

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра системного проектування

Затверджено

На засіданні
кафедри системного проектування
факультету електроніки та комп'ютерних
технологій
Львівського національного університету
імені Івана Франка
(протокол № 1 від 28.08.2023 р.)

Завідувач кафедри:



Роман ШУВАР

Силабус з навчальної дисципліни
“Методи та технології аналізу даних”,
що викладається в межах ОПП
“ Високопродуктивний комп'ютинг ”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 121 – Інженерія програмного забезпечення

Львів 2023 р.

| | |
|--|---|
| Назва дисципліни | Методи та технології аналізу даних |
| Адреса викладання дисципліни | Корпус факультету електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка, вул. Драгоманова 50, м. Львів, 79005, вул. Ген. Тарнавського 107, м. Львів, 79011 |
| Факультет та кафедра, за якою закріплена дисципліна | Факультет електроніки та комп'ютерних технологій Кафедра системного проєктування |
| Галузь знань, шифр та назва спеціальності | 12 Інформаційні технології 121 – Інженерія програмного забезпечення |
| Викладачі дисципліни | Юзевич Володимир Миколайович, докт. фіз.-мат. наук, професор, професор кафедри системного проєктування Огірко Ігор Васильович, докт. фіз.-мат. наук, професор, професор кафедри системного проєктування |
| Контактна інформація викладачів | volodymyr.yuzevych@lnu.edu.ua igor.ogirko@lnu.edu.ua |
| Консультації з питань навчання по дисципліні відбуваються | Консультації в день проведення лекційних занять/лабораторних занять (за попередньою домовленістю). Також можливі онлайн консультації. Для позгодження часу онлайн консультацій слід писати на електронну пошту викладача. |
| Сторінка дисципліни | https://moodle.elct.lnu.edu.ua/course/view.php?id=244 |
| Інформація про дисципліну | Дисципліна «Методи та технології аналізу даних» є вибірковою дисципліною зі спеціальності 121 Інженерія програмного забезпечення, ООП „Високопродуктивний комп'ютинг” яка викладається в 5 семестрі в обсязі 4,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS). |
| Коротка анотація дисципліни | Навчальну дисципліну розроблено таким чином, щоб надати студентам необхідні знання, обов'язкові для того, щоб розв'язувати типові задачі стосовно методів та технологій аналізу даних. У дисципліні представлено огляд основних підходів математичного моделювання та відповідних правил для ознайомлення студентів з новітніми інформаційними технологіями аналітичної обробки інформації; набуття практичних навичок використання методів і засобів інтелектуального аналізу даних. Також запропоновано методи побудови алгоритмів для розв'язування прикладних задач у контексті технології аналізу даних. |
| Мета та цілі дисципліни | Метою вивчення вибіркової дисципліни «Методи та технології аналізу даних» є ознайомлення студентів з інформацією про моделі, методи, технології та програмні засоби для розв'язування задач опрацювання даних з використанням підходів математичної статистики, а також з інформацією про алгоритми для розв'язування прикладних задач. <i>Цілі:</i> забезпечити знайомство студентів з загальною теорією методів аналізу даних; сформувати навички їх практичного використання; навчити студента вибирати і обґрунтувати технології аналізу даних, вказувати область їх застосування, записувати алгоритм розв'язування, реалізовувати його на персональному комп'ютері і аналізувати точність отриманого результату. |
| Література для | Основна література: |

| | |
|---|--|
| <p>вивчення дисципліни</p> | <ol style="list-style-type: none"> 1. Кучеров Д. П. Методи аналізу великих даних «Big Data». Київ. 2020. 237 с. 2. Ланде Д. В., Субач І. Ю., Гладун А. Я. Оброблення надвеликих масивів даних (Big Data) [Електронний ресурс]: навчальний посібник для використання у навчальному процесі з підготовки фахівців другого (магістерського) рівня вищої освіти зі спеціальності 122 «Комп'ютерні науки» / КПІ ім. Ігоря Сікорського. Електронні текстові дані. Київ: КПІ ім. Ігоря Сікорського, 2021. 168 с 3. Таран В.І., Гордієнко Ю.Г., Стіренко С.Г. Навчальний посібник з дисципліни “Технології Big Data” для студентів спеціальності 123 - “Комп'ютерна інженерія”. Київ: КПІ, 2022. 56 с. 4. Олещенко Л. М. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс]: навч. посіб. / КПІ ім. Ігоря Сікорського. Електронні текстові дані. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с. 5. Статистичні методи: Теорія оцінювання та статистичні гіпотези: Лабораторний практикум [Електронний ресурс]: навч. посіб. для студ. спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології» / КПІ ім. Ігоря Сікорського; уклад.: А. І. Жученко, Л. Д. Ярощук. Електронні текстові дані (1 файл: 1,14 Мбайт). Київ: КПІ ім. Ігоря Сікорського, 2020. 58 с. 6. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 2021. 92 с. ISBN 978-966-659-298-2. 7. Костюшко І. А., Любашенко Н. Д., Третиник В. В. <i>Методи обчислень</i>: Підручник. Київ: КПІ ім. Ігоря Сікорського, Вид. «Політехніка». 2021. 257 с. Інформаційні ресурси в Інтернеті 8. 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 9. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: https://www.netacad.com/courses/iot/big-data-analytics |
| <p>Обсяг курсу</p> | <p>Загальний обсяг: 135 годин – 4,5 кредитів. 64 години аудиторних занять. З них 32 години лекцій, 32 години лабораторних робіт та 71 година самостійної роботи</p> |
| <p>Очікувані результати навчання</p> | <p>Після завершення цього курсу студент буде:</p> <ul style="list-style-type: none"> - Знати основні поняття; проблеми курсу; типи моделей математичного моделювання; алгоритми розв'язування задач у контексті технології аналізу даних.; числові характеристики та довірчі інтервали для методів даних; класифікацію задач щодо опрацювання інформації про дані; алгоритми кластеризації даних. - Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності. - Знати спеціалізовані мови та технології програмування. - Вміти розв'язувати складні непередбачувані задачі і проблеми у спеціалізованих сферах професійної діяльності та/або навчання, що передбачають збирання, інтерпретацію та аналіз інформації (даних), |

вибір методів та інструментальних засобів, застосування інноваційних підходів щодо прийняття рішень.

- Вміти використовувати методи та засоби аналізу даних під час розв'язування прикладних проблем у спеціалізованих сферах професійної діяльності, в тому числі і з використанням відповідного програмного забезпечення у контексті технології аналізу даних.

Після вивчення даного курсу «Методи та технології аналізу даних» здобувачі набувають таких Загальних та Фахових компетентностей та Програмних результатів навчання:

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК04. Здатність спілкуватися іноземною мовою як усно, так і письмово.

ЗК05. Здатність вчитися і оволодівати сучасними знаннями.

ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК14. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

ФК16. Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.

ФК19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

ФК20. Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення.

ФК25. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.

ФК26. Здатність до алгоритмічного та логічного мислення.

ФК27. Здатність розробляти високопродуктивні програмні комплекси для вирішення задач наук про дані, систем штучного інтелекту, вбудованих та інших інноваційних систем.

ПРН01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

ПРН04. Знати і застосовувати професійні стандарти і інші нормативно-правові документи в галузі інженерії програмного забезпечення.

ПРН05. Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.

ПРН09. Знати та вміти використовувати методи та засоби збору, формулювання та аналізу вимог до програмного забезпечення.

ПРН10. Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування.

ПРН11. Вибирати вихідні дані для проектування, куруючись формальними методами опису вимог та моделювання.

ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.

ПРН14. Застосовувати на практиці інструментальні програмні засоби доменного аналізу, проектування, тестування, візуалізації, вимірювань та документування програмного забезпечення.

| | |
|---|--|
| | <p>ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.</p> <p>ПРН23. Вміти документувати та презентувати результати розробки програмного забезпечення.</p> <p>ПРН25. Вміти застосовувати інноваційні технологічні рішення при розробці високопродуктивних систем.</p> <p>ПРН27. Знати основи інженерії й аналізу даних та вміти вибрати оптимальні алгоритми і технології для розробки інноваційних рішень при розв'язанні задач наук про дані, вбудованих систем та систем штучного інтелекту.</p> |
| Ключові слова | Алгоритми аналізу даних, методи аналізу даних, технології аналізу даних, програмні засоби, . |
| Формат курсу | Очний |
| | Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем |
| Теми | Див. СХЕМА КУРСУ |
| Підсумковий контроль, форма | Залік в кінці семестру |
| Пререквізити | Для вивчення курсу студенти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Алгоритми та структури даних», «Чисельні методи», «Алгоритмізація і програмування», «Об'єктно-орієнтоване програмування». |
| Навчальні методи та техніки, які будуть використовуватися під час викладання курсу | Презентація, лекції, лабораторні роботи, обговорення, дискусія. Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація); дедуктивні методи на основі узагальнень; евристичні методи (проблемна лекція); інтерактивні методи (дискусія). |
| Необхідне обладнання | <p>Для проведення лекційних занять: комп'ютер (мінімальні характеристики: процесор Intel Core i3(4 ядра/8 потоків), 8ГБ оперативної пам'яті, 50ГБ вільного місця на диску, відеокарта Nvidia GEFORCE GT1030 2048MB), доступ до мережі Internet, засоби мультимедіа (в т. ч. проектор).</p> <p>Для проведення лабораторних занять: Комп'ютер (мінімальні характеристики: процесор Intel Core i3(4 ядра/8 потоків), 8ГБ оперативної пам'яті, 50ГБ вільного місця на диску, відеокарта Nvidia GEFORCE GT1030 2048MB), онлайн-доступ до кластера паралельних та розподілених обчислень та до GP GPU-сервера факультету. Необхідне програмне забезпечення включає в себе ОС Ubuntu 16 LTS, середовище розробки Geany/Code::Blocks, компілятор мови програмування C, бібліотеки OpenMP, MPI, набір програмного забезпечення CUDA Toolkit.</p> <p>Платформа Moodle.</p> |
| Критерії оцінювання (окремо для кожного виду навчальної діяльності) | <p>Оцінювання проводиться за 100-бальною шкалою.</p> <p>Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: до 25 балів за виконання модульного завдання – написання есе за обраною тематикою з переліку. • Змістовий модуль 2: до 25 балів за виконання модульного завдання – написання есе за обраною тематикою з переліку. • Виконання лабораторних робіт: до 40 балів. <p>Виконання залікової роботи: до 50 балів.</p> <p>Підсумкова максимальна кількість балів 100.</p> <p>Звітність за курс – залік у кінці семестру.</p> <p>Змістовий модуль – самостійна робота студента оформлена у</p> |

вигляді есе – робота друкованим текстом, рекомендованим обсягом до 10 сторінок (шрифт Times New Roman, 14). Есе включає в себе детальний розгляд обраної індивідуальної теми, приведення прикладів та лістингів коду програм, огляду технологій, літератури. Код програм повинен бути обов'язково прокоментований та пояснений, необхідно також продемонструвати його роботу у разі, якщо в якості прикладу наводяться не окремі елементи технології, а суцільна програма. Есе також повинно містити посилання на літературні джерела/інтернет ресурси, що були використані під час його написання. Фінальна версія есе здається студентом у електронному форматі .pdf викладачу для оцінки. Теми для змістових модулів див. у розділі **Питання до модульного контролю**.

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.

Відвідування занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та 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 за виконання лабораторних робіт. 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 лабораторних робіт (5 лабораторних робіт (16 завдань), максимальна кількість балів: до 50) відбувається шляхом оцінки роботи студента під час проведення лабораторної роботи в лабораторії (0-5 балів за один варіант роботи) та захисту звіту по виконаній лабораторній роботі (0-5 балів за один варіант роботи). У підсумку, всі набрані бали множаться на коефіцієнт (0,25) для переведення у 40 бальну шкалу.

Бали оцінювання лабораторних робіт нараховуються за наступним співвідношенням:

5 – студент у повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;

4 – студент достатньо добре розуміє розглянутий матеріал та принципи

написаного ним коду програми, присутні неточності та незначні помилки у відповідях на запитання по темі, код програми функціонує відповідно до завдання (або з несуттєвими недоліками);

3 – студент не досить добре розуміє розглянутий матеріал та написаний ним код програми, вагається та надає неточні/не конкретні відповіді на запитання по темі, код програми функціонує неточно, або з помірними недоліками;

2 – студент погано розуміє розглянутий матеріал та написаний ним код програми, студент в більшості надає помилкові відповіді на питання по темі, код програми функціонує з суттєвими недоліками;

1 - студент погано розуміє розглянутий матеріал та написаний ним код програми, код програми не функціонує належним чином;

0 - студент зовсім не засвоїв розглянутий матеріал, написаний ним код програми не відповідає темі/не функціонує взагалі.

Оцінювання змістових модулів (2 змістових модулі, 25 балів за кожний) відбувається шляхом написання студентом есе, тесту, написання програми, тощо.

Бали оцінювання змістових модулів нараховуються за наступним співвідношенням:

25-20 - розглянута тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно, які містять аналіз і систематизацію, аргументовані висновки. Засвідчено глибоке володіння матеріалом. Наведені приклади коду повністю робочі та відповідають темі. Можуть бути присутні несуттєві помилки та невідповідності;

20-15 - відтворюється значна частина розглянутої теми. Виявлено знання і розуміння основних положень навчальної дисципліни, проте присутні неточності та/або невідповідності основній темі. Наведені приклади коду частково робочі, проте в загальному відповідають темі;

15-10 - відстежується загальне розуміння розглянутої теми. Виявлені множинні неточності та невідповідності, пояснення наведеного коду відсутні, код функціонує із значними неточностями (або відсутні приклади запуску коду на виконання взагалі);

10-5 – студент погано розуміє розглянуту тему. Виявлені суттєві неточності та невідповідності. Наведені приклади коду з суттєвими недоліками, або не відповідають темі;

5 – 0 – студент взагалі не розуміє розглянуту тему. Тему не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи. Наведений код не робочий, або відсутній як такий.

У підсумку, всі набрані бали двох модулів множаться на коефіцієнт (0.2) для переведення у 10 бальну шкалу.

Критерії отримання додаткових балів:

Нарахування додаткових балів відбувається за написання тез доповідей, наукових статей, участь у діяльності наукового гуртка, участь у наукових семінарах та круглих столах, конкурсах, участь у заходах неформальної освіти за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах на провідних ІТ компаніях за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни.

Питання до заліку чи

Методи аналізу даних.

| | |
|--------------------------|---|
| <p>екзамену</p> | <p>Застосування аналізу даних. Етапи аналізу даних. Технології аналізу даних. Терміни аналізу даних. Аналіз даних у сфері HR. Фактичний аналіз даних. Інтерпретація результатів аналізу даних. Пріоритети вимірювань у контексті аналізу даних. Відбір даних. Вибіркове спостереження. Непевність оцінок даних. Джерела непевності результату оцінювання даних. Класифікація складових непевності та відповідні означення. Типи даних. Рівні виміру, матриця даних. Частотні таблиці, центральна тенденція оцінювання даних. Діаграми даних. Асоціації в контексті аналізу даних. Структури для роботи з великими обсягами даних в Python. Інформація про дані в середовищі R. Характерні особливості даних у R та Python. Фактори, коди, списки і таблиці даних у середовищах R та Python. Функції та вектори для аналізу даних. Векторизовані обчислення в R та аналіз даних. Інтелектуальний аналіз даних. Підготовчі етапи процесу Data Mining. Дублювання даних. Очищення даних. Етапи очищення даних. Оцінювання даних з допомогою гістограм. Приклади використання гістограм. Кластеризація даних. Таксономія і аналіз даних. Агломеративні методи для аналізу даних.. Способи побудови (англ. <i>sorting strategies</i>) дендрограм. Дослідження набору даних методами кластеризації. Алгоритми кластеризації. Іриси Андерсона (Фішера) та їх аналіз. Оцінка якості чіткої кластеризації. Критерії оцінки якості кластеризації. Нормалізована статистика Хаббарда. Індекс Данна в контексті аналізу даних. Індекс оцінки силуета в контексті аналіз даних.. Панельні дані і їх переваги. Аналіз панельних даних. Кореляційний та регресійний аналіз. Множинна регресія. Фрустрація. Часові ряди: означення, типологія, мотивація досліджень, класифікація, тренди. Моделі ковзного середнього (Moving Average (MA)). Модель стаціонарного часового ряду на основі авторегресії (AR).</p> |
| <p>Опитування</p> | <p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p> |

**Схема курсу “Методи та технології аналізу даних”
для студентів спеціальності 121 – Інженерія програмного забезпечення**

| Тиж. | Тема, план, короткі тези | Форма діяльності (заняття) лекція, самостійна, дискусія, групова робота) | Література. Ресурси в інтернеті | Завдання, год | Термін виконання |
|------|--|--|---------------------------------|---------------|------------------------|
| 1 | Вступ. Дані. Методи аналізу даних. Мета, завдання і предмет дисципліни. Застосування аналізу даних. Етапи. | Лекція | [1], [2], [3], [5], Сайт курсу | 2 | кінець поточного тижня |
| 1 | Підготовка до роботи з даними. Налаштування робочого середовища. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 2 | Терміни. Аналіз даних у сфері HR. Фактичний аналіз даних. Інтерпретація результатів. Пріоритети вимірювань. | Лекція | [1], [2], [3], [6], Сайт курсу | 2 | кінець поточного тижня |
| 2 | Лаб. 2. Обмін результатами та початковий аналіз даних. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 3 | Відбір даних. Вибіркове спостереження. Непевність. Джерела непевності результату. Класифікація складових непевності та їхні означення. | Лекція | [1], [2], [5], Сайт курсу | 2 | кінець поточного тижня |
| 3 | Лаб. 2b. Реєстрація на сайті GitHub, створення репозиторію. Створення веб-додатку із використанням модуля Srupe, який дозволяє обрати часовий ряд типу VCI, TCI, VNI. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 4 | Типи даних. Рівні виміру, матриця даних, частотні таблиці, центральна тенденція. Діаграми. Асоціації. | Лекція | [1], [2], [5], Сайт курсу | 2 | кінець поточного тижня |
| 4 | Лаб. 3a. Структури для роботи з великими обсягами даних в Python | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 5 | Інформація про дані в середовищі R. Характерні особливості R та Python. | Лекція | [1], [2], [3], Сайт курсу | 2 | кінець поточного тижня |
| 5 | Лаб. 3b. Використання як <code>numpy</code> <code>array</code> , так і <code>dataframe</code> , аналізування часових витрат на виконання процедур (профілювання часу виконання). | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 6 | Фактори, коди, списки і таблиці даних у середовищах R та Python. | Лекція | [1], [2], [3], Сайт курсу | 2 | кінець поточного тижня |
| 6 | Лаб. 3c. Опис набору даних (відомостей) щодо основних витрат електричної енергії домогосподарств. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 7 | Функції та вектори для аналізу даних. Векторизовані обчислення в R. | Лекція | [1], [2], [3], [5], Сайт курсу | 2 | кінець поточного тижня |
| 7 | Лаб. 3d. Опис набору даних щодо витрат домогосподарств, у яких загальна активна споживана потужність перевищує 5 кВт. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 8 | Інтелектуальний аналіз даних. | Лекція | [1], [2], [4], [5], | 2 | кінець |

| | | | | | |
|----|--|--------------------|-----------------------------------|---|------------------------|
| | Підготовчі етапи процесу Data Mining. Дублювання даних. Очищення даних. Етапи очищення даних. | | Сайт курсу | | поточного тижня |
| 8 | Лаб. 3е. Опис набору даних щодо витрат домогосподарств, у яких вольтаж перевищує 235 В. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 9 | Оцінювання даних з допомогою гістограм. Приклади використання гістограм. | Лекція | [1], [2], [3], [5], Сайт курсу | 2 | Кінець поточного тижня |
| 9 | Лаб. 3ф. Опис набору даних щодо витрат домогосподарств і обчислення середніх величин для різних груп споживання електричної енергії. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 10 | Кластеризація даних. Таксономія. Агломеративні методи. Способи побудови (англ. <i>sorting strategies</i>) дендрограм. | Лекція | [1], [2], [3], Сайт курсу | 2 | кінець поточного тижня |
| 10 | Лаб. 4а. Лінійна регресія даних Інтернет-трафіку. 1 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 11 | Дослідження набору даних методами кластеризації. Алгоритми. Іриси Андерсона (Фішера). | Лекція | [1], [2], [5], Сайт курсу | 2 | кінець поточного тижня |
| 11 | Лаб. 4б. Лінійна регресія даних Інтернет-трафіку. 2 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 12 | Оцінка якості чіткої кластеризації. Критерії оцінки якості. Нормалізована статистика Хаббарда. Індекс Данна. Індекс оцінки силуета. | Лекція | [1], [2], [4], [5], Сайт курсу | 2 | кінець поточного тижня |
| 12 | Лаб. 4с. Лінійна регресія даних Інтернет-трафіку. 3 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 13 | Панельні дані і їх переваги. Аналіз панельних даних. | Лекція | [1], [2], [4], [7], Сайт курсу | 2 | кінець поточного тижня |
| 13 | Лаб. 5а. Виявлення аномалій Інтернет-трафіку. 1 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 14 | Кореляційний та регресійний аналіз. Множинна регресія. Фрустрація. | Лекція | [1], [2], [4], [6], Сайт курсу | 2 | кінець поточного тижня |
| 14 | Лаб. 5б. Виявлення аномалій Інтернет-трафіку. 2 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 15 | Часові ряди: означення, типологія, мотивація досліджень, класифікація, тренди | Лекція | [1], [2], [4], [7] Сайт курсу | 2 | кінець поточного тижня |
| 15 | Лаб. 5с. Виявлення аномалій Інтернет-трафіку. 3 част. | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |
| 16 | Моделі ковзного середнього (Moving Average (MA)). Модель стаціонарного часового ряду на основі авторегресії (AR). | Лекція | [4], [7], Сайт курсу | 2 | кінець поточного тижня |
| 16 | Підсумкове заняття | Лабораторна робота | Сайт курсу | 2 | кінець поточного тижня |