

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра математичної статистики і диференціальних рівнянь



Затверджено

На засіданні
кафедри математичної статистики і
диференціальних рівнянь
механіко-математичного факультету
Львівського національного університету
імені Івана Франка
(протокол № 1 від 22.06.2023 р.)

Завідувач кафедри: Олег БУГРІЙ

Силабус з навчальної дисципліни
“Аналіз потокових даних”,
що викладається в межах ОПШ “ Статистичний аналіз даних ”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 112 - Статистика

Львів 2023 р.

| | |
|--|---|
| Назва дисципліни | Аналіз потокових даних |
| Адреса викладання дисципліни | Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1, 79000 |
| Факультет та кафедра, за якою закріплена дисципліна | Механіко-математичний факультет Кафедра математичної статистики і диференціальних рівнянь |
| Галузь знань, шифр та назва спеціальності | 11 - Математика та статистика 112 - Статистика |
| Викладачі дисципліни | Бокало Т.М., кандидат фізико-математичних наук, асистент кафедри математичної статистики і диференціальних рівнянь |
| Контактна інформація викладачів | taras.bokalo@lnu.edu.ua , https://new.mmf.lnu.edu.ua/employee/bokalo_t_m |
| Консультації з питань навчання по дисципліні відбуваються | Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю). Головний корпус ЛНУ ім. І. Франка, каб. 267. м. Львів, вул. Університетська, 1 |
| Сторінка курсу | https://new.mmf.lnu.edu.ua/course/apotd-112-bak23 |
| Інформація про дисципліну | Дисципліна “Аналіз потокових даних” є вибірковою дисципліною з спеціальності 112 – Статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 8-му семестрі в обсязі 5-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS). |
| Коротка анотація дисципліни | Курс надає базові знання про концепції, методи, мови та системи для прийому, обробки та аналізу даних, які надходять для прийняття рішень у реальному часі (потокові дані). Курс спрямований на приборкання швидкісних розмірів великих даних, враховуючи обсяг і різноманітність вимірів даних. |
| Мета та цілі дисципліни | <i>Мета:</i> формування у майбутніх спеціалістів повноцінних теоретичних знань та практичних навичок по застосуванню основних підходів формування системи для прийому, обробки та аналізу даних, які надходять для прийняття рішень у реальному часі (потокові дані). <i>Цілі:</i> викласти основні теоретичні положення щодо вміння приборкати швидкісні розміри великих даних, враховуючи обсяг і різноманітність вимірів даних. |
| Література для вивчення дисципліни | <ol style="list-style-type: none"> 1. Jay Kreps. I Love logs: Event data, stream processing, and data integration., O’Reilly Media, Inc., 2014 2. Gwen Shapira, Todd Palino, Rajini Sivaram, Krit Petty. Kafka: The Definitive Guide, 2nd Edition., O’Reilly Media, Inc., 2021 3. Fabian Hueske, Vasiliki Kalavri. Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications 1st Edition., O’Reilly Media, Inc., 2019 4. Spark Structured Streaming [Electronic resource] / Spark. – URL https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html |

| | |
|---|---|
| | <p>5. Tyler Akidau, Slava Chernyak, and Reuven Lax. Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing., O'Reilly Media, Inc., 2018</p> <p>6. Kafka Documentation [Electronic resource] / Apache Kafka. – URL https://kafka.apache.org/documentation/</p> <p>7. Flink Documentation [Electronic resource] / Apache Flink. – URL https://nightlies.apache.org/flink/flink-docs-release-1.15/</p> |
| Обсяг курсу | Загальний обсяг: 150 годин. Аудиторних занять: 78 год., з них 39 год. лекційних та 39 год. лабораторних занять. Самостійної роботи: 72 год. |
| Очікувані результати навчання | <p>У результаті вивчення даного курсу студент буде:</p> <p>знати: Як визначати проблеми, які можна вирішити за допомогою методів великих даних, пов'язаних із швидкістю, і застосовувати технології аналізу поточкових даних для вирішення реальних проблем;</p> <p>вміти: Враховуючи конкретні ситуації на проєкті, студенти зможуть визначити та реалізувати рішення для аналізу поточкових даних для задачі та застосувати його до реальних потоків даних із соціальних мереж та датчиків IoT.</p> |
| Ключові слова | Потокові дані, аналіз, великі дані, Spark |
| Формат курсу | Очний |
| Теми | Див. Схема курсу |
| Підсумковий контроль, форма | Залік у кінці семестру |
| Пререквізити | <p>Для вивчення даного курсу студенти потрібні базові знання з:</p> <ul style="list-style-type: none"> • Database management. • SQL. • Machine Learning. |
| Навчальні методи та техніки, які будуть використовуватися під час викладання курсу | Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація); дедуктивні методи на основі узагальнень; евристичні методи (проблемна лекція); інтерактивні методи (дискусія) |
| Необхідне обладнання | <p>Комп'ютер (мінімальні характеристики: процесор Intel Core i3, 4ГБ оперативної пам'яті), доступ до мережі Internet.</p> <p>Необхідне програмне забезпечення включає в себе ОС Windows 10/Mac OS, програмні додатки - MS Teams, dbeaver (DBMS), IDE зручне для компіляції коду</p> |
| Критерії оцінювання (окремо для кожного виду навчальної діяльності) | <p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: 20% семестрової оцінки за активну роботу на заняттях, 80% семестрової оцінки за виконання практичних аудиторних і домашніх завдань, максимальна кількість балів 35. • Змістовий модуль 2: 20% семестрової оцінки за активну роботу на заняттях, 80% семестрової оцінки за виконання практичних аудиторних і домашніх завдань, максимальна кількість балів 35. |

- Контрольне тестування: максимальна кількість балів 30.
Підсумкова максимальна кількість балів 100.

Академічна доброчесність: Очікується, що роботи студентів будуть оригінальними дослідженнями чи міркуваннями. Списування та втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні завдань є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.

Жодні форми порушення академічної доброчесності не толеруються.

Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні/лабораторні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів робіт, передбачених курсом.

Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.

Політика виставлення балів. Враховуються бали, набрані при поточному контролі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.

Оцінювання практичних робіт (2 змістових модулі містять по 5 практичних робіт кожен, загалом 10 самостійних практичних робіт, максимальна кількість балів: 100) відбувається шляхом оцінки роботи студента під час захисту написаної студентом вдома практичної роботи (0-10 балів за одну роботу).

Бали оцінювання аудиторного виконання практичних робіт нараховуються за наступним співвідношенням:

10 – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі;

7 – студент достатньо розуміє розглянутий матеріал та принципи написаного ним матеріалу, присутні неточності та незначні помилки у відповідях на запитання по темі;

5 – студент не досить добре розуміє розглянутий матеріал та написаний ним матеріал, вагається та надає неточні/не конкретні відповіді на запитання по темі;

2 – студент погано розуміє розглянутий матеріал, студент в більшості надає помилкові відповіді на питання по темі;

0 - студент зовсім не засвоїв розглянутий матеріал.

Оцінювання контрольного тестування (завдання з тематики кожного змістового модуля) відбувається шляхом оцінки письмових відповідей студента на поставлені запитання.

Відсотки нарахування балів оцінювання відповіді на кожне запитання

| | |
|--|---|
| | <p>нараховуються за наступним співвідношенням: 75-100% – тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно; 50-75% – відтворюється значна частина розглянутої теми, проте присутні неточності та/або невідповідності; 25-50% – виявлено множинні неточності та невідповідності, пояснення відсутні чи частково помилкові; 0-25% – тему майже не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи.</p> <p>Критерії оцінювання результатів неформальної освіти: Нарахування балів відбувається за публікацію студентом тез доповідей на конференціях, наукових статей, за участь студента у діяльності наукових гуртків, семінарів, круглих столів, конкурсів, участь у заходах неформальної освіти, за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах провідних ІТ компаній за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни</p> |
| Питання для контрольного тестування | Створення та вичитування повідомлень за допомогою інструментів командного рядка Kafka, створення дашбордів аналітики в реальному часі за допомогою Flink, створення програми Spark Streaming, порівняння досвіду розробки з Flink. |
| Опитування | Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу. |

**Схема курсу “Аналіз поточкових даних”
для студентів спеціальності 112 - Статистика**

| Тижні | Лекційний курс | | Лабораторні заняття | | К-сть год СР | Літе- ратура |
|-------|--|-----------|---|-----------|--------------|------------------------|
| | Назва теми | К-сть год | Назва теми | К-сть год | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | Введення в поточкові дані та архітектуру | 3 | Налаштування простого пайплайну поточкових дани | 3 | 5 | [1]-[7], Сайт курсу |
| 2 | Основи Apache Kafka | 3 | Знайомство з Kafka як поточною платформою | 3 | 5 | [1]-[7], Сайт курсу |
| 3 | Формати серіалізації даних у Kafka | 3 | Створення повідомлень за допомогою інструментів командного рядка Kafka | 3 | 5 | [1]-[7], Сайт курсу |
| 4 | Основи поточної обробки | 3 | Аналіз повідомлень за допомогою інструментів командного рядка Kafka | 3 | 5 | [1]-[7], Сайт курсу |
| 5 | Apache Flink Essentials | 3 | Розробка програми Flink, впровадження відмовостійких поточкових конвеєрів | 3 | 5 | [1]-[7], Сайт курсу |
| 6 | Аналітика в реальному часі за допомогою Apache Flink | 3 | Створення дашбордів аналітики в реальному часі за допомогою Flink | 3 | 5 | [1]-[7], Сайт курсу |
| 7 | Apache Kafka Connect для інтеграції даних | 3 | Вивчення функцій обробки стану | 3 | 6 | [1]-[7], Сайт курсу |
| 8 | Розширені концепції поточної обробки | 3 | Створення програми Spark Streaming | 3 | 6 | [1]-[7], Сайт курсу |
| 9 | Інтеграція з Apache Spark Streaming | 3 | Порівняння досвіду розробки з Flink | 3 | 6 | [1]-[7], Сайт курсу |
| 10 | Моніторинг та оптимізація | 3 | Налаштування інструментів моніторингу для Kafka та Flink. | 3 | 6 | [1]-[7], Сайт курсу |
| 11 | Безпека поточкових даних | 3 | Визначення та оптимізація поточкового стріму | 3 | 6 | [1]-[7], Сайт курсу |
| 12 | Тематичні дослідження та найкращі практики | 3 | Аналіз і тиражування тематичних досліджень | 3 | 6 | [1]-[7], Сайт курсу |
| 13 | Підсумковий колоквіум. Підсумкове заняття | 3 | Впровадження практик у приклад поточкового проекту | 3 | 6 | [1]-[7], Сайт курсу |
| | Разом | 39 | | 39 | 72 | |
| | Викладач: Бокало Т.М. | | Викладач: Бокало Т.М. | | | |