

СИЛАБУС

Базова інформація про дисципліну	
Назва дисципліни	SE011 Прикладний аналіз даних / Applied Data Analysis
Рівень фахової передвищої освіти	Фаховий молодший бакалавр
Галузь знань	12 «Інформаційні технології»
Спеціальність	121 «Інженерія програмного забезпечення»
Освітня програма	Інженерія програмного забезпечення
Семестр	6 семестр (на базі повної загальної середньої освіти) 8 семестр (на базі базової середньої освіти)
Курс	3 курс (на базі повної загальної середньої освіти) 4 курс (на базі базової середньої освіти)
Анотація курсу	Цей 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://78.137.2.119:2929/course/view.php?id=562
Мова викладання	українська
Лектор курсу	Марченко Станіслав Віталійович, спеціаліст першої категорії канали комунікації: СДН «Moodle»: повідомлення в чаті E-mail: sv.marchenko1989@gmail.com
Місце дисципліни в освітній програмі	
Освітня програма	http://csbc.edu.ua/documents/otdel/moop_pr23.pdf
Інтегральна компетентність	Здатність вирішувати типові спеціалізовані задачі інженерії програмного забезпечення, що вимагає застосування положень і методів відповідних наук (математики, інформатики, інформаційних технологій, тощо) та може характеризуватися певною невизначеністю умов; нести відповідальність за результати своєї діяльності; здійснювати контроль інших осіб у визначених ситуаціях.

Базова інформація про дисципліну	
Назва дисципліни	SE011 Прикладний аналіз даних / Applied Data Analysis
Рівень фахової передвищої освіти	Фаховий молодший бакалавр
Галузь знань	12 «Інформаційні технології»
Спеціальність	121 «Інженерія програмного забезпечення»
Перелік загальних компетентностей (ЗК)	ЗК05. Знання та розуміння предметної області та розуміння професійної діяльності. ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел. ЗК07. Здатність застосовувати знання у практичних ситуаціях.
Перелік спеціальних компетентностей (СК)	СК01. Здатність алгоритмічно та логічно мислити. СК02. Здатність вдосконалювати знання і навички в галузі інформаційних технологій та усвідомлення важливості навчання протягом усього життя. СК09. Здатність вибирати та використовувати ефективні інструментальні засоби розробки програмного продукту. СК10. Здатність реалізовувати всі етапи життєвого циклу програмного забезпечення.
Перелік програмних результатів навчання	РН04. Використовувати знання математичних методів на рівні, необхідному для розв'язання типових задач програмної інженерії. РН05. Розробляти та супроводжувати програмне забезпечення. РН08. Аналізувати вимоги до програмного забезпечення. РН10. Обирати та застосовувати ефективні методи оптимізації алгоритмів. РН12. Впроваджувати і супроводжувати програмні продукти. РН14. Розуміти предметну область, застосовувати знання у професійній діяльності. РН15. Аналізувати та узагальнювати необхідну інформацію з різних джерел та ресурсів для розв'язання професійних задач з урахуванням сучасних досягнень інформаційних технологій.
Опис дисципліни	
Структура навантаження на студента	Загальна кількість годин – 90 Кількість кредитів – 3 Кількість лекційних годин – 24 Кількість практичних занять – 24 Кількість годин для самостійної роботи студентів – 42 Форма підсумкового контролю – іспит
Методи навчання	За подачею навчального матеріалу: методи готових знань, дослідницький метод. З огляду на мету навчання: методи здобуття нових знань, метод формування умінь і навичок, метод застосування

Базова інформація про дисципліну	
Назва дисципліни	SE011 Прикладний аналіз даних / Applied Data Analysis
Рівень фахової передвищої освіти	Фаховий молодший бакалавр
Галузь знань	12 «Інформаційні технології»
Спеціальність	121 «Інженерія програмного забезпечення»
	знань на практиці, методи закріплення знань, умінь і навичок, методи перевірки і оцінювання знань, умінь і навичок.
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.

Базова інформація про дисципліну	
Назва дисципліни	SE011 Прикладний аналіз даних / Applied Data Analysis
Рівень фахової передвищої освіти	Фаховий молодший бакалавр
Галузь знань	12 «Інформаційні технології»
Спеціальність	121 «Інженерія програмного забезпечення»
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.
Використання ІІІ	Використання ІІІ під час виконання завдань регламентується Положенням «Про використання ІІІ в освітньому процесі ЧДБК». Завдання містять маркування щодо регламенту використання ІІІ.
Підсумковий контроль	Іспит
Система оцінювання Система оцінювання підсумкової успішності студентів поділяється на поточний контроль та семестровий контроль. Поточний контроль здійснюється протягом семестру і охоплює всі види аудиторної роботи (практичні, семінарські заняття) та виконання індивідуальних завдань. Максимальна кількість балів, яку студент може набрати за цей вид контролю, становить 100. Підсумковий контроль Відбувається в формі іспиту. Розрахунок підсумкової оцінки Підсумкова оцінка (О) розраховується як сума балів за роботу в семестрі (S) та балів за іспит (Т), кожен з яких має максимальну вагу 100 балів. Вагові коефіцієнти для обох компонентів однакові та дорівнюють 0,5. Формула: $O = (S \times 0.5) + (T \times 0.5)$	

Види навчальної роботи	Загальна кількість балів
Практична робота за темами 1, 2, 3, 5 по 10 балів	40
за темою 4 по 6 балів	6
за темою 7 по 12 балів	12
Модульні контрольні (2 к.р. по 6 балів)	12
Індивідуальна самостійна робота (задачі)	30
Разом	100

Критерії оцінювання для кожного виду навчальної роботи
Критерії оцінювання практичних робіт 100% – звіт з практичної роботи подано вчасно з дотриманням умов академічної доброчесності, термінів звітування, вимог до оформлення та коректності документування, всі завдання розв’язано, захищено усно (якщо передбачено умовою завдань). 90-99% – звіт з практичної роботи подано вчасно з дотриманням умов академічної доброчесності, термінів звітування, наявністю зауважень до оформлення та коректності документування, всі завдання розв’язано, захищено усно (якщо передбачено умовою завдань).

80-89% – звіт з практичної роботи подано вчасно з дотриманням умов академічної доброчесності, термінів звітування, наявністю значних зауважень до оформлення та коректності документування, всі завдання розв’язано, захищено усно (якщо передбачено умовою завдань).

50-79% – звіт з практичної роботи подано вчасно з дотриманням умов академічної доброчесності, термінів звітування, вимог до оформлення та коректності документування, всі завдання розв’язано, проте не захищено усно (якщо це передбачено умовою завдань).

Відсотковий показник визначається стосовно сумарної кількості балів, отриманої за практичну роботу за результатами звітування. Порухення умов академічної доброчесності та/або визначених термінів звітування може позбавити здобувача освіти права подачі звіту за відповідною темою.

Критерії оцінювання модульних робіт

Модульні контрольні роботи складаються з набору закритих тестових завдань та автоматично перевіряються засобами СДН Moodle.

Критерії оцінювання індивідуальних робіт (проектів)

Індивідуальні роботи охоплюють сукупність завдань на розширення та поглиблення розглянутих тем. Критерії їх оцінювання аналогічні критеріям для завдань практичних робіт.

Критерії оцінювання іспиту

За екзамен з предмету можна набрати 100 балів максимально. Білет складається з 3 питань, двох теоретичних та одного теоретико-практичних. За відповідь на теоретичне питання можна отримати максимально 30 балів, виконане теоретико-практичне завдання – 40 балів.

Таблиця 1. Бали за відповідь на теоретичне питання

Бали	Критерії
30	Студент (студентка) правильно, повно, чітко і логічно відповідає на всі поставлені питання
20-29	Правильно і повно, інколи з деякою неточністю та з допомогою навідних (пояснювальних) питань відповідає на всі поставлені питання
1-19	Відповідь студента (студентки) неповна і неточна, на навідні питання відповідає не зовсім точно, відповідь не є чіткою і логічною
0	Не може відповісти на питання взагалі

Таблиця 2. Бали за виконання теоретико-практичного завдання

Бали	Критерії
40	Студент (студентка) самостійно деталізує вхідні дані до завдання, аналізує алгоритм виконання завдання та особливості реалізації. Програмна реалізація чи візуалізація повністю відповідає поставленому завданню, і студент (студентка) здатний вносити в неї невеликі корективи.
25-39	Студент (студентка) самостійно деталізує вхідні дані до завдання, але не може обґрунтувати їх доцільність, аналізує алгоритм виконання завдання. Програмна реалізація чи візуалізація переважно відповідає поставленому завданню.
10-24	Студент (студентка) знає алгоритм виконання, потрібні синтаксичні інструменти. Програмна реалізація чи візуалізація частково відповідає поставленому завданню.
1-9	Студент (студентка) знає алгоритм виконання, потрібні синтаксичні інструменти, проте програмну реалізацію чи візуалізацію здійснити не може.
0	Завдання не виконано взагалі

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

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. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, 2nd Edition. Sebastopol: O'Reilly Media, 2020. 368с.
4. Downey A. B. Think Bayes: Bayesian Statistics in Python, 2nd Edition. Sebastopol: O'Reilly Media, 2021. 338с.
5. Stepanek H. Thinking in Pandas: How to Use the Python Data Analysis Library the Right Way. Portland: Apress, 2020. 188с.
6. 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.
7. Kuhn M., Johnson K. Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton: CRC Press, 2020. 298с.
8. 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.
9. Harrison M. Effective Pandas: Patterns for Data Manipulation (Treading on Python). 2021. 497с.
10. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка». 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. 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.
3. Bargagliotti A., Franklin Ch., Statistics and Data Science for Teachers. American Statistical Association, 2021. 234с.
4. De Mesquita E. B., Fowler A. Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis. Princeton University Press, 2021. 378с.
5. Vora D. R., Bhatia G. S. Python Machine Learning Projects: Learn how to build Machine Learning projects from scratch. London: BPB Online, 2023. 272с.
6. Grus A. Data Science from Scratch: First Principles with Python. Sebastopol: O'Reilly Media, 2019. 384с.
7. 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.
8. Skiena S. S. The Data Science Design Manual (Texts in Computer Science). Springer, 2017. 454с.
9. Hackeling G. Mastering Machine Learning with scikit-learn – Second Edition. Birmingham: Packt Publishing, 2017. 250с.

10. Smith T. Supervised Machine Learning with Python: Develop rich Python coding practices while exploring supervised machine learning. Birmingham: Packt Publishing, 2019. 156с.
11. Brownlee J. Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python. Edition: v1.4. 2019. 274с.

Інтернет ресурси

1. Duchesnay E., Löfstedt T., Younes F. Statistics and Machine Learning in Python [Електронний ресурс]. URL: <https://duchesnay.github.io/pystatsml/> (дата звернення: 29.08.2025р.).
2. Intro to Data Analysis [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-data-analysis--ud170> (дата звернення: 29.08.2025р.).
3. Intro to Inferential Statistics [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-inferential-statistics--ud201> (дата звернення: 29.08.2025р.).
4. Intro to Statistics [Електронний ресурс]. URL: <https://www.udacity.com/course/intro-to-statistics--st101> (дата звернення: 29.08.2025р.).
5. 50+ Python Projects for Data Science in 2025 [Електронний ресурс]. URL: <https://www.projectpro.io/article/python-projects-for-data-science/462> (дата звернення: 29.08.2025р.).
6. Level up with the largest AI & ML community [Електронний ресурс]. URL: <https://www.kaggle.com/> (дата звернення: 29.08.2025р.).
7. Spurious correlations [Електронний ресурс]. URL: <https://www.tylervigen.com/spurious-correlations> (дата звернення: 29.08.2025р.).
8. Спеціалізація Машинне навчання [Електронний ресурс]. URL: <https://www.coursera.org/specializations/machine-learning-introduction> (дата звернення: 29.08.2025р.).
9. Спеціалізація Прикладна наука про дані [Електронний ресурс]. URL: <https://www.coursera.org/specializations/applied-data-science#courses> (дата звернення: 29.08.2025р.).