



Кафедра комп'ютерної інженерії та інформаційних технологій

СИЛАБУС

Базова інформація про дисципліну	
Назва дисципліни	SE011 / Прикладний аналіз даних / Applied Data Analysis
Рівень вищої освіти / фахової передвищої освіти	Фахова передвища
Галузь знань	12 «Інформаційні технології»
Спеціальність	121 «Інженерія програмного забезпечення»
Освітня програма	Інженерія програмного забезпечення
Семестр	8 семестр
Факультет / відділення	Інженерії програмного забезпечення
Курс	4 курс (9 кл), 3 курс (11 кл)
Анотація курсу	Даний курс знайомить студентів зі статистичним аналізом даних та моделями, побудованими на даних. Також він формує практичні навички використання бібліотечних інструментів мови програмування Python для побудови проєктів з базовими моделями машинного навчання. У якості інструментальної бази використовуються мова програмування Python, бібліотеки numpy, pandas, scipy, scikit-learn.
Сторінка курсу в MOODLE	http://78.137.2.119:1919/m72/course/view.php?id=1150
Мова викладання	Українська
Лектор курсу	Марченко С.В., спеціаліст I категорії канали комунікації: СДН «Moodle»: повідомлення в чаті E-mail: numosmanius@gmail.com
Місце дисципліни в освітній програмі	
Освітня програма	http://csbc.edu.ua/documents/otdel/koop_pr.pdf

<p>Перелік загальних компетентностей (ЗК)</p>	<p>ЗК05. Знання та розуміння предметної області та розуміння професійної діяльності.</p> <p>ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>ЗК07. Здатність застосовувати знання у практичних ситуаціях.</p>
<p>Перелік спеціальних компетентностей (СК)</p>	<p>СК01. Здатність алгоритмічно та логічно мислити.</p> <p>СК02. Здатність вдосконалювати знання і навички в галузі інформаційних технологій та усвідомлення важливості навчання протягом усього життя.</p> <p>СК09. Здатність вибирати та використовувати ефективні інструментальні засоби розробки програмного продукту.</p> <p>СК10. Здатність реалізовувати всі етапи життєвого циклу програмного забезпечення.</p>
<p>Перелік програмних результатів навчання</p>	<p>РН04. Використовувати знання математичних методів на рівні, необхідному для розв'язання типових задач програмної інженерії.</p> <p>РН05. Розробляти та супроводжувати програмне забезпечення.</p> <p>РН08. Аналізувати вимоги до програмного забезпечення.</p> <p>РН10. Обирати та застосовувати ефективні методи оптимізації алгоритмів.</p> <p>РН12. Впроваджувати і супроводжувати програмні продукти.</p> <p>РН14. Розуміти предметну область, застосовувати знання у професійній діяльності.</p> <p>РН15. Аналізувати та узагальнювати необхідну інформацію з різних джерел та ресурсів для розв'язання професійних задач з урахуванням сучасних досягнень інформаційних технологій.</p>
<p>Опис дисципліни</p>	
<p>Структура навантаження на студента</p>	<p>Загальна кількість годин – 90</p> <p>Кількість кредитів – 3</p> <p>Кількість лекційних годин – 24</p> <p>Кількість практичних занять – 24</p>

	<p>Кількість годин для самостійної роботи студентів – 42</p> <p>Форма підсумкового контролю – екзамен</p>
Методи навчання	<p>За подачею навчального матеріалу: методи готових знань, дослідницький метод.</p> <p>З огляду на мету навчання: методи здобуття нових знань, метод формування умінь і навичок, метод застосування знань на практиці, методи закріплення знань, умінь і навичок, методи перевірки і оцінювання знань, умінь і навичок.</p>
Зміст дисципліни	
Тема 1. Вступ до аналізу даних	Дані та формулювання запитань до них. Задачі аналізу даних. Збір даних. Базові операції з датафреймами та серіями. Фільтрація, групування та агрегація даних. Ймовірність для аналізу даних. Описова статистика та очистка даних.
Тема 2. Статистичний аналіз даних	Статистичні розподіли. Вибірка з розподілів. Статистичні критерії. Баєсове висновування. Перевірка нормальності. Кореляційний аналіз. Параметричні критерії та дисперсійний аналіз. Непараметричні критерії.
Тема 3. Розвідувальний аналіз даних	Розвідувальний аналіз та візуалізація даних. Базові види діаграм для візуалізації даних. Вибір графіків для візуалізації даних та додаткові інструментальні засоби візуалізації даних. Модульна контрольна робота 1.
Тема 4. Задача кластеризації та методи вимірювання відстаней	Вимірювання відстаней. Кластеризація на основі щільності. Кластеризація на основі розподілу. Кластеризація на основі центроїдів. Ієрархічна кластеризація. Спектральна кластеризація.
Тема 5. Регресійний аналіз даних	Лінійна регресія. Регресія як підбір параметрів. Спрощення моделей за допомогою регуляризації. Регресійний аналіз на базі методу опорних векторів. Регресійний аналіз на базі дерев рішень.
Тема 6. Задачі класифікації	Регресія для задачі класифікації. Логістична регресія. Проблеми логістичної класифікації.

	Класифікація на основі методу опорних векторів. Класифікація на основі дерев рішень. Модульна контрольна робота 2.					
Політика дисципліни						
Політика відвідування	Регулярне відвідування всіх видів занять, своєчасність виконання самостійної роботи. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання зорганізується в он-лайн формі за погодженням із керівником курсу.					
Політика щодо дедлайнів та перескладання	Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.					
Академічна доброчесність	У випадку недотримання політики академічної доброчесності (плагіат, самоплагіат, фабрикація, фальсифікація, списування, обман, хабарництво) передбачено повторне проходження оцінювання.					
Система оцінювання						
<p>Поточний контроль здійснюється протягом семестру під час проведення практичних, семінарських та інших видів занять і оцінюється сумою набраних балів (максимальна сума – 100 балів; мінімальна сума, що дозволяє студенту отримати атестацію з предмету – 60 балів); підсумковий/ семестровий контроль, проводиться у формі іспиту, відповідно до графіку навчального процесу. Підсумкова оцінка за умови іспиту виставляється як загальна сума балів набраних за результатами поточного (70%) та підсумкового контролю.</p> <p style="text-align: center;">Накопичування рейтингових балів з навчальної дисципліни</p>						
Види навчальної роботи						Мах кількість балів
Аудиторна						
Практичні роботи (6 тем)						60
	T1	T2	T3	T4	T5	T6
	10	10	10	6	10	14
Тестування (2 x 5 балів)						10
Індивідуальна						
Захист практичного завдання до самостійної роботи						30
	T1	T2	T3	T4	T5	T6
	-	-	15	-	-	15
РАЗОМ						100

Шкала оцінювання		
ECTS	Бали	Зміст
A	90-100	Бездоганна підготовка в широкому контексті
B	80-89	Повні знання, міцні вміння
C	70-79	Хороші знання та вміння
D	65-69	Задовільні знання, стереотипні вміння
E	60-64	Виконання мінімальних вимог діяльності в стандартних умовах
FX	35-59	Слабкі знання, відсутність умінь
F	1-34	Необхідний повторний курс

Список рекомендованих джерел

Основна література

1. Liu Y. Python Machine Learning By Example: Build intelligent systems using Python, TensorFlow 2, PyTorch, and scikit-learn, Third Edition. Birmingham: Packt Publishing, 2020. 504с.
2. Campesato O. Managing Datasets and Models. Dulles: Mercury Learning and Information, 2023. 368с.
3. Brownlee J. Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python. Edition: v1.4. 2019. 274с.
4. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, 2nd Edition. Sebastopol: O'Reilly Media, 2020. 368с.
5. Downey A. B. Think Bayes: Bayesian Statistics in Python, 2nd Edition. Sebastopol: O'Reilly Media, 2021. 338с.
6. Stepanek H. Thinking in Pandas: How to Use the Python Data Analysis Library the Right Way. Portland: Apress, 2020. 188с.
7. Harrison M., Petrou T. Pandas 1.x Cookbook: Practical recipes for scientific computing, time series analysis, and exploratory data analysis using Python, Second Edition. Birmingham: Packt Publishing, 2020. 606с.
8. Grus A. Data Science from Scratch: First Principles with Python. Sebastopol: O'Reilly Media, 2019. 384с.
9. Géron A. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Farnham: O'Reilly UK, 2019. 600с.
10. Kuhn M., Johnson K. Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton: CRC Press, 2020. 298с.

11. Walker M. Python Data Cleaning Cookbook: Modern techniques and Python tools to detect and remove dirty data and extract key insights. Birmingham: Packt Publishing, 2020. 416с.
12. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. Харків: ХНУ імені В. Н. Каразіна, 2020. 301 с.

Допоміжна література

1. Downey A. B. Think Stats – Second Edition. Sebastopol: O'Reilly Media, 2015. 208с.
2. Harrison M. Effective Pandas: Patterns for Data Manipulation (Treading on Python). 2021. 497с.
3. Automated Machine Learning: Methods, Systems, Challenges (The Springer Series on Challenges in Machine Learning) / За ред. F. Hutter, L. Kotthoff, J. Vanschoren. Springer, 2019. 220с.
4. Bargagliotti A., Franklin Ch., Statistics and Data Science for Teachers. American Statistical Association, 2021. 234с.
5. De Mesquita E. B., Fowler A. Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis. Princeton University Press, 2021. 378с.
6. Vora D. R., Bhatia G. S. Python Machine Learning Projects: Learn how to build Machine Learning projects from scratch. London: BPB Online, 2023. 272с.
7. Skiena S. S. The Data Science Design Manual (Texts in Computer Science). Springer, 2017. 454с.
8. Hackeling G. Mastering Machine Learning with scikit-learn – Second Edition. Birmingham: Packt Publishing, 2017. 250с.
9. Smith T. Supervised Machine Learning with Python: Develop rich Python coding practices while exploring supervised machine learning. Birmingham: Packt Publishing, 2019. 156с.

Інформаційні ресурси

1. Duchesnay E., Löfstedt T., Younes F. Statistics and Machine Learning in Python [Електронний ресурс]. URL: <https://duchesnay.github.io/pystatsml/> (дата звернення: 29.08.2023р.).

2. Intro to Data Analysis [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-data-analysis--ud170> (дата звернення: 29.08.2023р.).
3. Intro to Inferential Statistics [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-inferential-statistics--ud201> (дата звернення: 29.08.2023р.).
4. Intro to Statistics [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-statistics--st101> (дата звернення: 29.08.2023р.).
5. 20 Python Projects for Data Science in 2023 [Електронний ресурс]. URL: <https://www.projectpro.io/article/python-projects-for-data-science/462> (дата звернення: 29.08.2023р.).
6. Level up with the largest AI & ML community [Електронний ресурс]. URL: <https://www.kaggle.com/> (дата звернення: 29.08.2023р.).
7. Spurious correlations [Електронний ресурс]. URL: <https://www.tylervigen.com/spurious-correlations> (дата звернення: 29.08.2023р.).
8. Спеціалізація Машинне навчання [Електронний ресурс]. URL: <https://www.coursera.org/specializations/machine-learning-introduction> (дата звернення: 29.08.2023р.).
9. Спеціалізація Прикладна наука про дані [Електронний ресурс]. URL: <https://www.coursera.org/specializations/applied-data-science#courses> (дата звернення: 29.08.2023р.).