

Чернівецький національний університет імені Юрія Федьковича
Факультет математики та інформатики
Кафедра математичного моделювання

“ЗАТВЕРДЖУЮ”
Декан факультету
математики та інформатики


/Мартинюк О.В./
“ 25 ” 06 2024 року

РОБОЧА ПРОГРАМА
навчальної дисципліни

DataMining з використанням мови Python
обов'язкова

Освітньо-професійна програма	Системний аналіз
Спеціальність	124 – Системний аналіз
Галузь знань	12 – Інформаційні технології
Рівень вищої освіти	перший (бакалаврський)
	Факультет математики та інформатики
Мова навчання	українська

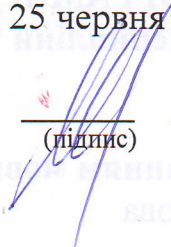
Чернівці 2024 рік

Робоча програма навчальної дисципліни "DataMining з використанням мови Python" складена відповідно до освітньо-професійної програми "Системний аналіз", затвердженої Вченою радою Чернівецького національного університету імені Юрія Федьковича 31 травня 2021 р., протокол № 6 (зі змінами від 29 квітня 2024 р., протокол № 7).

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

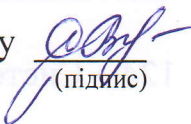
Погоджено з гарантом ОП і затверджено на засіданні кафедри математичного моделювання.

Протокол № 18 від 25 червня 2024 року.

Завідувач кафедри  Черевко І.М.
(підпис) (прізвище та ініціали)

Схвалено методичною радою факультету математики та інформатики.

Протокол № 11 від 25 червня 2024 року.

Голова методичної ради факультету  Сікора В.С.
(підпис) (прізвище та ініціали)

1. 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.

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.

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

Постреквізити. Набуті компетенції можна використовувати при підготовці кваліфікаційних робіт та вивченні навчальних дисциплін ОП "Системний аналіз" (магістр): "Прогнозування в системному аналізі", "Інтелектуальні системи прийняття рішень".

2. 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:

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК1. 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.

ФК6. 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.

ФК10. 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.

3. Опис навчальної дисципліни

3.1. Загальна інформація

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	4	7	3.0	90	30	–	–	30	30	–	залік

3.2. 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						

3.3. Теми лабораторних занять

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

3.4. Зміст завдань для самостійної роботи

Самостійна робота студентів (СРС) 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).

Студенти можуть отримувати до 1 бала в рахунок виконання завдань СРС / ІНДЗ під час кожного лекційного заняття за правильні відповіді на запитання лектора, активне обговорення багатоваріантних підходів до рішення представленої лектором проблеми (для активізації пошукової та дослідної діяльності студентів).

4. Освітні технології, методи навчання і викладання навчальної дисципліни

Методи навчання та викладання: лекції, лабораторні заняття, електронне навчання з використанням системи Moodle, тестування, виконання завдань ІНДЗ.

5. Критерії та засоби оцінювання результатів навчання з навчальної дисципліни

5.1. Критерієм підсумкового оцінювання є досягнення студентом мінімальних порогових рівнів оцінок (балів) за кожним передбаченим результатом навчання.

Мінімальний пороговий рівень оцінки варто визначати за допомогою якісних критеріїв і трансформувати його в мінімальну позитивну оцінку використовуваної числової (рейтингової) шкали.

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 кількості 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			Відвідування занять		
T1	T2	T3	T4	T5	T6	ЗМ1, ЗМ2	30	100
10	10	10	10	10	10	10		

5.2. Шкала оцінювання: національна та ЄКТС

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

5.3. Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання є: стандартизовані тести; аналітичні звіти з лабораторних робіт; реферати; презентації результатів виконаних завдань та досліджень ІНДЗ, усний контроль у вигляді індивідуального та фронтального опитування на лекціях та лабораторних заняттях.

6. Форми поточного та підсумкового контролю

Формами поточного контролю є усна чи письмова (тестування, реферат, лабораторна робота, ІНДЗ) відповідь студента.

Формою підсумкового контролю є залік.

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. Юрченко І.В. DataMining з використанням Python. Навчальний посібник.– Чернівці: Чернівецький національний університет імені Юрія Федьковича, 2024.– 143 с.
<https://archer.chnu.edu.ua/xmlui/handle/123456789/10067>
3. 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.
4. Ясинський В.К., Юрченко І.В. Прикладний статистичний аналіз. Методичні рекомендації до лабораторних робіт.– Чернівці: Рута, 2008.– 84 с.
5. Юрченко І.В. Прикладний статистичний аналіз з використанням Python.– Чернівці: Технодрук, 2021.– 102 с.
6. 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.
7. 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.
8. Бібліотека Scikit-learn мови Python [Електронний ресурс]. Джерело доступу: <https://scikit-learn.org/stable/>
9. 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 проєктування ознак та кластеризації для машинного навчання // Дипломна робота на здобуття ступеня магістра за освітньою програмою "Інформаційні технології та управління проєктами" зі спеціальності 122 "Комп'ютерні науки".– Чернівці: Чернівецький національний університет імені Юрія Федьковича, 2022.– 85 с.

8. Інформаційні ресурси

<http://moodle.chnu.edu.ua>
<https://matplotlib.org/>
<https://plotly.com/>
<https://numpy.org/>
<https://scipy.org/>
<https://scikit-learn.org/stable/>

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.