



СИЛАБУС ДИСЦИПЛІНИ «МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ У СФЕРІ ПРИРОДОКОРИСТУВАННЯ»

Ступінь вищої освіти - Магістр

Спеціальність 122 «Комп'ютерні науки»

Освітня програма «Інформаційні управляючі системи і технології»

Рік навчання 1, семестр 1

Форма навчання денна

Кількість кредитів ЄКТС 4

Мова викладання українська

Лектор курсу

Контактна інформація

лектора (e-mail)

Сторінка курсу в eLearn

Кириченко Віктор Вікторович

Кафедра комп'ютерних наук, к.15, ауд.237

e-mail v.kyrychenko@nubip.edu.ua

<https://elearn.nubip.edu.ua/course/view.php?id=2827>

ОПИС ДИСЦИПЛІНИ

(до 1000 друкованих знаків)

Інформаційні управляючі системи, що вивчаються сучасною наукою, з великими труднощами піддаються дослідженню звичайними (вербальними) теоретичними методами. Прямий експеримент над ними неможливий. Ціна помилок і прорахунків велика, тому математичне моделювання є неминучою складовою науково-технічного прогресу.

Сучасний спеціаліст у сфері природокористування повинен знати і уміти використовувати в повсякденній роботі новітні економіко-математичні методи і моделі. Швидкий розвиток і широке застосування засобів обчислювальної техніки зумовлюють вимоги до підготовки сучасного спеціаліста у сфері природокористування, який повинен за допомогою сучасних пакетів прикладних програм уміти аналізувати складні природно-економічні та соціальні явища.

Моделювання та прогнозування у сфері природокористування – одна з базових дисциплін підготовки фахівців у сфері природокористування. Ця дисципліна заснована на фундаменті знань з таких дисциплін як "Теорія ймовірностей, імовірнісні процеси і математична статистика", "Статистичні методи", "Методи та системи штучного інтелекту", "Програмування Python".

Мета – розширення та поглиблення теоретичних знань та набуття професійних компетентностей щодо прогнозування соціально-економічних процесів, процесів у сфері природокористування та моделювання складних систем за допомогою статистичних методів та моделей і сучасних інформаційних технологій.

Предметом навчальної дисципліни є: теоретичні та практичні питання щодо розробки прогнозів та побудови моделей в умовах ринкової економіки в сфері природокористування на підставі широкого використання сукупності математичних методів і моделей.

Основними завданнями навчальної дисципліни є: визначення основних особливостей моделювання та прогнозування складних інформаційно-управляючих систем; ознайомлення з існуючими статистичними методами та моделями; дослідження соціально-економічних процесів у сфері природокористування за допомогою, економетричних моделей, моделей кластерного та дискримінантного аналізів, багатофакторних індексних моделей, аналіз слабко формалізованих ситуацій за допомогою експертного аналізу.

Методологія та методика навчальної дисципліни базується на роботах вітчизняних і закордонних вчених з питань статистичного моделювання і прогнозування соціально-економічних систем.

Набуття компетентностей

Інтегральна компетентність:

Здатність розв'язувати задачі дослідницького та/або інноваційного характеру у сфері комп'ютерних наук.

Загальні компетентності:

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

Фахові (спеціальні) компетентності:

СК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

Програмні результати навчання

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.

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.

РН6. Розробляти концептуальну модель інформаційної або комп'ютерної системи.

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН8. Розробляти математичні моделі та методи аналізу даних (включно з великим).

СТРУКТУРА КУРСУ

Тема	Години (лекції/лаб. /самост.)	Результати навчання	Завдання	Оціню вання
Модуль 1 Проблеми представлення атомарних та багатовимірних структур даних				
Тема 1. Вступ до Data Science. Знайомство з бібліотекою Numpy.	2/2/10	Ознайомитись з видами задач машинного навчання, опанувати інструменти для інтерактивної роботи з Python - IPython, Jupyter Notebook та Google Colaboratory, ознайомитись з бібліотекою NumPy, навчитись працювати з типом даних Array, вивчити базові математичні функції Array, навчитесь створювати масиви та виконувати базові операції над ними в бібліотеці NumPy.	Виконання лабораторної роботи №1.	24

Тема 2. NumPy. Лінійна алгебра і статистика.	2/2/10	Отримати вступ до лінійної алгебри та розібрати на прикладах, як ці математичні функції працюють у бібліотеці NumPy, навчитися використовувати математичні та статистичні функції в NumPy, а також працювати з NumPy-матрицями та масивами різної величини.	Захист лабораторної роботи №1.	16
Тема 3. Знайомство з бібліотекою Pandas	2/2/10	Навчитися працювати з даними за допомогою бібліотеки Pandas, ознайомитись зі структурами даних у Pandas (Series та Dataframe) і навчитися проводити базові операції над ними, навчитися створювати, імпортувати й експортувати табличні дані в Pandas Dataframe, навчитися швидко та зручно працювати з табличними даними за допомогою Pandas.	Виконання лабораторної роботи №2.	24
Тема 4. Візуалізація даних	4/4/10	Ознайомитись з бібліотекою Matplotlib, навчитися будувати базові візуалізації в Python за допомогою Matplotlib, дізнатися про три способи візуалізувати дані в Pandas Data Frame, навчитися будувати просунуті візуалізації: heatmap, box plot, pair plot, cumulative plot, навчитися будувати базові інтерактивні візуалізації.	Захист лабораторної роботи №2.	16
Модульний контроль				20
Усього за модулем 1	10/10/40	100		

Модуль 2 Управління кубами OLAP				
Тема 5. Exploratory Data Analysis (EDA) та очищення даних	2/2/10	Дізнатися, що таке Exploratory Data Analysis (EDA) та його основні компоненти, Навчитися швидко розуміти структуру набору даних будь-якого розміру, дізнатися, що робити з пропущеними даними та дублікатами, навчитися проводити EDA за допомогою Pandas, SweetViz, Pandas Profiling.	Виконання лабораторної роботи №3.	24
Тема 6. Пошук і видалення пропущених значень	2/2/10	Дізнатися, що таке пропущені значення та чому їх треба опрацьовувати, ознайомитись з типами пропущених значень (випадкові, наслідки системної помилки тощо), навчитися виявляти пропущені значення у вибірці даних, розглянути різні способи заповнення пропущених значень.	Захист лабораторної роботи №3.	16
Тема 7. Задача прогнозування. Лінійна регресія. Гرادієнтний спуск	2/2/10	Розібрати модель простої лінійної регресії та її математичну основу, ознайомитись з методом найменших квадратів і методом градієнтного спуску, навчитись тренувати модель лінійної регресії за допомогою бібліотеки scikit-learn, дізнатись, як оцінювати якість моделі лінійної регресії, та навчитись працювати з метриками MSE і RMSE.	Виконання лабораторної роботи №4.	24
Тема 8. Інші типи регресій	4/4/10	Розібрати модель багатовимірної лінійної регресії та	Захист лабораторної роботи №4.	16

		навчитися будувати її за допомогою scikit-learn, навчитися інтерпретувати коефіцієнти лінійної регресії, розібрати модель поліноміальної регресії, дізнатись, як оцінювати якість моделі за допомогою метрик MAE, MAPE, R-squared, розібрати моделі гребінцевої регресії, регресії за методом LASSO та регресії «еластична мережа».		
Модульний контроль				20
Усього за модулем 2	10/10/40			100
Усього за навчальну частину				70
Екзамен				30
Усього за курс	20/20/80			100

ПОЛІТИКА ОЦІНЮВАННЯ

Політика щодо дедлайнів та перескладання:	Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку. Перескладання модулів відбувається із дозволу лектора за наявності поважних причин (наприклад, лікарняний).
Політика щодо академічної доброчесності:	Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних девайсів). Курсові роботи, реферати повинні мати коректні текстові посилання на використану літературу
Політика щодо відвідування:	Відвідування занять є обов'язковим. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись індивідуально (в он-лайн формі за погодженням із деканом факультету)

ШКАЛА ОЦІНЮВАННЯ СТУДЕНТІВ

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів, курсових проєктів та заліків	
	екзаменів	заліків
90-100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
0-59	незадовільно	не зараховано

Рекомендовані джерела інформації

1. Математичні методи моделювання: підручник / Усов А.В., Савельєва О.С., Становська І.І., Перпері А.О. під наук. ред. О. Л. Становського; Одес. нац. політехн. ун-т. Одеса, 2020. 500 с. URL: <http://dspace.opu.ua/jspui/handle/123456789/11400>.
2. Моделювання та прогнозування стану довкілля [підручник] / І.І. Яскевець, Н.М. Протас, Т.Ю. Осипова, Д.Ю. Касаткін // - К.: НУБіП України, 2018.- 566 с.
3. Пасічник Т.В. Моделювання та прогнозування стану довкілля. Навчальний посібник. -К: «Магнолія 2006», - 2023. – 200 с.
4. Скіп Б. Моделювання та прогнозування стану довкілля. Навчальний посібник, Ч І. Чернівці "Рута", –2004. – 65с.
5. Скіп Б. Моделювання та прогнозування стану довкілля. Навчальний посібник, Чернівці "Рута", II –2005. – 56 с.
6. Скіп Б.В., Філіпчук Т.В., Моделювання та прогнозування стану довкілля. Практикум, Чернівці "Рута", –2006. – 68 с.

Інтернет-джерела та відео матеріали

1. Python Download: <https://www.python.org/downloads/>
2. Installing Jupyter: <https://jupyter.org/install>
3. Google Colab: <https://colab.research.google.com/>
4. Array Programming with NumPy: <https://realpython.com/numpy-array-programming/>
5. Introduction to Numerical Computing with NumPy: <https://www.youtube.com/watch?v=ZB7BZMhfPgk>
6. Обернена матриця: https://uk.wikipedia.org/wiki/%D0%9E%D0%B1%D0%B5%D1%80%D0%BD%D0%B5%D0%BD%D0%B0_%D0%BC%D0%B0%D1%82%D1%80%D0%B8%D1%86%D1%8F
7. Linear Algebra for Machine Learning: <https://machinelearningmastery.com/linear-algebra-machine-learning-7-day-mini-course/>
8. Bessel's correction: https://en.wikipedia.org/wiki/Bessel%27s_correction
9. Intro to Pandas data structures: https://pandas.pydata.org/docs/user_guide/dsintro.html#dsintro
10. Pandas Cheat Sheet for Data Science in Python: <https://www.datacamp.com/cheat-sheet/pandas-cheat-sheet-for-data-science-in-python>
11. Panda's docs: <https://pandas.pydata.org/>
12. Matplotlib Tutorials: <https://matplotlib.org/stable/tutorials/index.html>
13. Pyplot vs Object Oriented Interface: <https://matplotlib.org/matplotlibblog/posts/pyplot-vs-object-oriented-interface/>
14. Pyplot: <https://plotly.com/python/>
15. Seaborn: <https://seaborn.pydata.org/>
16. Data Understanding for Machine Learning: Assessment & Exploration: <https://towardsdatascience.com/data-understanding-for-machine-learning-assessment-exploration-aca1adc1cb6>
17. Fast EDA in Jupyter & Colab notebooks using Sweetviz 2: <https://towardsdatascience.com/fast-eda-in-jupyter-colab-notebooks-using-sweetviz-2-0-99c22bcb3a1c>

18. What is Exploratory Data Analysis?: <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>
19. Advanced exploratory data analysis (EDA) with Python: <https://medium.com/epfl-extension-school/advanced-exploratory-data-analysis-eda-with-python-536fa83c578a>
20. Different types of missing data: <https://medium.com/analytics-vidhya/different-types-of-missing-data-59c87c046bf7>
21. Ways to impute missing values in the data: <https://medium.com/analytics-vidhya/ways-to-impute-missing-values-in-the-data-fc38e7d7e2c1>
22. Machine Learning Basics: Model, Cost function and Gradient Descent: <https://medium.com/@dhartidhami/machine-learning-basics-model-cost-function-and-gradient-descent-79b69ff28091>
23. Метод найменших квадратів: https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0%D0%B9%D0%BC%D0%B5%D0%BD%D1%88%D0%B8%D1%85_%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D1%96%D0%B2
24. Regression Metrics for Machine Learning: [https://machinelearningmastery.com/regression-metrics-for-machine-learning/#:~:text=There%20are%20three%20error%20metrics,Mean%20Absolute%20Error%20\(MAE\)](https://machinelearningmastery.com/regression-metrics-for-machine-learning/#:~:text=There%20are%20three%20error%20metrics,Mean%20Absolute%20Error%20(MAE))
25. scikit-learn: <https://scikit-learn.org/stable/>
26. Linear Regression: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression
27. All about Feature Scaling: <https://dlantsst5xjbh.cloudfront.net/data-science-python/text-8.pdf>
28. Overfitting and underfitting in machine learning: <https://www.superannotate.com/blog/overfitting-and-underfitting-in-machine-learning>