. Пошук нових знань – методи інтелектуального аналізу.

Дискримінантний аналіз і модель логістичної регресії. Процедура дискримінантного аналізу. Помилки дискримінації. Дискримінантний аналіз в пакеті MS EXCEL. Дискримінантний аналіз в пакеті Statistica. Логістична регресія. Оцінки якості моделей класифікації – чутливість, специфічність, точність.

Кластерний аналіз. Сутність, поняття та завдання кластерного аналізу. Методи визначення відстані при побудові кластерів. Ієрархічні методи кластерного аналізу. Кластеризація методом k-середніх.

Основи факторного аналізу. Сутність факторного аналізу і процедура його проведення. Метод головних компонентів в факторному аналізі. Визначення кількості факторів. Сутність і методи ротації факторів. Змістовна інтерпретація факторів. Використання факторів в подальшому аналізі.

Методи аналізу дожиття. Основні поняття аналізу дожиття: подія, час життя, повне та цензуроване спостереження. Загальна характеристика методів аналізу дожиття. Побудова таблиці життя. Крива Каплана-Мейера та її інтерпретація. Медіана дожиття. Порівняння двох груп – лог-ранк тест. Регресійна модель пропорційних ризиків Кокса.

Методи інтелектуального аналізу даних. Сутність та класифікація методів інтелектуального аналізу даних (ІАД). Залачі інтелектуального аналізу даних. Зв'язок методів інтелектуального аналізу з іншими напрямками математики та ІТ. Етапи проведення ІАД. Напрями застосування ІАД.

Методи та інструменти пошуку асоціативних правил. Математична постановка задачі пошуку асоціативних правил. Характеристики асоціативних правил: підтримка, достовірність, підйом. Класифікація асоціативних правил. Етапи алгоритму пошуку асоціативних правил. Порівняння алгоритмів пошуку асоціативних правил. Вибір інструментів пошуку асоціативних правил. Система PolyAnalyst. Модулі Market Basket Analysis (BA) та Transactional Basket Analysis (TB).

Технологія пошуку асоціативних правил в програмі Statistica. Завантаження даних для аналізу. Вибір «Правила асоціації» (Association Rules) в меню інтелектуального аналізу

даних (Data Mining), щоб відобразити панель запуску «Правила асоціації» (Association Rules Startup Panel). Вибір змінних і параметрів пошуку. Аналіз та візуалізація результатів.

Сучасні тенденції розвитку методів аналізу даних. Вплив розвитку економіки і збільшення обсягів даних на розвиток методів аналізу даних. Технології штучного інтелекту (Artificial intelligence – AI), машинного навчання (Machine Learning – ML), індустріального Інтернету речей (Industrial Internet of Things – IIoT), аналітика великих обсягів даних (Big Data).

РОЗДІЛ IV. СТАТИСТИЧНИЙ АНАЛІЗ ВЕЛИКИХ ДАНИХ

Вступ до аналітики великих даних. Концептуальне розуміння даних та веб-даних; класифікація даних як структурованих, напів-, мульти- та неструктурованих даних; характеристики, типи, класифікації та методи обробки великих даних. Масштабованість, масово паралельна обробка (MPP), розподілені, хмарні та мережеві обчислення. Рівні проектування архітектури обробки даних для управління даними та аналітики. Джерела даних, якість даних, попередня обробка даних та експорт сховища даних до хмари. Зберігання та аналіз даних; порівняння між традиційними системами, такими як система керування реляційними базами даних (RDBMS), корпоративними серверами, сховищами даних та підходами до зберігання та аналітики великих даних. Варіанти використання та застосування великих даних у різних галузях.

Hadoop та його екосистема. Ядро Hadoop, компоненти екосистеми Hadoop, а також потокові інтерфейси та інтерфейси каналів для вхідних даних до MapReduce. Розподілена файлова система Hadoop (HDFS) та фізична організація вузлів для обчислень у кластерах великомасштабних файлів.

Фреймворк та модель програмування MapReduce. Фреймворк MapReduce Daemon та модель програмування MapReduce. Функції Hadoop YARN, управління та планування ресурсів, а також паралельна обробка завдань програми. Функції інструментів екосистеми Hadoop.

Управління великими даними NoSQL, MongoDB та Cassandra. Концептуальне розуміння сховищ даних NoSQL, рішень для великих даних, моделей без схем та підвищеної гнучкості маніпулювання даними. Шаблони архітектури даних NoSQL, а саме: пари ключ-значення, табличні, сімейства стовпців, BigTable, стовпцеві стовпці записів (RC), оптимізовані стовпці рядків (ORC) та Parquet, сховища даних документів, об'єктів та графів, а також варіації архітектурних шаблонів. Управління сховищами даних NoSQL, застосування та проблеми обробки у великих даних. Вирішення проблеми аналізу великих даних за допомогою архітектури "без спільного доступу", чотири способи, за допомогою яких NoSQL обробляє проблеми великих даних. Застосування бази даних MongoDB та команди запитів. Використання бази даних Cassandra, модель даних, клієнтів та інтеграція їх з Hadoop

Технології MapReduce, Hive and Pig. Відображення в MapReduce завдання за допомогою сховища ключ-значення, групування за ключами, скорочування завдання за допомогою об'єднувачів та дії у разі відмови вузлів. Створення програм MapReduce для обчислень, таких як підрахунок, підсумовування, та алгоритмів реляційних алгебраїчних операцій, проєкцій, об'єднань, перетинів, природних з'єднань, групування, операцій агрегації та множення матриць. Архітектура Hive, встановлення, порівняння сховища даних Hive з традиційними базами даних. Застосування HiveQL для запитів, сортування, агрегації, запитів скриптів, MapReduce Joins та підзапитів. Архітектура Pig, команди оболонки Grunt, модель даних, Pig Latin, розробка скриптів та розширюваність за допомогою UDF.

Spark та аналітика великих даних. Архітектурні особливості Spark, компоненти програмного стеку та їхні функції. Кроки аналізу за допомогою Spark, Spark разом з Python, розширені функції, UDFS, векторизовані UDF, згруповані векторизовані UDF та бібліотеки аналітики Python. Завантаження Spark, початок програмування за допомогою Spark, оболонки Spark, контекст Spark, розробка та тестування коду, програмування за допомогою RDD та застосування MLib. Процеси ETL за допомогою вбудованих функцій, операторів та конвеєрів ETL. Аналітика, звітність даних/інформації та методи візуалізації.

Платформа SparkStreaming для аналізу потоків даних та аналітики в реальному часі. Поняття потоку даних, моделі, архітектури, управління, джерел, запитів та проблем обробки потоків. Методи вибірки, фільтрації та підрахунку окремих елементів у поточних обчисленнях. Методи пошуку частих наборів елементів у потоці, обробки великих наборів даних та аналізу правил асоціації. Платформа аналітики в реальному часі Apache SparkStreaming та застосування аналітики в реальному часі для аналізу настроїв та аналізу цін на акції.

Аналітика з використанням графів для великих даних та платформа Spark GraphX. Моделювання бази даних як графу та представлення графів за допомогою трійок. Поняття про графи, організацію графових мереж, вибір графів для аналітики; випадки використання графової аналітики. Параметри графів, методи, діагностика та рішення, статистичні моделі, StatsModel, графова аналітика на основі ймовірностей. Технічні складнощі аналізу графів. Використання Apache Spark GraphX, платформи графової аналітики великих даних. Функції, архітектура та компоненти. Застосування їх для графової аналітики.

КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАНЬ

При оцінюванні якості виконання екзаменаційних завдань враховуються загальні вимоги до них, що забезпечують максимальну оцінку:

- здатність до застосування знань, їх диференціювання, інтеграції та уніфікації аналізу фактів, подій, прогнозу результатів;
- правильність та повнота рішень;
- грамотність, лаконізм і логічна послідовність викладу;
- оформлення відповідно до чинних стандартів.

Підсумкову оцінку визначає екзаменаційна комісія, користуючись такими критеріями.

<i>Оцінка</i>	<i>Критерії оцінювання</i>
A 90-100 (відмінно)	ставиться у випадках, коли студент має систематичні та глибокі знання навчального матеріалу, вміє без помилок виконувати практичні завдання, передбачені програмою державного іспиту, знає основну літературу, подає матеріал у логічній послідовності, робить узагальнення і висновки, наводить практичні приклади у контексті тематичного теоретичного матеріалу.
B 82-89 (добре)	ставиться, якщо студент повністю засвоїв навчальний матеріал, знає основну літературу, вміє виконувати практичні завдання, подає матеріал у логічній послідовності, робить певні узагальнення і висновки, але не наводить практичних прикладів у контексті тематичного теоретичного матеріалу або допускається незначних помилок у формулюванні термінів, категорій, невеликих арифметичних помилок у розрахунках при розв'язанні практичних завдань.
C 75-81 (добре)	ставиться, якщо студент засвоїв навчальний матеріал, знає основну літературу, вміє виконувати практичні завдання, подає матеріал у логічній послідовності, не робить певні узагальнення і висновки.
D 67-74 (задовільно)	ставиться у випадках, коли студент засвоїв матеріал не у повному обсязі, дає неповну відповідь на поставлене теоретичне запитання.
E 60-66 (задовільно)	ставиться у випадках, коли студент засвоїв матеріал не у повному обсязі, дає неповну відповідь на поставлене теоретичне запитання, допускається грубих помилок у розрахунках при розв'язанні практичних завдань.

Оцінка "Незадовільно" ставиться студентові, який не засвоїв навчальний матеріал, дає неправильні відповіді на поставлені теоретичні запитання, не вміє або неправильно виконує розрахунки при розв'язанні практичних завдань.

НАВЧАЛЬНО-МЕТОДИЧНА ЛІТЕРАТУРА

1. Бакун, В.В. Теорія ймовірностей, випадкові процеси та математична статистика [Електронний ресурс]: підручник для здобувачів ступеня бакалавра за спеціальністю 122 Комп'ютерні науки / В. В. Бакун ; КПІ ім. Ігоря Сікорського, 2022. – 286 с.
2. Герич М.С., Синявська О.О. Математична статистика: навч. посіб. Ужгород: ДВНЗ “УжНУ”, 2021. – 146 с.
3. Данілов В.Я. Статистична обробка даних: навчальний посібник. К.: КНУ ім. Тараса Шевченка, 2019. – 156 с.
4. Завгородній О.І., Соловиченко О.В. Стороженко І.П., Левкін Д.А., Сичова Т.В. Математична статистика: основи теорії та методика розв'язування задач з варіантами індивідуальних завдань для студентів першого (бакалаврського) рівня вищої освіти денної та заочної форм навч. інж. спец. Держ. біотехнолог. ун-т. Харків: ДБТУ, 2022. – 57 с.
5. Математична статистика. Обробка дослідних даних методами математичної статистики: рекомендації до виконання індивідуального завдання [Електронний ресурс]: навчальний посібник для здобувачів ступеня бакалавра за освітньою програмою «Системний аналіз і управління» спеціальності 124 Системний аналіз / КПІ ім. Ігоря Сікорського; уклад.: І.Ю. Каніовська, О.В. Стусь.– Київ : КПІ ім. Ігоря Сікорського, 2023. – 128 с.
6. Математична статистика: навч. посібник / П.І. Бідюк, Б.П. Ткач, Т. Харрінгтон. – К.: ДП «Вид. дім «Персонал», 2018. – 248 с.
7. Перегуда О.В. Статистична обробка даних: навч. посібник [Електронний ресурс]. / О.В. Перегуда, О.А. Капустян, О.Б. Курилко. – Київ: КНУ ім. Тараса Шевченка, 2022. – 103 с.
8. Синявська О. О. Математична статистика: збірник завдань до типових розрахункових робіт для студентів математичного факультету. – Ужгород: ДВНЗ “УжНУ”, 2020. – 44 с.
9. Благун І.С., Кічор В.П., Селюченко Н.Є. Економічне прогнозування: теоретичні та прикладні аспекти: підручник за ред. В.П. Кічора. Львів: Растр-7, 2020. – 290 с.
10. Благун І.С., Кічор В.П., Скворцов Д.І. Статистичний аналіз і моделювання соціально-економічних об'єктів та процесів. Львів: Растр-7, 2022. – 400 с.
11. Голубова Г.В. Опорний конспект лекцій з навчальної дисципліни «Статистичне моделювання та прогнозування» для студентів освітньо-професійної програми «Прикладна статистика та бізнес аналітика» денної форми навчання. К.: НАСОА. 2022. – 113с.
12. Єріна А.М. Статистичне моделювання та прогнозування: Навч. посібник. – К.: КНЕУ, 2001. – 170 с.
13. Касьяненко В.О., Старченко Л.В. Моделювання та прогнозування економічних процесів. Київ: Унів. кн., 2023. – 185 с.
14. Кічор В.П. Економіко-статистичне моделювання: навч. посіб. Львів: Растр-7, 2017. – 350 с.
15. Основи статистичного моделювання: навч. посіб. за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП «Рута», 2022. – 604 с.
16. Шпігельхальтер Д. Мистецтво статистики. Прийняття аргументованих рішень на основі даних. Київ: КМ-БУКС, 2023. – 384 с.
17. Бідюк П.І., Данилов В.Я., Жиров О.Л. Прикладна статистика. – НТУУ «КПІ», 2022. – 180 с.
18. Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2018. – 482 с.
<http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf>
19. Литвин В.В., Пасічник В.В., Нікольський Ю.В. Аналіз даних та знань: підручник. - Л.: Магнолія, 2020.- 276 с.
20. Навчальний посібник з дисциплін “Аналіз даних” та “Аналіз даних в управляючих системах” для студентів спеціальності 126 - Інформаційні системи та технології / Гавриленко О.В., – Київ: КПІ, 2020. – 85с.
21. Методи аналізу даних у психологічних дослідженнях : навчальний посібник / В. М. Краєвський, Я. О. Остапенко, Т. М. Паянок, Н. В. Параниця ; Державний податковий

університет. – Ірпінь, 2024. – 144 с.

22. Болюбаш Н. М. Інтелектуальний аналіз даних : навч. посіб. / Н. М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. – 320 с.

23. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.

24. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. Київ : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018. 300 с.

25. Steven S. Skiena Computer Science Department Stony Brook University Stony Brook, NY USA, 2017 453 p.

26. Agbiniya J. Applied Data Analytics - Principles and Applications. River Publisher, 2020. 370 p. ISBN: 9788770220965.

27. Jamsa Kris. Introduction to Data Mining and Analytics: With Machine Learning in R and Python. Jones & Bartlett Learning, 2021. 668 p. ISBN 1284180905.

28. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. / Д.В. Ланде, І.Ю. Субач, А.Я. Гладун. / - Київ: ТОВ "Інжиніринг", 2021. - 168 с. ISBN 978-966-2344-83-7

29. Талах М.В. Технології обробки Big Data. Навчальний посібник/ Талах М.В., – Чернівці: Чернівецький нац.ун-т, 2024. – 454 с

30. Технології роботи з Big Data [Електронний ресурс] : методичні рекомендації до лабораторних робіт для студентів спеціальності 126 "Інформаційні системи та технології" освітньої програми "Інформаційні системи та технології" першого (бакалаврського) рівня / уклад. І. В. Кобзев, Н. М. Кобзева, О. В. Тесленко. – Харків : ХНЕУ ім. С. Кузнеця, 2023. – 83 с

31. Таран В. І., Гордієнко Ю. Г., Стіренко С. Г. ТЕХНОЛОГІЇ BIG DATA Практикум Київ КПІ ім. Ігоря Сікорського 2022 56 с.

32. Олещенко Л.М. ТЕХНОЛОГІЇ ОБРОБЛЕННЯ ВЕЛИКИХ ДАНИХ Конспект лекцій Київ КПІ ім. Ігоря Сікорського 2021 227 с.

33. Кучеров Д.П. Методи аналізу великих даних «Big Data» / Д.П. Кучеров. – Київ. – 2020 - 237 с.

34. Статистика: основи теорії та практикум: Навчальний посібник /Григорків В.С., Вінничук О.Ю., Григорків М.В., Маханець Л.Л. –Чернівці : Чернівець. нац. ун-т, 2022. – 304 с

35. Rodgers, W. J., Morris-Mathews, H., Romig, J. E., & Bettini, E. (2021). Observation Studies in Special Education. A Synthesis of Validity Evidence for Observation Systems. <https://doi.org/10.3102/00346543211042419>.

36. Data Science & Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data EMC Education Services [https://aitskadapa.ac.in/e-books/AI&DS/BIG%20DATA/Data%20Science%20_%20Big%20Data%20Analytics%20\(%20PDFDrive%20\).pdf](https://aitskadapa.ac.in/e-books/AI&DS/BIG%20DATA/Data%20Science%20_%20Big%20Data%20Analytics%20(%20PDFDrive%20).pdf)

37. Raj Kamal, Preeti Saxena BIG DATA ANALYTICS Introduction to Hadoop, Spark, and Machine-Learning <https://jcer.in/jcer-docs/E-Learning/Digital%20Library%20/E-Books/Big%20Data%20Analytics%20by%20RaJ%20Kamal,%20Preethi%20Saxena,TMH.pdf>

38. BERNARD MARR BIG DATA: USING SMART BIG DATA, ANALYTICS AND METRICS TO MAKE BETTER DECISIONS AND IMPROVE PERFORMANCE <https://bernardmarr.com/wp-content/uploads/2022/05/Big-Data.pdf>

ІНФОРМАЦІЙНІ РЕСУРСИ В МЕРЕЖІ ІНТЕРНЕТ

1. Coursera: "Statistical Inference" –курс від Johns Hopkins University.
2. <http://www.nbuv.gov.ua> – Національна бібліотека України імені В. І. Вернадського.
3. <http://www.nbuv.gov.ua> – Національна бібліотека України імені В.І. Вернадського.
4. https://prometheus.org.ua/course/course-v1:IRF+Stat101+2016_T3 – Prometheus: Аналіз даних та статистичне виведення на мові R
5. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч.

посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергєєв-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с.

6. Основи статистики та аналізу даних. <https://socialdata.org.ua/manual/manual4/>

7. A brief introduction to two data processing architectures — Lambda and Kappa for Big Data // Електронний ресурс. Режим доступу: <https://towardsdatascience.com/a-briefintroduction-to-twodata-processing-architectures-lambda-and-kappa-for-big-data4f35c28005bb>

8. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>