

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 серпня 2023 р.)

Завідувач кафедри _____



Роман ШУВАР

Силабус з навчальної дисципліни
«Засоби та технології аналітики даних»,
що викладається в межах ОП «Комп'ютерні науки»
другого (магістерського) рівня вищої освіти
для здобувачів зі спеціальності
122 – Комп'ютерні науки

Львів 2023

Назва дисципліни	Засоби та технології аналітики даних
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра системного проектування
Галузь знань, шифр та назва спеціальності	12 Інформаційні технології, 122 Комп'ютерні науки та інформаційні технології
Викладачі дисципліни	Огірко Ігор Васильович, докт. фіз.-мат. наук, професор, професор
Контактна інформація викладачів	igor.ogirko@lnu.edu.ua https://electronics.lnu.edu.ua/employee/ohirko-i-v
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю): кімн. 317, корпус факультету електроніки та комп'ютерних технологій, м. Львів, вул. Драгоманова, 50. Також можливі он-лайн консультації через MS Teams або систему електронного навчання Moodle. Для погодження часу онлайн консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://moodle.elct.lnu.edu.ua/enrol/index.php?id=246 https://electronics.lnu.edu.ua/course/informatsijna-arhitektura
Інформація про дисципліну	Дисципліна «Засоби та технології аналітики даних» є вибірковою дисципліною з спеціальності 122 – Комп'ютерні науки для освітньої програми «Комп'ютерні науки», яка викладається в 1 семестрі в обсязі 7 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено таким чином, щоб надати учасникам необхідні знання та практичні навички щодо аналітичної роботи з реальними даними соціальної, економічної природи, тощо. Типові задачі щодо лінійних та нелінійних методів опису даних, прогнозування стаціонарних на нестаціонарних часових рядів, виділення з останніх типових значущих компонентів розглядатиметься у чіткому поєднанні з реальними задачами з аналітики даних та типовими прикладами.
Мета та цілі дисципліни	Метою і цілями викладання навчальної дисципліни є формування системи фундаментальних знань щодо аналізу даних статистичних показників соціальних, економічних та поведінкових процесів, а, також, підходи до прогнозування часових рядів вищезначених процесів

<p>Література для вивчення дисципліни</p>	<p>Основна література:</p> <ol style="list-style-type: none"> 1. Р.Є. Майборода, О.В. Сугакова АНАЛІЗ ДАНИХ ЗА ДОПОМОГОЮ ПАКЕТА R : Навчальний посібник: К.: КНУ імені Т. Шевченко, 2015. – 65 с. 2. https://cran.r-project.org/doc/manuals/r-release/R-intro.html 3. Прикладна економетрика : навч. посіб. : у двох частинах. / Л. С. Гур'янова, Т. С. Клебанова, С. В. Прокопович та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2016. 4. Gartner Identifies the Top 10 Data and Analytics Trends. URL: https://www.gartner.com/en/newsroom/press-releases/2023-05-09-gartner-identifies-the-top-ten-data-and-analytics-trends-for-20230 5. Walter Enders. Applied Econometric Time Series. Wiley, 2-d edition, 2004. – 460 p. 6. Diggle P.J. Time Series: A Biostatistical Introduction. – Oxford, 1990. 7. Mingda Zhang Time Series: Autoregressive models AR, MA, ARMA, ARIMA http://people.cs.pitt.edu/~milos/courses/cs3750/lectures/class16.pdf 8. Ibid: Markov models: http://people.cs.pitt.edu/~milos/courses/cs3750/lectures/class15.pdf
--	---

<p>Обсяг курсу</p>	<p>210 годин занять. З них 32 години лекцій, 48 годин лабораторних робіт та 130 годин самостійної роботи.</p>
<p>Очікувані результати навчання</p>	<p>Після завершення цього курсу студент буде:</p> <p>Знати:</p> <ul style="list-style-type: none"> • <i>основи збору даних, підходи до обробки останніх ;</i> • адекватні моделі для аналізу соціально-економічних трендів, • алгоритми , економіко-математичні методи та моделі для соціально-економічних та демографічних трендів. <p>Вміти:</p> <ul style="list-style-type: none"> • вільно володіти мовою програмування R для статистичних обрахунків, візуалізації даних. • реалізовувати адаптивні методи для часової динаміки реальних соціально-економічних явищ; • розраховувати чисельні характеристики часових рядів, програмно знаходити та уключати аномальні дані часових рядів ; • надавати оціночні рішення в економіко-математичних моделях. <p>Після вивчення курсу здобувачі набудуть таких компетентностей і програмних результатів:</p> <p>ЗК5. Здатність вчитися й оволодівати сучасними знаннями.</p> <p>СК1. Усвідомлення теоретичних засад комп'ютерних наук.</p> <p>СК2. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.</p> <p>СК5. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних і комп'ютерних систем різного призначення.</p> <p>СК6. Здатність застосовувати існуючі та розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.</p> <p>СК7. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог із урахуванням наявних ресурсів і обмежень.</p> <p>СК8. Здатність розробляти та реалізовувати проєкти зі створення програмного забезпечення, у т. ч. в непередбачуваних умовах, за нечітких вимог і необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проєктом.</p> <p>СК9. Здатність розробляти та адмініструвати бази даних і знань.</p>

	<p>СК10. Здатність оцінювати та забезпечувати якість ІТ-проектів, інформаційних і комп'ютерних систем різного призначення, застосовувати міжнародні стандарти оцінки якості програмного забезпечення інформаційних і комп'ютерних систем, моделі оцінки зрілості процесів розробки інформаційних і комп'ютерних систем.</p> <p>СК13. Здатність застосовувати методи і підходи штучного інтелекту, інтелектуального аналізу та науки про дані та підходів оптимізації до розв'язання конкретних проблем комп'ютерних наук.</p> <p>РН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.</p> <p>РН2. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.</p> <p>РН4. Управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів.</p> <p>РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.</p> <p>РН8. Розробляти математичні моделі та методи аналізу даних (включно з великими).</p> <p>РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).</p> <p>РН10. Проектувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.</p> <p>РН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.</p> <p>РН12. Проектувати та супроводжувати бази даних та знань.</p> <p>РН16. Виконувати дослідження у сфері комп'ютерних наук.</p> <p>РН20. Володіти методами та засобами штучного інтелекту, інженерії та аналізу даних, розпізнавання образів і адаптивного опрацювання інформації, аналізу та обробки природної мови, моделювання та оптимізації.</p> <p>РН21. Створювати нові системи даних, високорівневі вбудовані системи, спеціалізовані комп'ютерні системи та інтелектуальні системи із застосуванням базових знань апаратного і програмного забезпечення мікроконтролерів і мікрокомп'ютерів.</p>
--	--

Ключові слова	Мова R, часові ряди, збір даних, прогнозна аналітика.
Формат курсу	Очний
	Проведення лекцій, практичних робіт та консультації для кращого розуміння тем.
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Екзамен в кінці семестру

Пререквізити	Для вивчення курсу студенти потребують базових знань у галузі 12 – Інформаційні технології, зокрема з таких предметів: вища математика, дискретна математика, теорія алгоритмів, алгоритмізація і програмування, об'єктно-орієнтоване програмування..
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, практичні роботи, обговорення, дискусія.
Необхідне обладнання	Мультимедіа, платформа Moodle, комп'ютерне програмне забезпечення: мова програмування Python, open-source веб-інтерактивне обчислювальне середовище Jupyter, Community Edition PyCharm, середовище RStudio для програмування мовою R
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою. Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • практичні роботи: 50% семестрової оцінки; максимальна кількість балів $8 \times 6 + 2 = 50$ (8 практичних робіт, де 2 додаткові бали – за якісне виконання всіх робіт). • контрольні заміри (2 модулі): 50% семестрової оцінки; максимальна кількість балів $2 \times 25 = 50$. <p>Загалом упродовж семестру 100 балів.</p> <p>Оцінки за практичні заняття розподіляються наступним чином: виконання практичних завдань – 60%, відповіді на запитання викладача по темі заняття – 40%.</p> <p>Контрольні заміри проводяться у формі тестових завдань по 25 балів за кожний модуль.</p> <p>Академічна доброчесність: Очікується, що практичні та контрольні роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і практичні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися усіх термінів визначених для виконання усіх видів робіт, передбачених курсом.</p> <p>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.</p> <p>Політика виставлення балів. Враховуються бали набрані на поточному тестуванні, самостійній роботі 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</p>

	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. Жодні форми порушення академічної доброчесності не толеруються.
Питання до контрольних робіт	Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-сторінці курсу (система електронного навчання Moodle: https://moodle.elct.lnu.edu.ua/course/view.php?id=246).
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання, лабораторна робота, самостійна робота, год.	Термін виконання
1	Основи синтаксису R: а саме змінні, вектори, матриці, фрейми даних, елементарні дії з ними, функції, стандартні бібліотеки та модулі. Дані, цикли та умови в R. Статистичні функції в R візуалізація, генератори розподілів тощо.	Лекція	1, 2,	Вступне заняття. Академічна доброчесність. Цикли та умови в R, 4	3 тиж. семестру
2	Датасети R та методики збору даних для задач аналітики соціальних та економічних явищ. Алгоритми кластеризації використовуючи функціонал R та оцінюватиме збіжність (якість) для методів із заданою кількістю кластерів: k-means, k-medians.	Лекція	1, 2	Реалізація алгоритмів кластеризації в R, 4	5 тиж. семестру
3	R для оцінки кореляції даних (cor), регресії 1-го порядку (lm) та підводитиме до застосунків регресій 2-го та вищих порядків (сімейство функцій nls), множинних регресій. Застосування для завдань аналітики даних.	Лекція	1, 2, 3, 4	Кореляція та лінійна регресія в R, 4	7 тиж. семестру
4	Нелінійні функції для застосунків регресій. Природничі та суспільні реальні дані: динаміка населення, ціноутворення товарів, ресурсів, обіг рітейлу інше.	Лекція	1, 2, 3, 4	Нелінійна регресія в R, 4	8 тиж. семестру
5	Понятійний апарат часових рядів (ЧР). Основні складові останніх, як то: тренди, сезонність, циклічність, нерегулярні вклади в ЧР, елементи аналізу ЧР, інформативні функції для ЧР, означення сильної та слабкої стаціонарності ЧР. Сімейство методів для моделювання часових рядів ARIMA, що враховують фактори,	Лекція	3-8	Авторегресія в R, 4	10 тиж. семестру

	викиди і структурні зміни часового ряду. Методи сімейства, їх можливості та недоліки. Однопараметричний AR метод: задачі, збіжність, оцінки, R коди.				
6	Moving average (MA метод): означення, задачі, збіжність, аналогії із AR. Оцінюватимуться R функції для MA із стандартної та альтернативних бібліотек R.	Лекція	3-8	Moving Average (MA) в R, 4	11 тиж. семестру
7	Модель ARMA: параметризація, детрендинг, диференціювання тощо. Це дозволить логічно ввести методи для нестационарних рядів- ARIMA.	Лекція	3-8	Моделі ARMA в R, 4	12 тиж. семестру
8	Нестационарна ARIMA модель (акронімом Auto Regressive Integrated Moving Average). Встановлюватимуться особливості методу та його застосування до конкретної задачі аналітики. Стратегія вибору для моделі ARIMA кількості диференціювань d , значень для авторегресії та ковзного середнього p і q , додаткові значення для сезонності, ілюструватимуться реальні задачі	Лекція	3-8	Реалізація ARIMA в R, 4	13-14 тиж. семестру
9	Альтернативні методи для аналітики часових рядів, як то Марківські ланцюжки. Розглядатимуться основні означення, властивості та задачі аналітики, що вивіщуються використовуючи Марківські ланцюжки. Ймовірнісний підхід до прогнозування часових рядів, а саме - за допомогою динамічних Баєсівських мереж. Останні дозволитимуть прогнозувати стани багатьох змінних використовуючи апіорні дані до обчислення апостеріорних ймовірностей станів.	Лекція	3-8	Написання та захисти есеїв, що стосуватимуться альтернативних (до тих, що описувалися в лекціях курсу) методів прогнозування аналітики. Захист есе дозволить набрати додаткові (до 10) бали поточної успішності. Вибір тематики доповіді обмежуватиметься лише прогнозування аналітикою, 4	16 тиж. семестру