

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
“ Сучасні методи аналізу даних ”,
що викладається в межах ОПШ
“ Статистичний аналіз даних ”
другого (магістерського) рівня вищої освіти для здобувачів з
спеціальності 112 - Статистика

Львів 2023 р.

Назва дисципліни	Сучасні методи аналізу даних
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 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	Головатий Юрій Данилович, доктор фізико-математичних наук, доцент, професор кафедри математичної статистики і диференціальних рівнянь; Доманська Олена Вікторівна, кандидат фізико-математичних наук, доцент кафедри математичної статистики і диференціальних рівнянь
Контактна інформація викладачів	olena.domanska@lnu.edu.ua , https://new.mmf.lnu.edu.ua/employee/domanska_o_v ; Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю): ауд. 267, головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Сторінка курсу	https://new.mmf.lnu.edu.ua/course/smad-112-mag
Інформація про дисципліну	Дисципліна “Сучасні методи аналізу даних” є нормативною дисципліною з спеціальності 112-Статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 1-му семестрі в обсязі 6-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Курс розроблено для ознайомлення студентів з сучасними методами аналізу даних, використовуючи Python, Pandas, Numpy та машинне навчання. Студенти оволодіють навиками розробки практичних програм для вирішення ряду конкретних проблем, зокрема, у фінансовій галузі
Мета та цілі дисципліни	<i>Мета:</i> ознайомити студентів з основними підходами до сучасних методів аналізу даних <i>Цілі:</i> ознайомити студентів з підходами до використання Python для ефективного аналізу даних, новітніми пакетами Python для різностороннього аналізу даних з різних джерел (CSV, Excel, баз даних, веб тощо). Створення програм для вирішення проблем у фінансовій галузі, зокрема, розподіл активів, аналіз ризиків, інвестиційна ефективність, тощо
Література для вивчення дисципліни	Рекомендована література 1) Abdullah Karasan. Machine Learning for Financial Risk Management with Python. Algorithms for modelling Risk. O’Reilly, 2022. 2) Yves Hilpisch. Python for Finance. Mastering Data-driven Finance. O’Reilly, 2019. 3) James Ma Weiming. Mastering Python for Finance. Packt, 2015.

	<p><i>Додаткова література та інтернет-ресурси</i></p> <p>4) Онлайн курс “Complete Python and Machine Learning in Financial Analysis” на Udemy, https://www.udemy.com/course/python-and-machine-learning-in-financial-analysis.</p>
Обсяг курсу	Загальний обсяг: 180 годин. Аудиторних занять: 64 год., з них 32 год. лекційних та 32 год. лабораторних занять. Самостійної роботи: 116 год.
Очікувані результати навчання	<p>У результаті вивчення даного курсу студент повинен:</p> <p>знати: Сучасні підходи до аналізу даних, як застосовувати машинне навчання та глибоке навчання до аналізу даних, зокрема, у фінансовій галузі. Які є основні проблеми при роботі з фінансовими ризиками, як кількісно збалансувати винагороду та ризик;</p> <p>вміти: використовувати сучасні інструменти та технології для аналізу даних, такі як Python, Pandas, Numpy, машинне навчання та глибоке навчання.</p> <p>В результаті засвоєння матеріалу даного курсу студент набуде таких загальних (ЗК) і спеціальних (СК) компетентностей:</p> <p>ЗК-2. Здатність застосувати знання у практичних ситуаціях, генерувати нові ідеї, розробляти проекти та управляти ними.</p> <p>ЗК-5. Здатність до пошуку, обробки та аналізу інформації з різноманітних джерел.</p> <p>СК-2. Здатність будувати математичні моделі реальних явищ і процесів в різних предметних галузях, досліджувати їх засобами математики та статистики.</p> <p>СК-4. Здатність використовувати сучасні інформаційні технології та комп’ютерні методи статистичних досліджень.</p> <p>СК-6. Володіти специфічними статистичними методами для візуалізації та аналізу великих даних, а також даних в мережах складної структури.</p> <p>і здобуде такі програмні результати навчання (РН):</p> <p>РН-2. Володіти знаннями й розумінням основних засад теоретичної і прикладної статистики.</p> <p>РН-3. Застосовувати методи теорії ймовірностей, математичної статистики і теорії випадкових процесів до дослідження випадкових явищ.</p> <p>РН-5. Володіти базовими знаннями в галузі інформатики й сучасних інформаційних технологій, комп’ютерними інструментами аналізу даних, прогнозування і прийняття рішень.</p> <p>РН-7. Володіти навичками збору та зберігання інформації, а також її захисту з урахуванням аспектів соціальної та етичної відповідальності.</p> <p>РН-8. Знати методологію та організацію наукових досліджень в галузі статистики та аналізу даних.</p> <p>РН-10. Володіти базовими знаннями в галузі економічних та соціальних наук.</p>
Ключові слова	Аналіз даних, Python, Pandas, Numpy, машинне навчання, глибоке навчання, управління ризиками
Формат курсу	Очний
Теми	<p>Тема 1. Технічний аналіз в Python</p> <p>Тема 2. ООП з Python</p> <p>Тема 3. Аналіз даних з Pandas</p>

	<p>Тема 4. Вирішення проблем з якістю даних за допомогою Pandas</p> <p>Тема 5. Обробка великих масивів даних за допомогою Numpy</p> <p>Тема 6. Фінансові дані та робота з ними</p> <p>Тема 7. Моделювання часових рядів</p> <p>Тема 8. Вступ до багато-факторних моделей</p> <p>Тема 9. Моделювання волатильності ринку використовуючи GARCH моделі</p> <p>Тема 10. Симуляції Монте-Карло для фінансів</p> <p>Тема 11. Оцінка ризиків (Value at Risk) використовуючи метод Монте Карло</p> <p>Тема 12. Розподіл активів використовуючи Python</p> <p>Тема 13. Передбачення кредитного дефолту за допомогою машинного навчання</p> <p>Тема 14. Моделі машинного навчання для фінансів</p> <p>Тема 15. Вступ до глибокого навчання</p> <p>Тема 16. Глибоке навчання для фінансів</p>
Підсумковий контроль, форма	Іспит у кінці семестру
Пререквізити	Для вивчення даного курсу студентам потрібні базові знання з: <ul style="list-style-type: none"> - програмування; - машинного навчання.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	<p>Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація);</p> <p>дедуктивні методи на основі узагальнень;</p> <p>евристичні методи (проблемна лекція);</p> <p>інтерактивні методи (дискусія).</p> <p>Створення команди курсу в месенджерах для призначення завдань, консультацій і спілкування зі студентами. Створення презентацій лекційних та практичних занять, запис відео лекцій та практичних. Створення бази завдань та проведення тестувань і контрольних робіт на платформі Kahoot та Kaggle.</p>
Необхідне обладнання	Комп'ютер із необхідним програмним забезпеченням, доступ до Internet мережі.
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: 10% семестрової оцінки за контрольну роботу виконання домашніх завдань, максимальна кількість балів 10. • Змістовий модуль 2: 10% семестрової оцінки за контрольну роботу та виконання домашніх завдань, максимальна кількість балів 10. • Змістовий модуль 3: 10% семестрової оцінки за контрольну роботу та виконання домашніх завдань, максимальна кількість балів 10. • Змістовий модуль 4: 10% семестрової оцінки за контрольну роботу та виконання домашніх завдань, 10% семестрової оцінки за проект, максимальна кількість балів 20. • іспит: 50% семестрової оцінки, максимальна кількість балів 50. <p>Підсумкова максимальна кількість балів 100.</p> <p>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</p>

	<p>викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні 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.</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 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. Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до заліку чи екзамену.	Для проведення тестування та іспиту створено банк питань в середовищі Kahoot, а також розроблені практичні завдання
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

**Схема курсу “ Сучасні методи аналізу даних ”
для студентів спеціальності 112 – Статистика**

Тижні	Лекційний курс		Практичні заняття	
	Назва теми	К-сть год	Назва теми	К-сть год
1	2	3	4	5
1	Технічний аналіз в Python	2	Типи та структури даних в Python. Списки, кортежі, словники, набори. Функціональне програмування	2
2	ООП з Python	2	Об'єкти в Python. Основи класів. Поняття abstractions, inheritance, polymorphism and encapsulation.	2
3	Аналіз даних з Pandas	2	Series та Data Frames як основні об'єкти в Pandas. Базова аналітика з Pandas. Візуалізація даних.	2

4	Вирішення проблем з якістю даних за допомогою Pandas	2	Очистка даних від nan. Виявлення аутлаєрів. Операції групування та сортування. Об'єднання кількох дата фреймів.	2
5	Обробка великих масивів даних за допомогою Numpy	2	Основи Numpy. Масиви з багатьма вимірами. Створення та індексування масивів. Базові математичні операції. Бродкастинг.	2
6	Фінансові дані та робота з ними	2	Фінансові дані. Завантаження даних з різноманітних джерел. Статистичний аналіз. Кореляційний аналіз.	2
7	Моделювання часових рядів	2	Вступ до часових рядів. Декомпозиція часових рядів. Перевірка на стаціонарність. Моделювання часових рядів методами ARIMA.	2
8	Вступ до багато-факторних моделей	2	Імплементація CAPM моделі в Python. Побудова трьох-факторної моделі для аналізу портфоліо активів.	2
9	Моделювання volatility ринку використовуючи GARCH моделі	2	Обґрунтування волатильності прибутковості акцій за допомогою моделей ARCH та GARCH. Побудова CCC-GARCH моделі.	2
10	Симуляції Монте-Карло для фінансів	2	Моделювання динаміки курсу акцій за допомогою геометричного броунівського руху.	2
11	Оцінка ризиків (Value at Risk) використовуючи метод Монте Карло	2	Ціноутворення на опціони з використанням моделі Монте Карло. Оцінка ризику за методом Монте Карло.	2
12	Розподіл активів використовуючи Пайтон	2	Оцінка ефективності базового портфеля. Пошук ефективного порогу за допомогою оптимізації використовуючи scipy.	2

13	Передбачення кредитного дефолту за допомогою машинного навчання	2	Завантаження даних, аналіз даних, робота з відсутніми даними та з аутлаєрами. Побудова дерева рішень. Тренування моделі з іншими гіперпараметрами.	2
14	Моделі машинного навчання для фінансів	2	Класифікатори для передбачення фінансових показників. Дерева рішень, бустингові моделі. Робота з незбалансованими даними.	2
15	Вступ до глибокого навчання	2	Архітектура нейронних мереж. Імплементация нейронних мереж за допомогою PyTorch.	2
16	Глибоке навчання для фінансів	2	Глибоке навчання для табличних даних. Конволюційні та рекурентні нейронні мережі для передбачення часових рядів.	2
	Разом	32		32
	Викладач: Доманська О.В.		Викладач: Доманська О.В.	