

Чернівецький національний університет імені Юрія Федьковича

Факультет математики та інформатики

Кафедра математичного моделювання

СИЛАБУС
навчальної дисципліни

DataMining з використанням мови Python
обов'язкова

Освітньо-професійна програма Інформаційні технології та управління проектами

Спеціальність 6.124 – Системний аналіз

Галузь знань 12 – Інформаційні технології

Рівень вищої освіти перший (бакалаврський)

Факультет математики та інформатики

Мова навчання українська

Розробник: Юрченко Ігор Валерійович,
доцент кафедри математичного моделювання,
кандидат фіз.-мат. наук, доцент

Профайл викладача <http://matmod.fmi.org.ua/pro-kafedru/spivrobotnyky/yurchenko-igor-valeriyovich/>

Контактний тел. 0372-58-48-25

E-mail: i.yurchenko@chnu.edu.ua

Сторінка курсу в Moodle <https://moodle.chnu.edu.ua/course/view.php?id=2062>

Консультації Онлайн-консультації: п'ятниця з 13:00 до 14:20.
Очні консультації: за попередньою домовленістю.

1. Анотація дисципліни (призначення навчальної дисципліни).

Навчальна дисципліна призначена для ознайомлення студентів з основними теоретичними та практичними аспектами обробки масивів даних, інтелектуального аналізу даних (DataMining) за допомогою широко розповсюджених бібліотек мови Python.

2. Мета навчальної дисципліни: ознайомити студентів із сучасним станом обробки масивів даних, основними задачами DataMining, їх базовими компонентами, методами, що застосовуються при обробці даних, методами використання бібліотек мови Python для розв'язання задач галузі DataMining, навчити студентів розв'язувати прикладні задачі обробки та аналізу даних різної розмірності з використанням комп'ютерних технологій.

3. Пререквізити. Навчальні дисципліни: “Теорія ймовірностей та математична статистика”, “Бібліотеки мови Python”.

4. Результати навчання. У результаті вивчення навчальної дисципліни студент повинен

знати: предметну область, основні задачі галузі DataMining; методи та засоби обробки та аналізу даних з використанням бібліотек мови Python,

вміти: застосовувати бібліотеки мови Python для розв'язання прикладних задач обробки та аналізу даних різної розмірності.

Дисципліна формує такі **компетенції** у відповідності до стандарту вищої освіти [1] та освітньої програми:

К02. Здатність застосовувати знання у практичних ситуаціях.

К07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

К17. Здатність використовувати системний аналіз як сучасну міждисциплінарну методологію, що базується на прикладних математичних методах та сучасних інформаційних технологіях і орієнтована на вирішення задач аналізу і синтезу технічних, економічних, соціальних, екологічних та інших складних систем.

К22. Здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації, обробки інформації, інтелектуального аналізу даних.

К26. Здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них.

Наведені результати навчання за відповідною дисципліною співвідносяться із такими **програмними результатами навчання:**

ПР12. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

ПР13. Проектувати, реалізовувати, тестувати, впроваджувати, супроводжувати, експлуатувати програмні засоби роботи з даними і знаннями в комп'ютерних системах і мережах.

5. Опис навчальної дисципліни

5.1. Дидактична карта навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Теми лекційних занять	Змістовий модуль 1.											
Тема 1. Математичне моделювання та аналіз даних програмними засобами Python.	13	4	–	4	–	5						
Тема 2. Методи та засоби візуалізації результатів аналізу даних програмними засобами Python.	17	6	–	6	–	5						
Тема 3. Методи та засоби статистичного аналізу даних у бібліотеці SciPy.	15	5	–	5	–	5						
Разом за змістовим модулем 1	45	15	–	15	–	15						
Теми лекційних занять	Змістовий модуль 2.											
Тема 4. Лінійні моделі в задачах інтелектуального аналізу даних.	15	5	–	5	–	5						
Тема 5. Методи визначення якості математичних моделей.	15	5	–	5	–	5						
Тема 6. Ансамблеві математичні моделі.	15	5	–	5	–	5						

Разом за ЗМ 2	45	15	–	15	–	15					
Усього годин	90	30	–	30	–	30					

5.2. Теми лабораторних занять

№	Назва теми
1	Засоби реалізації математичного моделювання та аналізу даних DataMining (NumPy, SciPy, Pandas).
2	Методи та засоби візуалізації результатів аналізу даних (Matplotlib, Pandas, Plotly).
3	Застосування статистичних методів та засобів в задачах аналізу даних (NumPy, SciPy).
4	Використання лінійних моделей в задачах інтелектуального аналізу даних (Scikit-learn).
5	Застосування методів визначення якості моделей в задачах інтелектуального аналізу даних (Scikit-learn, Metrics).
6	Використання ансамблевих математичних моделей у задачах інтелектуального аналізу даних (Scikit-learn).

5.3. Зміст завдань для самостійної роботи

Самостійна робота складає 30 годин. Розподіл самостійної роботи за видами навчальних робіт:

- підготовка до лекційних занять – 10 годин;
- підготовка до лабораторних занять – 15 годин;
- підготовка до підсумкового модуль-контролю – 5 годин.

Самостійна робота складається з повторення матеріалу, засвоєного на лекціях, самостійного опанування частини теоретичного матеріалу, роботи з контрольними запитаннями та завданнями.

Студенти можуть отримати до 5 балів в рахунок ІНДЗ, якщо самостійно зареєструються на безкоштовному курсі платформи Prometheus "Машинне навчання", пройдуть навчання, отримають відповідний сертифікат і надішлють його на сайт дистанційного навчання викладачу разом зі скріншотом успішності на курсі. Кількість балів буде виставлена пропорційно до навчальних результатів студента (згідно зі статистикою сайту Prometheus та після захисту студентом матеріалів про виконання завдань курсу).

Також додаткові бали (до 10 балів за курс з урахуванням прогресу проходження курсу у відсотках та після захисту студентом детального звіту про виконання завдань курсу платформи Coursera і отримання верифікованого сертифікату) можна одержати за проходження курсів платформи Coursera за тематикою DataMining (за попереднім погодженням з викладачем).

6. Система контролю та оцінювання

Види та форми контролю

Формами поточного контролю є усна чи письмова (тестування, реферат, звіт з лабораторних робіт, ІНДЗ) відповідь студента.

Формою підсумкового контролю є залік (7 семестр).

Засоби оцінювання

Усний контроль у вигляді індивідуального та фронтального опитування на лекціях та лабораторних заняттях, захист лабораторних робіт та індивідуального навчально-дослідницького завдання; письмовий контроль у вигляді контрольних робіт, тестів, підсумкове тестове опитування.

Критерії оцінювання результатів навчання з навчальної дисципліни

Звіт з лабораторної роботи, який студенти завантажують на сайт, повинен мати таку структуру:

1) файл у форматі Word, що містить титульний аркуш (назва університету, факультету, кафедри, навчальної дисципліни, хто виконав, хто перевірів, Чернівці - 2023) , назву лабораторної роботи, умову кожного завдання, код програми (не скріншот), скріншоти виконання з демонстрацією різних можливих, в т.ч. особливих, випадків роботи програми;

2) усі файли розробленого студентом проекту, що містять коди програм;

3) всі вищезазначені файли архівуються в один архів, назва архіву має складатися із прізвища студента латиницею, номера лабораторної та номера варіанту, наприклад: Ivanenko_lab_1_var_21.arj.

У разі невиконання студентами вищезазначених правил оформлення кількість отриманих студентом балів при перевірці буде знижуватися. Якщо студент отримав суттєві зауваження до програмного коду і переробив програму та звіт, надіславши його на повторне оцінювання, максимально можлива кількість отриманих ним балів буде щонайменше на 1 (один) бал меншою, ніж максимально можлива кількість балів за цю роботу.

Якщо програма працює і демонструє вірні результати, за це надається 50% балів, передбачених на відповідне завдання. Решту 50% студент може отримати на захисті роботи, відповідаючи на запитання викладача стосовно коду програми, аналізу отриманих результатів, зроблених студентом висновків щодо ходу роботи та ефективності запропонованих методів аналізу даних.

Тестування проводиться з використанням сайту дистанційного навчання (30 питань, 1 бал – 1 питання).

Розподіл балів, які отримують студенти

Поточне оцінювання						Кількість балів (модуль-контроль)	Сумарна к-ть балів
Змістовий модуль №1			Змістовий модуль №2				
T1	T2	T3	T4	T5	T6	ЗМ1, ЗМ2	
10	10	10	10	10	10	10	30
							100

Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
Відмінно	A (90-100)	відмінно
Добре	B (80-89)	дуже добре
	C (70-79)	добре
Задовільно	D (60-69)	задовільно
	E (50-59)	достатньо
Незадовільно	FX (35-49)	незадовільно
	F (1-34)	

7. Рекомендована література

7.1. Основна

1. Стандарт вищої освіти України перший (бакалаврський) рівень, галузь знань 12 – Інформаційні технології, спеціальність 124 – Системний аналіз // Затверджено і введено в дію наказом Міністерства освіти і науки України від 13.11.2018 р. № 1245.– 23 с.
URL: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/12/21/124-sistemn.analiz-bakalavr-1.pdf>
2. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с. ISBN 978-966-641-874-9.
3. Ясинський В.К., Юрченко І.В. Прикладний статистичний аналіз. Методичні рекомендації до лабораторних робіт.– Чернівці: Рута, 2008.– 84 с.
4. Юрченко І.В. Прикладний статистичний аналіз з використанням Python.– Чернівці: Технодрук, 2021.– 102 с.
5. Jake VanderPlas. Python Data Science Handbook. Essential Tools for Working with Data.– Beijing, Boston, Farnham, Tokyo: O’Reilly Media, Inc, 2016.– 576 p.– ISBN: 9-781-491-912-058.
6. Davy Cielen, Arno D.B. Meysman, Mohamed Ali. Introducing data science: BigData, Machine Learning, and more, using Python tools.– New York:

Manning Publications Co, Shelter Island, 2016.– 322 p.– ISBN: 9-781-633-430-037.

7. Бібліотека Scikit-learn мови Python [Електронний ресурс]. Джерело доступу: <https://scikit-learn.org/stable/>
8. Scikit-learn – Вікі [Електронний ресурс]. Джерело доступу: <https://uk.upwiki.one/wiki/Scikit-learn>

7.2. Допоміжна

1. Шевченко О.В. Методи обробки великих масивів даних в пакетному режимі // Магістерська дисертація на здобуття ступеня магістра за освітньо-професійною програмою «Інтелектуальні сервіс-орієнтовані розподілені обчислювання» зі спеціальності 122 «Комп'ютерні науки».– Київ: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", 2020.– 101 с.
2. Гуцуляк І.В. Застосування методів проектування ознак та кластеризації для машинного навчання // Дипломна робота на здобуття ступеня магістра за освітньою програмою "Інформаційні технології та управління проектами" зі спеціальності 122 "Комп'ютерні науки".– Чернівці: Чернівецький національний університет імені Юрія Федьковича, 2022.– 85 с.

8. Інформаційні ресурси

<http://moodle.chnu.edu.ua>

[Matplotlib — Visualization with Python](#)

[Plotly: Low-Code Data App Development](#)

[NumPy](#)

[SciPy](#)

[scikit-learn: machine learning in Python — scikit-learn 1.4.0 documentation](#)

9. Політика освітнього процесу

Здобувач зобов'язаний своєчасно та якісно виконувати всі отримані завдання; за необхідністю з метою з'ясування всіх незрозумілих під час самостійної та індивідуальної роботи питань, відвідувати консультації викладача.

Студенти мають дотримуватись правил академічної доброчесності відповідно до "Кодексу академічної доброчесності ЧНУ". Політика дотримання академічної доброчесності (відповідно до Закону України "Про вищу освіту") – викладання навчальної дисципліни ґрунтується на засадах академічної доброчесності – сукупності етичних принципів та визначених законом правил, якими мають керуватися учасники освітнього процесу під час навчання, викладання та провадження наукової (творчої) діяльності з

метою забезпечення довіри до результатів навчання та/або наукових (творчих) досягнень. Наявність академічного плагіату в студентських роботах є підставою для виставлення негативної оцінки. Списування студентів під час проведення модульної контрольної роботи є підставою для дострокового припинення її складання та виставлення негативної оцінки.

Складання/перескладання екзаменів відбувається за встановленим деканатом розкладом.