

Чернівецький національний університет імені Юрія Федьковича
(повне найменування закладу вищої освіти)

Навчально-науковий інститут фізико-технічних та комп'ютерних наук
(назва навчально-наукового інституту / факультету)

Кафедра комп'ютерних систем та мереж



"ЗАТВЕРДЖУЮ"

Директор навчально-наукового інституту
фізико-технічних та комп'ютерних наук
О. В. Ангельський

2022 року

РОБОЧА ПРОГРАМА навчальної дисципліни

Технології IoT & IoE для Big Data Analysis

(англійською мовою)

(назва навчальної дисципліни)

обов'язкова

(вказати: обов'язкова / вибіркова)

Освітньо-наукова програма Комп'ютерна інженерія
технологій Інтернету речей та кіберфізичних систем
(назва програми)

Спеціальність 123 Комп'ютерна інженерія
(вказати: код, назва)

Галузь знань 12 Інформаційні технології
(вказати: шифр, назва)

Рівень вищої освіти другий (магістерський)
(вказати: перший бакалаврський/другий магістерський)

фізико-технічних та комп'ютерних наук

(назва факультету/ навчально-наукового інституту,
на якому здійснюється підготовка фахівця за вказаною освітньо-професійною програмою)

Мова навчання англійська, українська

Робоча програма навчальної дисципліни

OK9 Технології IoT & IoE для Big Data Analysis (англійською мовою)

(назва навчальної дисципліни)

складена відповідно до освітньо-наукової програми

Комп'ютерна інженерія технологій інтернету речей і кіберфізичних систем,

123 Комп'ютерна інженерія

(назва освітньо-професійної програми, код та назва спеціальності)

12 Інформаційні технології, 15 квітня 2021 р.

(галузь знань, шифр та назва; дата останнього затвердження)

Розробники: Танасюк Юлія Володимирівна, доцент кафедри КСМ,

канд.фіз.-мат. наук, доцент

(П.І.Б. авторів, посада, науковий ступінь, вчене звання)

Погоджено з гарантом ОП і затверджено на засіданні

кафедри комп'ютерних систем та мереж

Протокол № 1 від "29" серпня 2022 року

Завідувач кафедри  (Воробець Г.І.)

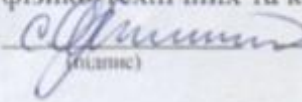
(підпис)

(прізвище та ініціали)

Схвалено методичною радою навчально-наукового інституту
фізико-технічних та комп'ютерних наук

Протокол № 1 від "31" серпня 2022 року

Голова методичної ради навчально-наукового інституту
фізико-технічних та комп'ютерних наук

 Струк Я. М.)

(підпис)

(прізвище та ініціали)

1. Мета навчальної дисципліни

Мета дисципліни «Технології IoT & IoE для Big Data Analysis»: формування у студентів здатностей застосовувати програмні методи та засоби оброблення даних для аналізу великих даних та прийняття управлінських рішень у різних галузях науки та бізнесу.

2. Результати навчання

У результаті вивчення навчальної дисципліни студент отримує компетентності, у результаті чого повинен:

2.1. Знати:

- архітектурні моделі Big Data;
- технології віртуалізації та контейнерної технології виконання програмного коду на сервері;
- технології Hadoop Big Data, MapReduce, HDFS, Kafka, Cassandra, Spark;
- можливості, що надає мова програмування Python для аналізу великих даних;
- технології імпортування даних з Інтернету та засоби для кореляційного аналізу в Pandas;
- засоби керування та оброблення даних, доступні у базах даних SQL і NoSQL;
- технології регресійного аналізу великих даних засобами Python;
- технології візуалізації даних Pyplot, Plotly, Scikit-Learn, Seaborn, Folium, Leaflet.js Pandas та Matplotlib;
- методи та засоби проведення аналізу великих даних з використанням хмарних, туманних і крайових обчислень;
- заходи для організації оброблення та візуалізації великих даних на базі Raspberry Pi.

2.2. Уміти:

- використовувати технології IoT та програмні засоби для обробки та аналізу великих даних.
- засобами мови програмування Python створювати повний цикл видобування, маніпулювання та подання даних із різних джерел;
- застосовувати пошуковий аналіз даних для генерування рішень на основі одержаних даних;
- подавати та візуалізувати великі обсяги даних;
- опрацьовувати слабоструктуровані та різнотипні дані, що надходять з різних джерел, засобами мови Python та баз даних NoSQL.

2.3. Набути компетентностей:

Загальні компетенції (ЗК):

- ЗК1. Здатність до адаптації та дій в новій ситуації.
- ЗК2. Здатність до абстрактного мислення, аналізу і синтезу.
- ЗК4. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ЗК6. Здатність виявляти, ставити та вирішувати проблеми.
- ЗК7. Здатність приймати обґрунтовані рішення.

ЗК8. Здатність спілкуватися іноземною мовою.

Фахові компетентності спеціальності (СК):

СК2. Здатність розробляти алгоритмічне та програмне забезпечення, компоненти комп'ютерних систем та мереж, Інтернет додатків, кіберфізичних систем з використанням сучасних методів і мов програмування, а також засобів і систем автоматизації проектування.

СК4. Здатність будувати та досліджувати моделі комп'ютерних систем та мереж.

СК6. Здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема з метою підвищення їх ефективності.

СК9. Здатність представляти результати власних досліджень та/або розробок у вигляді презентацій, науково-технічних звітів, статей і доповідей на науково-технічних конференціях.

СК10. Здатність ідентифікувати, класифікувати та описувати роботу програмно-технічних засобів, комп'ютерних систем, мереж та їхніх компонентів;

СК11. Здатність обирати ефективні методи розв'язування складних задач комп'ютерної інженерії, критично оцінювати отримані результати та аргументувати прийняті рішення.

СК12. Здатність використовувати методи аналізу, ідентифікації й синтезу комп'ютерних систем та мереж, кіберфізичних систем, засобів Інтернету речей та IT-інфраструктур.

СК13. Здатність застосовувати технології IoT, IoE, мобільні і гібридні IoT обчислення для аналізу великих даних, вирішувати завдання комп'ютерної інженерії та науково-прикладного застосування комп'ютерних засобів з використанням штучного інтелекту, хмарних технологій, IoT, комп'ютеризованих інформаційно-вимірювальних систем і комплексів.

Програмні результати навчання (РН):

РН2. Знаходити необхідні дані, аналізувати та оцінювати їх.

РН6. Аналізувати проблематику, ідентифікувати та формулювати конкретні проблеми, що потребують вирішення, обирати ефективні методи їх вирішення.

РН7. Вирішувати задачі аналізу та синтезу комп'ютерних систем та мереж.

РН9. Розробляти програмне забезпечення для вбудованих і розподілених застосувань, мобільних і гібридних систем.

РН10. Здійснювати пошук інформації в різних джерелах для розв'язання задач комп'ютерної інженерії, аналізувати та оцінювати цю інформацію.

РН12. Вільно спілкуватись усно і письмово українською мовою та однією з іноземних мов (англійською, німецькою, італійською, французькою, іспанською) при обговоренні професійних питань, досліджень та інновацій в галузі інформаційних технологій.

Змістовий модуль 2. <i>Архітектурні моделі Big Data</i>												
Тема 3.	19	3	-	4	-	12						
Тема 4.	27	3	-	6	-	18						
Тема 5.	26	3	-	8	-	15						
Разом за ЗМ 2	72	9		18	-	45						
Усього годин	120	15	-	30	-	75						

3.3. Тематика лабораторних занять

№	Назва теми (завдання)	Кількість годин
1.	Аналіз та візуалізація даних у Python.	4
2.	Кореляційний аналіз у Python.	4
3.	Описова статистика у Python.	4
4.	Побудова лінійної регресії та прогнозування.	4
5.	Візуалізація статистичних даних.	6
6.	NoSQL бази даних для розподіленої обробки великих даних	8
Разом:		30

3.4. Самостійна робота студентів

№	Назва теми	Кількість годин
1	Дослідження джерел відкритих даних. Завантаження датасету та збереження даних в форматі csv.	6
2	Програмування Python та SQLite. Призначення утиліти csvsql.	4
3	Статистичні підходи до аналітики великих даних. Використання Pandas. Імпорт даних з мережі Інтернет. Описова статистика в Pandas. Засоби для кореляційного аналізу в Pandas.	12
4	Регресійний аналіз даних в Python	5
5	Модуль Рурplot. Інструмент Plotly. Типи візуалізації даних.	5
6	Технології Hadoop Big Data. Розподілена обробка MapReduce. HDFS.	4
7	Kafka: розподілена потокова платформа обробки великих даних. Переваги Cassandra.	8
8	Проблема обчислювальної функції. Технологія Spark. Порівняння Spark та MapReduce.	12
9	Lambda та Карра архітектури оброблення великих даних.	8
10	Машинне навчання.	4
11	Захист великих даних.	4
12	Етика використання даних	3
Разом:		75

4. Технічне й програмне забезпечення/обладнання.

Для виконання лабораторних робіт та завдань із самостійної роботи на кафедрі КСМ в комп'ютерних класах 8 к. ЧНУ наявні комп'ютери з такою конфігурацією:

- Motherboard Asus Prime H310M-A R2.0
- CPU Intel Pentium Gold G5400 (BX80684G5400) s1151 BOX
- SSD Apacer AS350 Panther 240GB 2.5" SATAIII TLC (AP240GAS350-1)
- Memory HyperX DDR4-2400 8192MB PC4-19200 Fury Black (HX424C15FB2/8)
- Case GameMax ET-207 400 Вт
- Keyboard Defender Element HB-520 PS/2 Black (45520)
- Mouse 2E MF107 USB Black (2E-MF107UB)
- Monitor 21.5" Philips.

Апаратне та програмне забезпечення для виконання лабораторних робіт:

- 1 бездротовий маршрутизатор (універсальний бренд) з підтримкою WPA2
- Raspberry Pi 3
- Cisco Packet Tracer v8.1 та вище
- Jupyter Notebook.

5. Форми і методи навчання

Форми навчання – це проблемні й оглядові лекції, лабораторні заняття, заняття із застосуванням комп'ютерної та телекомунікаційної техніки, інтерактивні заняття з навчанням одних студентів іншими, відеолекції, відеоконференції засобами Google Meet, Cisco Webex, заняття з використанням системи електронного навчання Moodle та освітнього порталу Програми мережних академій Cisco - netacad.com.

Методи: проблемний виклад матеріалу, частково-пошукові та дослідницькі лабораторні практикуми, презентації, консультації і дискусії, робота в інтернет-класі: електронні лекції, лабораторні роботи, дистанційні консультації та ін., спрямовані на активізацію і стимулювання навчально-пізнавальної діяльності студентів.

Підходи до навчання: використовуються студентоцентрований, проблемно-орієнтований, діяльнісний, комунікативний, професійно-орієнтований, міждисциплінарний підходи.

Реалізація навчального процесу здійснюється під час лекційних, лабораторних занять, самостійної позааудиторної роботи з використанням сучасних інформаційних засобів навчання, програмних симуляторів, віртуальних середовищ колективної роботи, консультацій з викладачами.

Для формувань умінь та навичок застосовуються такі методи навчання:

- вербальні/словесні (лекція, пояснення, розповідь, бесіда, інструктаж);
- наочні (спостереження, ілюстрація, демонстрація);
- практичні (проведення експерименту, практики);
- пояснювально-ілюстративний, який передбачає пред'явлення готової інформації викладачем та її засвоєння студентами;
- репродуктивний (виконання лабораторних завдань за зразком);
- метод проблемного викладу матеріалу на лабораторних заняттях.

6. Критерії оцінювання результатів навчання з навчальної дисципліни

Шкала оцінювання: національна та ЄКТС

Оцінка за шкалою ЄКТС	Критерії	Пояснення	Оцінка за 100-бальною шкалою	Оцінка за національною шкалою (іспит/залік)
A	Відмінний рівень компетентностей у межах обов'язкового матеріалу, з можливими незначними недоліками	відмінно/ зараховано	90 – 100	відмінно/ зараховано
B	Достатньо високий рівень компетентностей у межах обов'язкового матеріалу без суттєвих (грубих) помилок	дуже добре/ зараховано	80-89	добре/ зараховано
C	В цілому добрий рівень компетентностей із незначною кількістю помилок	добре/ зараховано	70-79	
D	Посередній рівень компетентностей із значною кількістю недоліків, достатній для подальшого навчання або професійної діяльності	задовільно/ зараховано	60-69	задовільно/ зараховано
E	Мінімально можливий допустимий рівень компетентностей	достатньо/ зараховано	50-59	
FX	Незадовільний рівень компетентностей, з можливістю повторного перескладання за умови належного самостійного доопрацювання	(незадовільно/ не зараховано) з можливістю повторного складання	35-49	незадовільно/ не зараховано
F	Дуже поганий рівень компетентностей, що вимагає повторного вивчення дисципліни	(незадовільно/ не зараховано) з обов'язковим повторним вивченням дисципліни	1-34	

7. Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання є

- віртуальні середовища моделювання;
- індивідуальні та групові завдання на лабораторному обладнанні;
- контрольні роботи;
- тести;
- презентації та обґрунтування результатів виконаних завдань.

8. Форми поточного та підсумкового контролю

Формами поточного контролю рівня знань є усна та письмова відповідь студента при захисті виконаних лабораторних робіт, кількість отриманих балів при виконанні тестових завдань та контрольних робіт, а також кількість балів, набутих при виконанні практичних завдань у середовищі моделювання.

8.1. Розподіл максимально можливої кількості балів, які отримують студенти за виконання всіх видів навчальної діяльності

Поточне оцінювання (індивідуальна робота та самостійна робота)										Підсумковий контроль (іспит)	Сумарна кількість балів
Змістовий модуль 1				Змістовий модуль 2							
T1	TM1	T2	TM2	T3	TM3	T4	TM4	T5	TM5	40	100
5	10	5	10	5	5	5	5	5	5		

Змістовий модуль 1. Джерела та типи великих даних

T1. Аналіз і візуалізація даних у Python. (виконання та захист л/р №1 та на основі лекційного матеріалу, методичних розробок та матеріалів самостійної роботи – 5 балів).

TM1 – підсумковий тест за матеріалами Теми 1 (10 балів).

T2. Кореляційний аналіз у Python (виконання та захист л/р № 2 на основі лекційного матеріалу та матеріалів самостійної роботи – 5 балів).

TM2 – підсумковий тест за матеріалами Теми 2 (10 балів).

Змістовий модуль 2. Архітектурні моделі Big Data

T3. Описова статистика у Python. (виконання та захист л/р № 3 на основі лекційного матеріалу та матеріалів самостійної роботи – 5 балів).

TM3 – підсумковий тест за матеріалами Теми 3 (10 балів).

T4. Побудова лінійної регресії та прогнозування. (виконання та захист л/р № 4 на основі лекційного матеріалу та матеріалів самостійної роботи – 5 балів).

TM4 – підсумковий тест за матеріалами Теми 4 (5 балів).

T5. Класифікація за деревом рішення. Візуалізація статистичних даних. (виконання та захист л/р № 5 і 6 – 5 балів, на основі лекційного матеріалу, методичних розробок та матеріалів самостійної роботи).

TM5 – підсумковий тест за матеріалами Теми 5 (5 балів).

Підсумковий контроль (іспит) – 40 балів.

Формами підсумкового контролю рівня знань є складання тестів та виконання практичних завдань у віртуальному середовищі моделювання.

8.2. Перелік питань для підсумкового модуль-контролю

1. Інтернет Речей (IoT) та збільшення обсягів даних. Визначення великих даних. Приклади великих даних у реальному світі.
2. Відкриті дані. Приватність даних. Структуровані та неструктуровані дані. Хмарні та туманні обчислення. Дані в спокої та дані в русі. Інфраструктура великих даних. Розподілені дані та їх обробка.
3. Можливості інструментів аналізу даних. Роль Python в аналізі даних. Традиційна аналітика великих даних та аналітика нового покоління.
4. Життєвий цикл аналізу даних. Відкриті дані, їх формати та засоби обробки. Веб-скрепінг. Видобування, перетворення та завантаження даних.
5. Форматування даних про час та дату у Python.
6. Читання та запис файлів в Python. Взаємодія із зовнішніми додатками.
7. Основні операції SQL. Робота Python з SQLite. Призначення утиліти csvsql.
8. Статистичні підходи до аналітики великих даних. Використання Pandas.
9. Імпорт даних з файлів. Імпорт даних з мережі Інтернет.
10. Описова статистика в Pandas. Засоби для кореляційного аналізу в Pandas.
11. Оброблення відсутніх даних. Перетворення типів даних. Маніпулювання дата фреймами.
12. Регресійний аналіз. Типи регресійного аналізу. Застосування регресійного аналізу.
13. Помилки в аналізі даних та прогнозній аналітиці. Оцінка помилок регресії засобами Python. Призначення бібліотеки scikit-learn.
14. Проблеми класифікації. Алгоритми класифікації. Візуалізація класифікацій. Застосування та валідація класифікацій.
15. Модуль Rpyplot. Інструмент Plotly. Типи візуалізації даних. Візуалізація аномалій. Використання бібліотек Folium та Leaflet.js для побудови карт.
16. Історія розвитку мови R. Можливості мови R.
17. Об'єкти, пакети, функції. Вектори, матриці та операції над ними в R. Фактори, списки, фрейми та дії над ними.
18. Типи змінних величин. Вибірка та популяція. Описова статистика. Розподіл, дисперсія, мода, середнє арифметичне, медіана.
19. Кореляція та обумовленість а розрізі оброблення великих даних.
20. Архітектурні моделі інженерії Big Data.
21. Технології віртуалізації. Шари абстракції.
22. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. Інжиніринг даних.
23. Масштабованість великих даних.
24. Зберігання та оброблення даних в розподілених файлових системах. Розподілені бази даних.
25. Розподілена файлова система Hadoop (HDFS).
26. Проблема прийому даних. Розподілена потокова платформа Kafka. Переваги Cassandra.
27. Проблема обчислювальної функції. Технологія Spark.
28. Порівняння Spark та MapReduce. Spark і sparklyr для роботи з великими даними в R.
29. Lambda - архітектура. Переваги і недоліки Lambda-архітектури.
30. Карра - архітектура. Переваги і недоліки Карра-архітектури.

У разі отримання понад 70 балів за модуль студентіві видається сертифікат про успішне завершення курсу від Cisco Networking Academy.

9. Рекомендована література Основні (фахова)

1. Інтернет речей для індустріальних і гуманітарних застосунків. У трьох томах. Том 1. Основи і технології / За ред. В. С. Харченка. – Міністерство освіти і науки України, Національний аерокосмічний університет ХАІ, 2019. – 605 с.
2. IoT Fundamentals: Big Data & Analytics. Електронний ресурс: <https://www.netacad.com/courses/iot/big-data-analytics>
3. Ерл Т. Основи Big Data: концепції, алгоритми і технології. – К: Баланс бізнес букс, 2018. – 320 с.
4. О'Нил К. Big Data. Зброя математичного знищення. – К: Форс, 2019. – 336 с.
5. Провост Ф., Фоусет Т. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані. – К.: Наш Формат, 2019. 400 с.
6. Nokeri T. C. Data Science Solutions with Python. – Apress, 2022. 119 p.

Допоміжна

1. Klosterman S. Data science projects with Python: a case study approach to gaining valuable insights from real data with machine learning? 2nd edition. Packt Publishing Limited. 2021. 432 p.
2. Kats P. Learn Python by building data science applications. 1st edition. Packt Publishing. 2019. 482 p.
3. Bruce P. Practical statistics for data scientists: 50 essential concepts. 1st edition. O'Reilly Media. 2017. 318 p.
4. Muller A.C. Introduction to machine learning with Python: a guide for data scientists. 1st edition. O'Reilly Media. 2016. 402 p.
5. О'Нил К., Шатт Р. Data Science. Інсайдерська інформація для новачків. Включаючи мову R. Пітер. 2016. 368 с.

Інтернет-ресурси

1. Big Data: 20 free Big Data sources everyone should know. Електронний ресурс: <https://www.smartdatacollective.com/big-data-20-free-big-data-sources-everyone-should-know/>
2. Big Data & Analytics Tutorial. Електронний ресурс: https://www.tutorialspoint.com/big_data_tutorials.htm
3. Big Data Tutorial Library. Електронний ресурс: <https://data-flair.training/blogs/big-data-tutorials-home/>
4. Machine learning. Електронний ресурс: https://www.w3schools.com/python/python_ml_getting_started.asp
5. Python Machine Learning Tutorials. Електронний ресурс: <https://realpython.com/tutorials/machine-learning/>
6. Data Science Tutorial for Beginners. Електронний ресурс: <https://www.simplilearn.com/tutorials/data-science-tutorial>