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

BIG DATA
Big data

рівень вищої освіти: другий (магістерський)

тип дисципліни: вибіркова

мова(и) викладання: українська

Миколаїв – 2022

Робоча програма навчальної дисципліни "Big Data".

26 грудня 2022 року. –16 с.

Розробник:
О. В. Гайдаєнко доцент кафедри інформаційних управляючих систем та технологій

Проєкт робочої програми навчальної дисципліни "Big Data" розглянуто на засіданні кафедри інформаційних управляючих систем та технологій.
Протокол № 5 від «27» грудня 2022 р.
Завідувач кафедри_____/І. Л. Михелєв/

Проєкт робочої програми навчальної дисципліни "Big Data" розглянуто Науково-методичною комісією Навчально-наукового інституту комп'ютерних наук та управління проєктами.
Протокол № 5 від «27» грудня 2022 р.
Голова_____/Т. А. Фаріонова/

Проєкт робочої програми навчальної дисципліни "Big Data" погоджено з навчальним відділом.
Начальник відділу_____/А. В. Лабарткава/

Робоча програма навчальної дисципліни "Big Data" затверджена Навчально-методичною радою НУК.
Протокол № 10 від «28» грудня 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. Засоби діагностики результатів навчання та методів їх демонстрування 10
7. Форми поточного та підсумкового контролю 11
8. Критерії оцінювання результатів навчання 14
9. Засоби навчання 14
10. Рекомендовані джерела інформації 15

Вступ

Аннотація

Big Data (великі дані) – це поєднання структурованих, напівструктурованих та неструктурованих даних, які можуть бути видобуті для отримання інформації та використані в проектах машинного навчання, прогнозного моделювання та інших передових програм аналітики. Обсяги неоднорідної цифрової інформації, що швидко надходить, обробити традиційними інструментами неможливо. Сам аналіз даних дозволяє побачити певні та непомітні закономірності, які не може побачити людина. Це дозволяє оптимізувати всі сфери нашого життя – від державного управління до виробництва та телекомунікацій. Завдання сучасної епохи полягає в тому, щоб розібратися в це море даних. Саме тут аналітика великих даних входить у свої права. Big Data Analytics в основному включає збор даних з різних джерел, зміна їх таким чином, щоб вони стали доступними для використання аналітиками, і, нарешті, надання продуктів даних, корисних бізнесу. Навчальний курс формуватиме певну систему знань у студентів у галузі інформаційних технологій: про особливості зберігання та обробки інформації; сутності поняття Big Data та його значення; принципи та підходи до управління Big Data, а також сучасних технологіях зберігання даних та інструментаріях для їх аналізу.

Ключові слова: неструктуровані дані, великий обсяг даних, достовірність, мінливість, цінність, широкий вибір типів даних, генерування даних.

Abstract

Big Data is a combination of structured, semi-structured and unstructured data that can be mined for information and used in machine learning, predictive modeling and other advanced analytics applications. It is impossible to process the volumes of non-homogeneous digital information coming in quickly with traditional tools. Data analysis itself allows to see certain and imperceptible regularities that cannot be seen by a person. This allows us to optimize all areas of our life - from public administration to production and telecommunications. The challenge of the modern era is to make sense of this sea of data. This is where big data analytics comes into its own. Big Data Analytics basically involves gathering data from various sources, transforming it so that it becomes usable by analysts, and finally providing data products useful to the business. The training course will form a certain system of knowledge among students in the field of information technologies: about the peculiarities of information storage and processing; the essence of the concept of Big Data and its meaning; principles and approaches to Big Data management, as well as modern data storage technologies and tools for their analysis.

Keywords: unstructured data, big data, validity, variability, value, wide variety of data types, data generation.

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 – 1</td><td rowspan="2">1-й</td><td rowspan="2">1-й</td></tr>
<tr><td>Електроний адрес 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="12">Тижневих годин для денної форми навчання:
аудиторних – 4;
самостійної роботи студента – 6.</td><td rowspan="12">Освітній рівень: другий (магістерський)</td><td colspan="2">Лекції</td></tr>
<tr><td>30 год</td><td>10 год</td></tr>
<tr><td colspan="2">Практичні заняття</td></tr>
<tr><td>30 год</td><td>10 год</td></tr>
<tr><td colspan="2">Самостійна робота</td></tr>
<tr><td>90 год</td><td>130 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

Мета вивчення навчальної дисципліни «Big Data» - ознайомлення та формуваня у студентів фундаментальних знань з теорії та практики в області розробки і використання систем обробки і аналізу великих масивів даних.

Результатом вивчення дисципліни є формування наступних **здатностей**:

• аналізувати, верифікувати, оцінювати повноту інформації в ході професійної діяльності, при необхідності заповнювати та синтезувати недостатню інформацію, реалізовувати моделі та алгоритми прикладний математики у вигляді комп'ютерних програм;

• керування підготовкою та проведенням кількісних та якісних досліджень, розробляти плани, програми та інші матеріали прогнозно-аналітичного характеру.

• ставити завдання дослідження, вибирати методи експериментальної роботи, готувати базу для наукових досліджень.

3. Передумови для вивчення дисципліни

Передумови для вивчення дисципліни відсутні.

4. Очікувані результати навчання

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

Знання теоретичні основи методів аналізу великих даних та машинного навчання; типи завдань, які вирішуються з використанням методів аналізу великих даних та машинного навчання; сучасні технології аналізу великих даних та машинного навчання.

Вміння ставити завдання та розробляти програму дослідження з використанням аналізу великих даних та машинне навчання; обирати відповідні методи вирішення експериментальних та теоретичних завдань; застосовувати методи та підходи аналізу великих даних та машинного навчання для вирішення експериментальних та теоретичних завдань; інтерпретувати, представляти та застосовувати результати, отримані з використанням аналізу великих даних та

машинного навчання..

5. Програма навчальної дисципліни

Модуль 1. Великі дані

Змістовий модуль 1. Введення у машинне навчання та обробку даних

Тема 1. Введення у машинне навчання та обробку даних. Програмний інструментарій Data Mining та Machine Learning. [1] стор. 8-48; 135-286, [2]стор. 5-99, [7].

Тема 2. Основний апарат комбінаторики та математичної статистики. Регресійний аналіз та стиск даних. [3] стор. 187-300, [4] 101-404, [6].

Тема 3. Детектування викидів та аномалій. Очищення даних та технології регуляризації.. [5] стор. 28-259, 125-145

Тема 4. Технології кластеризації та класифікації. Нейронні мережі. Генетичні алгоритми. [1] стор. 99-200, [11].]

Тема 5. Виділення особливостей (Feature detection); нормалізація даних. Нечіткі множини. Байєсові мережі. [9] стор. 92-266, [6], [7], [8]

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. Введення у машинне навчання та обробку даних												
Тема 1. Введення у машинне навчання та обробку даних. Програмний інструментарій Data Mining та Machine Learning	**30**	6	6	0	18	-	**30**	2	2	0	26	-
Тема 2. Основний апарат комбінаторики та математичної статистики.	**30**	6	6	0	18	-	**30**	2	2	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
Регресійний аналіз та стиск даних												
Тема 3. Детектування викидів та аномалій. Очищення даних та технології регуляризації.	**30**	6	6	0	18	-	**30**	2	2		26	
Тема 4. Технології кластеризації та класифікації. Нейронні мережі. Генетичні алгоритми	**30**	6	6	0	18	-	**30**	2	2	0	26	-
Тема 5. Виділення особливостей (Feature detection); нормалізація даних. Нечіткі множини. Байєсові мережі	**30**	6	6	0	18	-		2	2		26	
Разом за ЗМ 1	**150**	**30**	**30**	**0**	**90**	**0**	**30**	**10**	**10**	**0**	**130**	**0**
Усього годин	**150**	**30**	**30**	**0**	**90**	**0**	**150**	**10**	**0**	**10**	**130**	**0**

Теми лабораторних занять

Лабораторні заняття непередбачені навчальним планом.

Теми практичних занять

Студентом виконується комплексна практична робота, відповідно до теми атестаційної магістерської роботи. Етапи комплексної практичної роботи за змістовними модулями:

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
1.	Ознайомлення з інтерфейсом користувача аналітичної платформи Deductor. Експорт, імпорт і візуалізація даних. Очищення даних. Сховище даних	4	2
2.	OLAP-технологія аналізу даних, Автокореляційний аналіз часового ряду. ABCD-аналіз даних.	6	2
3.	XYZ-аналіз даних. Прогнозування часового ряду за допомогою нейронної мережі.	6	2
4. .	Регресійне прогнозування часового ряду. Класифікація даних за допомогою дерев рішень. Кластеризація даних ієрархічними методами.	6	2
5.	Кластеризація даних за допомогою самоорганізуючихся карт Кохонена . Пошук асоціативних правил	8	2
	Разом	**30**	**10**

Самостійна робота

Розподіл часу самостійної роботи

№ з/п	Вид роботи	Кількість годин	
		денна форма	заочна форма
1	Підготовка до лекцій	15	15
2	Підготовка до практичних робіт	35	60
3	Підготовка до поточних модульних контрольних робот	40	55
	РАЗОМ	**90**	**130**

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

Набуття здобувачами компетенцій забезпечується через засвоєння навчального матеріалу у вигляді лекційних, практичних занять, самостійної роботи студентів та контрольних тестових робіт програмованого типу.

Основними **методами навчання** з дисципліни є:

1. **Пояснювально-ілюстративний метод**, за допомогою повідомлення та засвоєння інформації на лекційних та практичних заняттях словесними й наочними засобами.

2. **Проблемний метод**, коли на практичних заняттях утворюються пошукові ситуації, розвивається активність, самостійність, творчі здібності здобувачів освіти.

Як наочний матеріал на лекціях застосовуються мультимедійні слайди та лекційні демонстрації.

Для зручної та ефективної організації навчального процесу, зроблено ClassRoom з навчальної дисципліни, де розташовані завдання до практичних робіт та супроводжувальні матеріали. Доступ до класу надається на першому практичному занятті за університетською електронною поштою.

Засобами оцінювання та методами демонстрування результатів навчання з дисципліни "Сучасні засоби колективної розробки інформаційних систем" є:

– звіти з виконання етапів комплексної практичної роботи 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 чинним законодавством.

Форми контролю результатів навчальної діяльності студентів та їх оцінювання

Поточний контроль виконання етапів комплексної практичної роботи (письмовий контроль)

Виконання кожного з етапів комплексної практичної роботи має супровожуватися 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**
5	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.
4	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.
3	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.
2	Робота виконана з порушенням встановлених термінів. Студент виконує практичну згідно з інструкцією; складений звіт містить неточності у висновках та помилки.
1	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 25 балів (5 етапів комплексної практичної роботи по 5 балів).

Поточний модульний контроль (тестовий контроль)

Правильних відповідей,%	100	90	80	70	60	50	40	30	20	10	5
Бал	15	14	12	11	9	8	6	5	3	2	1

По завершенні змістовного модуля здійснюється поточний модульний контроль – тестування.

Форма контролю	**Максимальна кількість балів**	
	Денна форма	**Заочна форма**
Звіти з етапів комплексної практичної роботи (письмовий контроль)	5 х 5 балів = 25 балів	5 х 5 балів = 25 балів
Поточний модульний контроль (тестовий контроль)	5 х 15 балів = 75 балів	5 х 15 балів = 75 балів
Всього	**100**	**100**

8. Критерії оцінювання результатів навчання

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

9. Засоби навчання

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

Комп’ютерні засоби навчання: Deductor.

10. Рекомендовані джерела інформації

Основна література

1. Nathan Marz. Big Data. Principles and best practices of scalable real-time data systems. 2015 by Manning Publications Co. 330 p.

2. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Hardcover – February 19, 2013. 450 p.

3. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных / Год издания: 2017 Издательство: Питер 336 с.

4. Rajat Mehta. Big Data Analytics with Java. 2017 Packt Publishing. 418 p.

5. Business Analytics for Managers: Taking Business Intelligence Beyond. Gert H. N. Laursen Data Smart: Using Data Science to Transform Information into Insight John W. Foreman.

Допоміжна література

6. Handbook of Statistical Analysis and Data Mining Applications. John Elder IV

7. Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms) Daniel Covington

8. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. - Сиб.: Питер, 2001.

9. Глибовець М.М., Отецький О.В. Штучний інтелект. Підручник. - К: Вид.дім "КМ Академія", 2002, - 366с.

10. Джексон П. Введение в экспертные системы. М.: Изд.дом „Вильяме", 2001.

11. Лорьер Ж.-Л. Системы искусственного интеллекта. - М.: Мир, 1991,- 568с.

Інформаційні ресурси

12. Конспекти лекцій з дисципліни.
13. Підручники та навчальні посібники.

Розробник
доцент кафедри інформаційних
управляючих систем та технологій /О. В.Гайдаєнко/