

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра теорії оптимальних процесів

Затверджено

На засіданні
кафедри теорії оптимальних процесів
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № ____ від _____ 2020__ р.)

Завідувач кафедри Шахно С.М.

Силабус з навчальної дисципліни
“Інструментальні засоби аналізу даних”,
що викладається в межах ОПП Системний аналіз
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 124 – системний аналіз

Львів 2020 р.

Назва дисципліни	Інструментальні засоби аналізу даних
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Факультет прикладної математики та інформатики Кафедра теорії оптимальних процесів
Галузь знань, шифр та назва спеціальності	12 – інформаційні технології 124 – системний аналіз
Викладачі дисципліни	Гнатишин Олександра Павлівна, старший викладач кафедри теорії оптимальних процесів
Контактна інформація викладачів	oleksandra.hnatyshyn@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/hnatyshyn ; Головний корпус ЛНУ ім. І. Франка, каб. 269. м. Львів, вул. Університетська, 1
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю).
Сторінка курсу	
Інформація про дисципліну	Дисципліна “Інструментальні засоби аналізу даних” є дисципліною вільного вибору студента зі спеціальності 124 – системний аналіз для освітньої програми Системний аналіз, яка викладається в 7-му семестрі в обсязі 5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS.).
Коротка анотація дисципліни	Суть аналізу даних полягає в здобутті знання про об’єкт дослідження — виявлення корисної інформації, формулювання висновків, зважене прийняття рішень. Аналіз даних може мати багато аспектів та підходів, реалізовуватися за допомогою різних інструментів — в тому числі математичних, статистичних, за допомогою різноманітних способів візуалізації. Акцент в курсі зроблено на задачах класифікації, кластеризації, прогнозуванні та реалізації поставлених задач інструментальними засобами (R, Python, ...).
Мета та цілі дисципліни	<i>Метою</i> дисципліни “Інструментальні засоби аналізу даних” є формування базового уявлення, первинних знань, вмінь та навичок з аналізу даних як наукової та прикладної дисципліни. Виробити навички виявлення нових реальних неочевидних нетривіальних закономірностей і систематичних взаємозв’язків між змінними, які згодом можна буде застосувати до нових наборів даних. Освоєння студентами сучасних інструментальних засобів для аналізу даних. <i>Завданням курсу є</i> :ознайомити студентів з сучасними технологіями аналізу даних та програмними засобами, орієнтованими на реалізацію цих задач. Навчити студентів використовувати інструментальні засоби (R, Python, ...) для аналізу даних.
Література для вивчення дисципліни	1. <i>Gareth James</i> . An Introduction to Statistical Learning with Application in R// <i>Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani</i> . Springer, 2017. 2. <i>Matthew Mayo</i> Top Python Libraries for Data Science, Data Visualization &

	<p>Machine Learning//<i>Matthew Mayo</i>. KDnuggets, 2020</p> <p>3. <i>Afshine Amidi</i>. Super Study Guide: Data Science Tools// <i>Afshine Amidi, Shervine Amidi</i> . Software Tools/ Massachusetts Institute of Technology. 2020</p> <p>4. <i>Ian H. Witten</i>. Data Mining. Practical Machine Learning Tools and Techniques // <i>Ian H. Witten, Eibe Frank</i>. Elsevir, 2005</p> <p>5. <i>Jiawei Han</i>. Data Mining: Concepts and Tecniques.// <i>Jiawei Han, Micheline Kamber</i>. Morgan Kaufmann Publishers, 2011</p> <p>6. <i>А.А.Барсегян</i>. Технологи анализа данных: Data Mining, Visual Mining. Text Mining, OLAP.// <i>А.А.Барсегян, М.С.Куприянов, В.В.Степаненко И.И.Холод</i> Санкт-Петербург. „БХВ-Петербург”. 2007.-376с.</p> <p>7. <i>В.Дюк, А.Самойленко</i>. Data Mining: Ученый курс. // <i>В.Дюк, А.Самойленко</i> – СПб: Питер. 2001. - 368с.</p> <p>8. <i>В.П.Боровиков</i>. STATISTICA.Искусство анализа данных на компьютере: Для профессионалов – СПб.: Питер, 2003 – 688 с.</p> <p>9. . https://www.kdnuggets.com/</p> <p>10.https://www.kaggle.com/datasets</p> <p>11.https://colab.research.google.com/notebooks/intro.ipynb</p>
Обсяг курсу	Загальний обсяг: 150 годин. Аудиторних занять: 48 год., з них 32 год. лекцій та 16 години лабораторних робіт. Самостійної роботи: 102 год.
Очікувані результати навчання	Після завершення цього курсу студент буде : знати: основні задачі аналізу даних і методи їх розв’язування; вміти: застосовувати інструментальні засоби (R, Python, etc) для аналізу даних, зокрема для виявлення неочевидних закономірностей в даних..
Ключові слова	Аналіз даних, класифікація, кластеризація, прогнозування, регресія, бібліотека, R, Python
Формат курсу	Очний, дистанційний Проведення лекцій, лабораторних робіт і консультацій.
Теми	<ol style="list-style-type: none"> 1. Що таке аналіз даних. Дані. Інформація. Знання. 2. Задачі аналізу даних.. 3. Методи класифікації. 4. Методи кластерного аналізу. 5. Методи прогнозування. 6. Огляд комп’ютерних засобів аналізу даних. 7. Open Source інструменти для аналізу даних. 8. Реалізація методів аналізу даних засобами пакету R. 9. Бібліотеки Python для аналізу даних, візуалізації і машинного навчання. 10. Інструментальні засоби виявлення знань в даних. Система See 5. Побудова дерева розв’язків в системі See 5. 11. Побудова дедрограми інструментальними засобами. 12. Побудова регресійних моделей засобами R та Python.
Підсумковий контроль, форма	Залік у кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з <ul style="list-style-type: none"> - Теорії ймовірностей і математичної статистики; - Програмування; достатніх для сприйняття категоріального апарату і розуміння джерел
Навчальні методи та техніки,	Презентації, лекції Індивідуальні завдання

які будуть використовуватися під час викладання курсу	Групові проекти.
Необхідне обладнання	Комп'ютер із програмним забезпеченням (R, Python, etc) , Internet , проектор
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"> • 3 індивідуальні завдання , кожне максимально до 30% семестрової оцінки; максимальна кількість балів 90; • участь в обговоренні і дискусії при захисті завдань одногрупників – 10%: семестрової оцінки; максимальна кількість балів 10 <p>Підсумкова максимальна кількість балів 100.</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 індивідуальних завдань, передбачених курсом.</p> <p>Література. Тексти лекцій, вимоги до виконання індивідуальних завдань і основна література доступна студентам в системі MS TEAMS. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до заліку.	<ol style="list-style-type: none"> 1. Що таке аналіз даних. Дані. Інформація. Знання. 2. Задачі аналізу даних. 3. Методи класифікації. 4. Методи кластерного аналізу. 5. Методи прогнозування. 6. Огляд комп'ютерних засобів аналізу даних. 7. Open Source інструменти для аналізу даних. 8. Реалізація методів аналізу даних засобами RStudio. 9. Бібліотеки Python для аналізу даних, візуалізації і машинного навчання. 10. Інструментальні засоби виявлення знань в даних. Система See 5. Побудова дерева розв'язків в системі See 5. 11. Побудова дедрограми інструментальними засобами. 12. Побудова регресійних моделей засобами R та Python.
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.