

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «МЕТОДИ ТА ТЕХНІКИ АНАЛІЗУ ВЕЛИКИХ ДАНИХ»



Ступінь освіти	магістр
Освітня програма	Комп'ютерна інженерія
Тривалість викладання	3, 4 чверті
Заняття:	II семестр 2020/2021 н.р.
лекції:	1 година
лабораторні заняття:	2 години
Мова викладання	українська

Сторінка курсу в СДО НТУ «ДП»:

Інші додаткові ресурси: <https://www.netacad.com/courses/cybersecurity/ccna-security>

Кафедра, що викладає Інформаційних технологій та комп'ютерної інженерії

Інформація про викладача:



Викладач:

Кожевніков Антон Вячеславович

Доц. кафедри

Персональна сторінка

https://it.nmu.org.ua/ua/HR_staff/prepods/kozhevnykov.php

E-mail:

kozhevnykov.a.v@nmu.one

1. Анотація до курсу

Дані всюди. Важко уявити, скільки нових даних генерується кожен день. Дані можуть бути словами в книзі, вмістом електронної таблиці, зображеннями, відео, аудіо або потоком вимірювань, відправленим з пристрою моніторингу. Необроблені дані мають мало користі. Ми повинні обробити ці дані, а потім інтерпретувати вивід, щоб зробити їх корисним. Ці корисні дані тепер перетворилися в інформацію.

Коли даних так багато, що традиційні способи їх обробки, зберігання та аналізу неможливо використовувати, це називається великими даними. Великі дані вимагають нових методів та інструментів, щоб зробити їх значущими. Цей курс знайомить студентів з цими методами та інструментами, які допоможуть використовувати всю міць великих даних.

2. Мета та завдання курсу

Мета дисципліни – вивчення аналітичних методів та інструментів, які використовуються для обробки і візуалізації великих даних.

Предметом вивчення дисципліни є технології, методи та засоби обробки і візуалізації великих даних.

Завдання вивчення дисципліни:

- оволодіння основними поняттями інтелектуального аналізу даних;
- ознайомлення з новітніми інформаційними технологіями аналітичної обробки інформації;
- набуття практичних навичок використання методів і засобів інтелектуального аналізу даних.

3. Результати навчання

Студенти

Знають: основні поняття, методи, засоби, моделі та алгоритми інтелектуального аналізу великих даних.

Розуміють: принципи застосування технологій інтелектуального аналізу даних, перспективи і проблеми аналітики великих даних в зв'язку з Інтернетом речей.

Уміють: вільно орієнтуватися на сучасному ринку аналітичних програмних продуктів; використовувати Jupyter Notebooks для створення конвеєра даних для збору, аналізу та візуалізації даних; застосовувати моделі машинного навчання для автоматизації завдань.

Компетенції:

- студент спроможний розрізняти структуровані дані та неструктуровані дані;
- студент спроможний аналізувати дані за допомогою Python і SQLite;
- студент спроможний аналізувати дані, використовуючи основні статистичні методи і методи підготовки даних в Python за допомогою pandas;
- студент спроможний аналізувати дані за допомогою моделей машинного навчання.

4. Структура курсу

Види робіт за навчальним планом	Години
Аудиторні заняття, у т.ч.:	57
Лекції	17
Лабораторні роботи	34
Практичні заняття	-
Контрольні заходи	6
Самостійна робота, у т.ч.:	63
Підготовка до аудиторних занять	59
Підготовка до заліку	4
Всього:	120 (4 кредити)
Форма підсумкового контролю	залік

Назви змістових модулів і тем лекцій	Кількість годин			
	усього	у тому числі		
		лекції	лабор	самост. робота
Зм. модуль 1. Великі дані та Інтернет речей				
Пояснення значення важливості даних	7	1	2	4
Основні технології управління даними	7	1	2	4
Зм. модуль 2. Основи аналізу даних				
Моделі аналітики	7	1	2	4
Види аналізу даних	7	1	2	4
Життєвий цикл аналізу даних	7	1	2	4
Джерела даних та їх підготовка	7	1	2	4
Етика великих даних	7	1	2	4
Зм. модуль 3. Задачі інтелектуального аналізу даних				
Дослідницький аналіз даних	7	1	2	4
Статистичний аналіз	7	1	2	4
Аналіз з використання описової статистики	9	1	2	6
Аналіз за допомогою кореляції	13	2	4	7
Зм. модуль 4. Розширена аналітика даних і машинне навчання				
Прогностична аналітика	7	1	2	4
Оцінка моделі	7	1	2	4
Зм. модуль 5. Візуалізація великих даних				
Використовувати бібліотек Python для створення графічних відображень	7	1	2	4
Зм. модуль 6. Архітектура великих даних і інженерія даних				
Аналітика масштабування даних	7	1	2	4
Вступ до інженерії даних	7	1	2	4
Усього годин за рік	120	17	34	69

ЛАБОРАТОРНІ ЗАНЯТТЯ

Назва теми	Кількість годин
1. Базовий аналіз даних	7
2. Робота з Python та SQLite	7
3. Описова статистика на Python	8
4. Кореляційний аналіз на Python	8
5. Лінійна регресія на Python	8
6. Класифікація дерева рішень	8
7. Оцінка помилок відповідності в лінійній регресії	8
8. Лінійна регресія даних Інтернет-трафіку	8
9. Виявлення аномалії Інтернет-трафіку	8
10. Візуалізація даних у Excel	6
Всього	76

5 Технічне обладнання та/або програмне забезпечення

Технічні засоби навчання.
Комп'ютерний, клас.
Комп'ютерний клас з доступом до Internet.
Raspberry Pi.
Програмне забезпечення Excel.
Програмне забезпечення Jupyter Notebooks
Програмне забезпечення Python та бібліотеки: pandas, numpy, matplotlib, seaborn, sqlite3,
Дистанційна платформа MOODL.
Дистанційна платформа netacad.com.

6. Система оцінювання та вимоги

6.1. Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
75-89	добре
60-74	задовільно
0-59	незадовільно

6.2. Здобувачі вищої освіти можуть отримати **підсумкову оцінку** з навчальної дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів з поточного тестування та самостійної роботи складатиме не менше 60 балів.

Поточна успішність складається з оцінок за лекційну частину курсу та лабораторний практикум. Отримані бали додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за поточною успішністю здобувач вищої освіти може набрати 100 балів.

Максимальне оцінювання:

Теоретична частина	Лабораторна частина		Разом
	При своєчасному складанні	При несвоєчасному складанні	
50	50	30	100

Лабораторні роботи приймаються за результатами їх виконання та захисту.

Теоретична частина оцінюється за результатами здачі білету диференційного заліку, який містить 25 тестових питань.

6.3. Критерії оцінювання підсумкової роботи

Під час проведення диференційного заліку наприкінці четвертої чверті здобувачі вищої освіти складають відповідні тести, кожен з яких складається з 25 питань. На кожне питання надається 4 варіанти відповіді, серед яких лише 1 – вірний. Максимальна оцінка за тест складає 50 балів. Опитування за тестом проводиться з використанням системи дистанційної освіти Moodle.

6.4. Критерії оцінювання лабораторної роботи

Кожна лабораторна робота виконується на мережному обладнанні Cisco або в програмі Packet Tracer в випадку дистанційного навчання:

- **100 балів** – робота виконана повністю на 100%
- **90 балів** – робота виконана повністю, але містить неточності та/або недостатньо пояснень;
- **N балів** – відповідно N процентів виконано завдання.

Максимальна оцінка за лабораторну роботу складає 100 балів. Максимальна оцінка за лабораторний практикум – 50 балів за формулою: (середнє зважене)/2.

7. Політика курсу

7.1. Політика щодо академічної доброчесності

Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка". http://www.nmu.org.ua/ua/content/activity/us_documents/System_of_prevention_and_detection_of_plagiarism.pdf.

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікативна політика

Здобувачі вищої освіти повинні мати активовану університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання

Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4 Політика щодо оскарження оцінювання

Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

7.5. Відвідування занять

Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в он-лайн формі за погодженням з керівником курсу.

7.6. Студентоцентризований підхід

Для врахування інтересів та потреб студентів на початку вивчення курсу здобувачам вищої освіти пропонується відповісти у системі Moodle на низку питань щодо інформаційного наповнення курсу. Відповідно до результатів опитування формується траєкторія навчання з урахуванням потреб студентів.

Під час навчання студенти реалізують своє право вибору індивідуальних завдань лабораторних робіт.

Наприкінці вивчення курсу та перед початком сесії здобувачам вищої освіти пропонується анонімно заповнити у системі Moodle електронні анкети для оцінки рівня задоволеності методами навчання і викладання та врахування пропозицій стосовно покращення змісту навчальної дисципліни. За результатами опитування вносяться відповідні корективи у робочу програму та силябус

8 Рекомендовані джерела інформації

Базові

1. Курс Network Academy Cisco «Big Data & Analytics». [URL: <https://www.netacad.com/courses/iot/big-data-analytics>]
2. Лесна Н. С., Репка В. Б., Шатовська Т. Б. Інтелектуальний аналіз даних / Лесна Н. С., Репка В. Б., Шатовська Т. Б.. -Харків.: ХНУРЕ, 2003. -110 с.
3. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг) / Ситник В. Ф., Краснюк М. Т.. -К.: КНЕУ, 2007. -375 с.
4. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник. К., 2014. 599 с.

Допоміжні

1. Барсегян, А. А. , Куприянов М. С. Степаненко В. В., Холод И. И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / 2-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2010. 384 с:
2. Дюк В., Самойленко А. Data Mining: учебный курс. СПб: Питер, 2001. -368 с. Теоретические основы компьютерной безопасности. : Учеб. пособие для вузов./ Девянин Н.Н., Михальский О.О. и др. – М.: Радио и связь, 2000. – 192 с.