МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ КОРАБЛЕБУДУВАННЯ
імені адмірала Макарова
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
КОМП'ЮТЕРНИХ НАУК ТА УПРАВЛІННЯ ПРОЕКТАМИ

Код РПНД

ЗАТВЕРДЖЕНО
Проректор з науково-
педагогічної роботи
Слободян С.О.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Program of the Discipline

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В PYTHON
Data Mining in Python

рівень вищої освіти: другий (магістерський)

тип дисципліни: вибіркова

мова(и) викладання: українська

Миколаїв – 2022

Робоча програма навчальної дисципліни "Інтелектуальний аналіз даних в Python".

25 січня 2022 року. – с.

Розробники:
А. Ю. Гайда, доцент кафедри інформаційних управляючих систем та технологій

Проєкт робочої програми навчальної дисципліни "Інтелектуальний аналіз даних в Python" розглянуто на засіданні кафедри інформаційних управляючих систем та технологій.
Протокол № 6 від «26» січня 2022 р.
Завідувач кафедри_____ /І. Л. Михелєв/

Проєкт робочої програми навчальної дисципліни "Інтелектуальний аналіз даних в Python" розглянуто Науково-методичною комісією Навчально-наукового інституту комп'ютерних наук та управління проєктами.
Протокол № 6 від «27» січня 2021 р.
Голова_____ /Т. А. Фаріонова/

Проєкт робочої програми навчальної дисципліни "Інтелектуальний аналіз даних в Python" погоджено з навчальним відділом.
Начальник відділу_____ /А. В. Лабарткава/

Робоча програма навчальної дисципліни "Інтелектуальний аналіз даних в Python" затверджена Навчально-методичною радою НУК.
Протокол № 2 від «23» лютого 2022 р.

Голова НМР НУК_____ /Є. І. Трушляков/

© НУК, 2022

Зміст

Вступ.....4
1. Опис навчальної дисципліни.....6
2. Мета вивчення навчальної дисципліни.....7
3. Передумови для вивчення дисципліни.....7
4. Очікувані результати навчання.....7
5. Програма навчальної дисципліни.....8
6. Методи навчання, засоби діагностики результатів навчання та методи їх демонстрування.....11
7. Форми поточного та підсумкового контролю.....11
8. Критерії оцінювання результатів навчання.....13
9. Засоби навчання.....13
10. Рекомендовані джерела інформації.....14

Вступ

Анотація

Інтелектуальний аналіз даних в інформатиці (Видобуток даних, Data Mining) – процес виявлення корисних закономірностей та взаємозв’язків у великих наборах даних з використанням інструментів математичної статистики та штучного інтелекту (такого, як штучні нейронні мережі (ШНМ) та машинне навчання). Видобуток даних широко використовується у бізнесі (менеджмент, банківська справа, торгівля, страхування, реклама), наукових дослідженнях (матеріалознавство, медицина), державній безпеці (захист інформаційних систем), системах автоматизації (розпізнавання образів та голосу) тощо. Володіння сучасними інструментальними засобами і технологіями інтелектуального аналізу даних підвищує конкурентоспроможність фахівців на ринку праці.

Дисципліна ознайомлює студентів з сучасними методами і засобами аналізу даних та практичним застосуванням інструментального забезпечення мови програмування Python при вирішенні зазначених задач.

Завданням дисципліни є забезпечення теоретичної і практичної підготовки фахівців для розв’язання завдань розробки та супроводу корпоративних і спеціалізованих комп’ютерних інформаційних систем. Опанування курсу надає професійні компетенції розв’язувати складні спеціалізовані задачі та практичні проблеми в області комп’ютерних наук при здійсненні професійної діяльності, що передбачає проведення досліджень: знання сучасних інструментів первинної обробки і аналізу та візуалізації даних. Доцільність вивчення дисципліни підтверджується стрімким розвитком технологій розробки систем штучного інтелекту і широким застосуванням для вирішення зазначених задач мови програмування Python.

Ключові слова: комп’ютерні програми, видобуток даних, аналіз баз даних, інтелектуальний аналіз, математична модель, машинне навчання.

Abstract

Data Mining is the process of identifying useful patterns and relationships in large data sets using mathematical statistics and artificial intelligence tools (such as neural networks and machine learning). Data extraction is widely used in business (management, banking, trade, insurance, advertising), research (materials science, medicine), national security (protection of information systems), automation systems (image and voice recognition), etc. Possession of modern tools and technologies of data mining increases the competitiveness of professionals in the labor market.

The course acquaints students with modern methods and tools of data analysis and the practical application of instrumental software Python programming language in solving these problems.

The task of the discipline is to provide theoretical and practical training for solving problems of development and maintenance of corporate and specialized computer information systems. Mastering the course provides professional competencies to solve complex specialized problems and practical problems in the field of computer science in the implementation of professional activities involving research: knowledge of modern tools for primary processing and analysis and visualization of data. The expediency of studying the discipline is confirmed by the rapid development of technologies for the development of artificial intelligence systems and the widespread use of Python programming language to solve these problems.

Keywords: computer programs, data mining, database analysis, intelligent analysis, mathematical model, machine learning.

1. Опис навчальної дисципліни

<table>
<tr><td rowspan="2">Найменування показників</td><td rowspan="2">Галузь знань, спеціальність (освітня програма), освітній рівень</td><td colspan="2">Характеристика навчальної дисципліни</td></tr>
<tr><td>Денна форма навчання</td><td>Заочна форма навчання</td></tr>
<tr><td>Кількість кредитів – 5</td><td></td><td colspan="2">Вибіркова</td></tr>
<tr><td>Модулів – 1</td><td rowspan="5"></td><td colspan="2">Рік підготовки:</td></tr>
<tr><td>Змістових модулів – 5</td><td rowspan="2">1-й</td><td rowspan="2">1-й</td></tr>
<tr><td>Електрона адреса РПНД на сайті Університету:</td></tr>
<tr><td rowspan="2">Загальна кількість годин – 150</td><td colspan="2">Семестри</td></tr>
<tr><td>2-й</td><td>2-й</td></tr>
<tr><td rowspan="14">Тижневих годин для денної форми навчання: аудиторних – 4; самостійної роботи студента – 6.</td><td rowspan="14">Освітній рівень: другий (магістерський)</td><td colspan="2">Лекції</td></tr>
<tr><td>30 год</td><td>6 год</td></tr>
<tr><td colspan="2">Практичні заняття</td></tr>
<tr><td>30 год</td><td>6 год</td></tr>
<tr><td colspan="2">Самостійна робота</td></tr>
<tr><td>90 год</td><td>138 год</td></tr>
<tr><td colspan="2">Індивідуальні завдання</td></tr>
<tr><td>–</td><td>–</td></tr>
<tr><td colspan="2">Вид контролю: залік</td></tr>
<tr><td colspan="2">Форма контролю: комбінована (письмовий контроль, тестовий контроль)</td></tr>
</table>

2. Мета вивчення навчальної дисципліни

Метою вивчення навчальної дисципліни "Інтелектуальний аналіз даних Python" є ознайомлення студентів з основними відомостями щодо методів та засобів аналізу даних, побудови моделей аналізу, підготовки даних, візуалізації результатів аналізу, розробки прикладного програмного забезпечення аналізу даних з використанням засобів аналізу даних мови програмування Python, формування у студентів наступних **професійних компетентностей**:

- здатність виявляти і застосовувати інформаційні технології, які можуть бути застосовані для вирішення існуючих, нових або потенційних проблем підприємств;
- здатність до впровадження та застосування інструментального програмного забезпечення;
- здатність розробляти комп’ютерні моделі об’єктів і процесів інформатизації;
- здатність розробляти та впроваджувати сховища даних інформаційних систем, використовувати методи попередньої обробки і аналізу даних;
- розуміння основних напрямів розвитку корпоративних інформаційних систем і інструментального програмного забезпечення.

3. Передумови для вивчення дисципліни

Передумов немає.

4. Очікувані результати навчання

У результаті вивчення навчальної дисципліни студент отримає наступні знання та вміння, що забезпечать зазначені у меті професійні компетентності:

Знання сучасних методів та інструментальних засобів, пошуку у необроблених масивах даних невідомих корисних знань та закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів та різних інструментальних засобів, що використовуються для класифікації і кластеризації даних та при розробці інтелектуальних систем.

Вміння практичного застосуванням інструментального забезпечення аналізу даних при вирішенні задач розробки і експлуатації систем збору і обробки даних, класифікації і кластеризації, генерації звітів; здійснювати підготовку інформації, формулювати завдання для реалізації проєктів і програм збору і обробки даних; створювати моделі програмних проектів для автоматизації первинної обробки і аналізу даних; здійснювати первинну обробку даних засобами мови Python; використовувати інструментальні засоби візуалізації даних мови Python.

5. Програма навчальної дисципліни

Модуль 1. Інтелектуальний аналіз даних

Змістовий модуль 1. Технології і методи машинного навчання

Тема 1. Загальні відомості про машинне навчання. Інтелектуальний аналіз даних. Технології і методи інтелектуального аналізу даних. [2, 3]

Тема 2. Методи збору і первинної обробки даних. Методи візуалізації багатовимірних наборів даних. [2, 3]

Змістовий модуль 2. Засоби збору і первинної обробки даних в Python

Тема 3. Засоби збору і представлення даних машинного навчання у Python. Первинна обробка структурованих даних в Python. [1, 4]

Тема 4. Засоби візуалізації багатовимірних наборів даних і результатів їх аналізу в Python. [1, 4]

Змістовий модуль 3. Інтелектуальний аналіз даних засобами ШНМ в Python

Тема 5. Засоби побудови ШНМ в Python. Проблеми навчання і перенавчання ШНМ. Регуляризація ШНМ. [2, 3]

Тема 6. Особливості побудови ШНМ в Python в задачах класифікації і кластеризації, регресії і зниження розмірності даних. [2, 3]

Змістовий модуль 4. Машинне навчання в прикладних задачах оптимізації і моделювання

Тема 7. Засоби, моделі і методи машинного навчання в задачах оптимізації і моделювання. [2, 4]

Тема 8. Машинне навчання і глибокі нейронні мережі в Python. [2, 4]

Тематичний план навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	Денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	ср	інд		л	п	лаб	ср	інд
1	2	3	4	5	6	7	8	9	10	11	12	13
Змістовий модуль 1. Технології і методи машинного навчання												
Тема 1. Загальні відомості про машинне навчання. Інтелектуальний аналіз даних. Технології і методи інтелектуального аналізу даних.	**14**	4	4	0	6	0	**30**	2	2	0	26	0
Тема 2. Методи збору і первинної обробки даних. Методи візуалізації багатовимірних наборів даних.	**16**	4	4	0	8	0						
Разом за ЗМ 1	**30**	**8**	**8**	**0**	**14**	0	**30**	**2**	**2**	**0**	**26**	0
Змістовий модуль 2. Засоби збору і первинної обробки даних в Python												
Тема 3. Засоби збору і представлення даних машинного навчання у Python. Первинна обробка структурованих даних в Python.	**23**	4	4	0	15	0	**45**	2	2	0	41	0
Тема 4. Засоби візуалізації багатовимірних наборів даних і результатів їх аналізу в Python.	**22**	4	4	0	14	0						
Разом за ЗМ 2	**45**	**8**	**8**	**0**	**29**	**0**	**45**	**2**	**2**	**0**	**41**	**0**
Змістовий модуль 3. Інтелектуальний аналіз даних засобами ШНМ в Python												
Тема 5. Засоби побудови ШНМ в Python. Проблеми навчання і перенавчання ШНМ. Регуляризація ШНМ.	**18**	4	4	0	10	0	**45**	1	1	0	43	0
Тема 6. Особливості побудови ШНМ в Python в задачах класифікації і кластеризації, регресії і зниження розмірності даних.	**27**	4	4	0	19	0						
Разом за ЗМ 3	**45**	**8**	**8**	**0**	**29**	**0**	**45**	1	**1**	**0**	**43**	**0**
Змістовий модуль 4. Машинне навчання в прикладних задачах оптимізації і моделювання												
Тема 7. Засоби, моделі і методи машинного навчання в	**20**	4	4	0	12	0	**30**	1	1	0	26	0

Назви змістових модулів і тем	Кількість годин											
	Денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	ср	інд		л	п	лаб	ср	інд
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
задачах оптимізації і моделювання**.**												
Тема 8. Машинне навчання і глибокі нейронні мережі в Python.	**10**	2	2	0	6	0						
Разом за ЗМ 4	**30**	**6**	**6**	**0**	**18**	**0**	**30**	1	**1**	**0**	**26**	**0**
Усього годин	**150**	**30**	**30**	**0**	**90**	**0**	**150**	6	6	**0**	**138**	**0**

Теми практичних занять

Студентом виконуються практичні роботи відповідно до теми атестаційної магістерської роботи:

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
1	Збір і первинна обробка даних.	8	2
2	Регресія в ШНМ	8	2
3	Класифікація і кластеризація в ШНМ	8	1
4	Машинне навчання і оптимізація	6	1
	Разом	**30**	**6**

Самостійна робота

Розподіл часу самостійної роботи

№ з/п	Вид роботи	Кількість годин	
		денна форма	заочна форма
1	Підготовка до лекцій	30	15
2	Підготовка до практичних робіт	30	65
3	Підготовка до поточних модульних контрольних робот	30	58
	РАЗОМ	**90**	**138**

6. Методи навчання, засоби діагностики результатів навчання та методи їх демонстрування

Методами навчання є:

– проведення лекцій та практичних занять з поясненням нових понять, розкриттям та деталізацією матеріалу, контрольними питаннями по ходу проведення лекцій та практичних занять та відповідями на запитання, наданням рекомендацій щодо вибору додаткових джерел;

– демонстрація прикладів програмного коду та моделей, що ілюструють окремі поняття та співвідношення між ними;

– розробка спрощених рішень на основі наведених прикладів для підготовки до виконання практичних завдань.

Засобами оцінювання та методами демонстрування результатів навчання з дисципліни "Кросплатформне програмування" є:

– звіти з виконання лабораторних робіт та/або презентації результатів виконаних завдань та досліджень (письмовий контроль);

– поточні модульні контрольні роботи у формі тестування (тестовий контроль);

– студентські презентації та виступи, тези доповідей, статті відповідної до дисципліни тематики на наукових заходах (письмовий контроль).

7. Форми поточного та підсумкового контролю

Досягнення студента оцінюються за 100-бальною системою Університету.

Підсумкова оцінка навчального курсу включає в себе суму оцінок поточного контролю. У кожного компонента своя частка в загальній системі оцінок.

Поточний контроль проводиться на кожному практичному занятті та за результатами виконання завдань самостійної роботи. Він передбачає оцінювання теоретичної підготовки здобувачів вищої освіти із зазначеної теми (у тому числі, самостійно опрацьованого матеріалу) під час роботи на практичних заняттях.

Зарахування кредитів навчального курсу можливо тільки після досягнення результатів, запланованих РПНД, що виражається в одній з позитивних оцінок, передбачених чинними нормативними документами.

Форми контролю результатів навчальної діяльності студентів та їх оцінювання

Поточний контроль виконання етапів комплексної практичної роботи (письмовий контроль)

Виконання кожної з практичних робіт має супроводжуватися звітом та/або презентацією результатів виконаних завдань та досліджень, допускаються для зарахування етапу (етапів) студентські презентації та виступи, тези доповідей, статті відповідної до дисципліни тематики на наукових заходах.

Оцінювання практичних робіт Бал	**Критерії оцінювання**
15	Робота виконана у встановлений термін. Виконана самостійно, чітко сформульовані цілі, завдання та гіпотеза досліджень. Застосовувалися коректні методи обробки отриманих результатів. У висновках проведена коректна інтерпретація результатів.
12	Робота виконана у встановлений термін. Студент виконує практичну роботу згідно з інструкцією, іноді після консультації викладача; описує спостереження; в цілому правильно складає звіт та робить висновки.
9	Робота виконана з порушенням встановлених термінів. Студент виконує практичну роботу згідно з інструкцією, іноді після консультації викладача; описує спостереження; складає звіт, що містить неточності у висновках та помилки.
6	Робота виконана з порушенням встановлених термінів. Студент виконує практичну згідно з інструкцією; складений звіт містить неточності у висновках та помилки.
3	Робота виконана з порушенням встановлених термінів. Студент виконує практичну під керівництвом викладача; складений звіт містить неточності у висновках та помилки.
0	Робота не виконувалася.

Максимальна оцінка студента за даний компонент програми 60 балів (4 етапи комплексної практичної роботи по 15 балів).

Поточний модульний контроль (тестовий контроль)

Правильних відповідей,%	100	90	80	70	60	50	40	30	20	10
Бал	10	9	8	7	6	5	4	3	2	1

По завершенні змістовного модуля здійснюється поточний модульний контроль – тестування.

Форма контролю	Максимальна кількість балів	
	Денна форма	Заочна форма
Звіти з практичних робіт (письмовий контроль)	4х15балів = 60 балів	4х15балів = 60 балів
Поточний модульний контроль (тестовий контроль)	4х10балів = 40 балів	4х10балів = 40 балів
Всього	**100**	**100**

8. Критерії оцінювання результатів навчання

		Денна форма		Заочна форма	
		Вид роботи	**Бали**	**Вид роботи**	**Бали**
ЗМ 1	Т 1	Практична робота № 1	15	Практична робота № 1	15
	Т 2	Поточний модульний контроль	10	Поточний модульний контроль	10
ЗМ 2	Т 3	Практична робота № 2	15	Практична робота № 2	15
	Т 4	Поточний модульний контроль	10	Поточний модульний контроль	10
ЗМ 3	Т 5	Практична робота № 3	15	Практична робота № 3	15
	Т 6	Поточний модульний контроль	10	Поточний модульний контроль	10
ЗМ 4	Т 7	Практична робота № 4	15	Практична робота № 4	15
	Т 8	Поточний модульний контроль	10	Поточний модульний контроль	10

9. Засоби навчання

Технічні засоби навчання: персональні комп’ютери з підключенням до мережі Інтернет, мультимедійний проектор.

Комп’ютерні засоби навчання: інтерпретатор Python з пакетами математичної обробки і аналізу даних, табличний процесор з пакету LibreOffice або Microsoft Office, потоковий редактор AWK, бібліотеки SciPy, Tkinter.

10. Рекомендовані джерела інформації

Основна література

1. Підручник з Python [Електронний ресурс] / Режим доступу: \www/ URL: https://docs.python.org/uk/3/tutorial/index.html
2. Рассел, С. Штучний інтелект: сучасний підхід . Том 1. Рішення проблем: знання і міркування. 2-е вид. / С. Рассел, П. Норвіг. – К.: Діалектика. 2020. – 1408 с.
3. Rundell, Michael. Data Mining in Python: A Guide [Електронний ресурс] / Режим доступу: \www/ URL: https://www.springboard.com/blog/data-science/data-mining-python-tutorial/
4. Kumar, Ashutosh. Python Tools for Data Mining: Harnessing the Power of Python for Extracting Insights [Електронний ресурс] / Режим доступу: \www/ URL: https://www.taazaa.com/python-tools-for-data-mining/

Допоміжна література

5. Єремєєв, В.С. Теорія ймовірностей та математична статистика / В.С. Єремєєв, Д.О. Сосновських, О.В. Тітова. – Навчальний посібник: Мелітополь: ТОВ «Видавничий будинок ММД», 2009. – 188 с.
6. Discovering and Visualizing Patterns with Python [Електронний ресурс] / Режим доступу: \www/ URL: https://dzone.com/refcardz/data-mining-discovering-and

Інформаційні ресурси

1. Конспекти лекцій з дисципліни.
2. Підручники та навчальні посібники.

Розробник:
доцент кафедри інформаційних
управляючих систем та технологій,
к.т.н. А. Ю. Гайда