

Міністерство освіти і науки України  
Львівський національний університет імені Івана Франка  
Кафедра системного проектування

“ЗАТВЕРДЖУЮ”

Декан факультету  
електроніки та комп'ютерних технологій  
\_\_\_\_\_ доц.Юрій Фургала  
“ \_\_\_\_\_ ” \_\_\_\_\_ 2022 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

\_\_\_\_\_ «Методи та технології інженерії даних»  
(шифр і назва навчальної дисципліни)  
спеціальність \_\_\_\_\_ 121 – Інженерія програмного забезпечення  
(шифр і назва спеціальності)  
спеціалізація \_\_\_\_\_  
(назва спеціалізації)  
факультет \_\_\_\_\_ електроніки та комп'ютерних технологій  
(назва інституту, факультету, відділення)

2022-2023 навчальний рік

Робоча програма “Методи та технології інженерії даних” для студентів  
(назва навчальної дисципліни)  
галузі знань “12 – Інформаційні технології”  
за спеціальністю “121 Інженерія програмного забезпечення”

Розробник: Лідія ДЕМКІВ (кандидат фізико-математичних наук, доцент,  
доцент кафедри системного проектування)  
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Робочу програму схвалено на засіданні кафедри (циклової комісії) \_\_\_\_\_  
\_\_\_\_\_ системного проектування

Протокол від “ 30 ” 08 2022 року № 1

Завідувач кафедри \_\_\_\_\_ системного проектування



( Роман ШУВАР )

(підпис)

(прізвище та ініціали)

Ухвалено Вченою радою \_\_\_\_\_ факультету електроніки та комп'ютерних технологій

Протокол від “ 31 ” 08 2022 року № 28/22

### 1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни
		<i>денна форма навчання</i>
Кількість кредитів – <b>4</b>	Галузь знань <b>12</b> <b>Інформаційні технології</b>	Вибіркова
Модулів –	Спеціальність <b>121</b> – Інженерія програмного забезпечення	<i>Рік підготовки:</i> <b>3-й</b>
Змістових модулів – 2		<i>Семестр</i> <b>5-й</b>
Індивідуальне науково-дослідне завдання –		
Загальна кількість годин – <b>120</b>		
Тижневих годин для денної форми навчання: аудиторних – <b>4</b> самостійної роботи студента – <b>3,5</b>	Освітньо-кваліфікаційний рівень: <b>бакалавр</b>	<i>Лекції</i> <b>32 год.</b>
		<i>Практичні, семінарські</i> -----
		<i>Лабораторні</i> <b>32 год.</b>
		<i>Самостійна робота</i> <b>56 год.</b>
		<i>Вид контролю: іспит</i>

## 2. Мета та завдання навчальної дисципліни

Дисципліна “**Методи та технології інженерії даних**” є вибірковою дисципліною для спеціалізації 121 Інженерія програмного забезпечення

**Мета:** реалізувати знайомство студента із способами створення цілісних та чистих даних, теоретичними принципами галузі інженерії даних та практичними технологіями, які найбільш затребувані в інженерії даних.

**Цілі:** забезпечити знайомство студентів з особливостями обробки та зберігання різних типів даних, варіантами побудови сховищ даних; вивчити алгоритми попередньої обробки даних; опанувати теоретичний матеріал і практичне оволодіння сучасними графічно-інформаційними технологіями, комп'ютерними та програмними засобами створення цілісних даних, подання їх в графічній формі; вивчити методи і алгоритми обробки даних, визначити статистичні параметри даних; ознайомити з базовими концепціями обробки даних, які дозволять правильно структурувати дані для подальшого їхнього опрацювання, візуалізації і моделювання, управління програмною інфраструктурою та інтерфейсом систем обробки даних, теорією і проектуванням систем обробки даних.

В результаті вивчення даної дисципліни студент повинен вміти

**Знати:** принципи формування та збереження даних; принципи побудови сховищ даних Data Warehouse та DataLake, способи візуалізації даних; методи побудови, перетворення і аналізу графічної інформації; підходи до зберігання, представлення та обробки інформації в сучасних інформаційних системах;

**Вміти:** використовувати сучасні програмні засоби для попередньої обробки неідеальних реальних даних, запису даних у відповідні структури та сховища даних, моделювання та інтеграції даних; реалізовувати інтерактивні візуалізації даних; проводити необхідну попередню обробку даних для отримання чистих даних; визначати тип задачі аналізу та вирішувати її адекватно обраним методом з оптимально визначеними параметрами; оцінювати результати; робити змістовні висновки та інтерпретацію опрацювання даних.

Після вивчення даного курсу здобувачі набудуть таких Загальних та Фахових компетентностей та Програмних результатів навчання:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК16. Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.

ФК17. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

ФК18. Здатність розробляти архітектури, модулі та компоненти програмних систем.

ФК19. Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.

ФК20. Здатність дотримуватися специфікацій, стандартів, правил і рекомендацій в професійній галузі при реалізації процесів життєвого циклу.

ФК22. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

ПРН21. Знати, аналізувати, вибирати, кваліфіковано застосовувати засоби забезпечення інформаційної безпеки (в тому числі кібербезпеки) і цілісності даних відповідно до розв'язуваних прикладних завдань та створюваних програмних систем.

### **3. Програма навчальної дисципліни**

#### **Змістовний модуль 1 Отримання даних, розуміння даних, створення чистих даних**

*Тема 1.* Поняття даних. Визначення кількості даних та проблеми швидкого збільшення кількості даних в сучасному світі. Професії, що працюють з даними. Відкриті дані — це цінний ресурс, який допомагає посилити цифрову та "реальну" економіку країни. Засоби вилучення, перетворення і завантаження даних. Поняття ETL, pipeline, datalake, data warehouse.

*Тема 2.* Формати зберігання даних. Джерела даних: таблиці, файли, бази даних, web-сервіси. Зчитування та запис файлових даних. Типи даних за шкалами вимірювання.

*Тема 3.* Візуалізація даних як етап аналізу даних

*Тема 4.* Коваріація та кореляція даних. Кореляційний аналіз кількісних ознак. Побудова рекомендаційних систем.

*Тема 5.* Статистичний аналіз даних. Обчислення описової статистики даних. Статистичні моменти даних. Роботи з великою кількістю даних (оновлення статистики даних) Нормування та стандартизація даних

*Тема 6.* Розподіли числових даних. Нормальний розподіл даних. Види розподілів. Перевірка гіпотез. Види гіпотез. P-value. Параметричні та непараметричні тести.

*Тема 7.* Дослідницький аналіз даних. Етапи дослідницького аналізу даних.

*Тема 8.* Створення даних для бізнес аналітики (BI). Business intelligence – інструменти, технології і процеси, які використовуються для перетворення даних в інформацію. Збільшення цінності даних. Проблеми імплементації BI. Засоби BI – Tableau для створення бізнес аналітики.

#### **Змістовний модуль 2 Робота з даними у різних форматах. Формування сховищ даних**

*Тема 9.* Підготовка даних до класифікації. Редукція та класифікація даних. Метод головних компонент.

*Тема 10.* Побудова дерев класифікації. Random Forest. Порівняння методів класифікації.

*Тема 11.* Доступ до баз даних за допомогою модулів python. API Python для роботи з базами даних MySQL, SQLite, Принципи побудови DataLake.

*Тема 12.* Поняття ETL. Принципи побудови сховищ даних DataWareHouse.

*Тема 13.* Робота з нереляційними базами даних. JSON-формат та особливості його використання. Створення та парсинг формату. Типи даних json та масиви. Особливості зчитування json у python. Підготовка даних до кластеризації.

*Тема 14.* Графові бази даних. Графи та структури даних на графах. Представлення та візуалізація даних за допомогою графів

*Тема 15.* Особливості роботи з сирими raw даними. Чисельні системи представлення даних та частота дискретизації даних

*Тема 16.* Pipeline - трубопровід для опрацювання даних. Схеми побудови трубопроводів. Зберігання та оброблення даних в розподілених файлових системах. Особливості роботи з BigData.

#### 4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	Денна форма					
	Усього	у тому числі				
		Л	П	лаб	інд	Ср
1	2	3	4	5	6	7
<b>Змістовий модуль 1. Створення та статистичний аналіз даних</b>						
<i>Тема 1.</i> Поняття даних. Визначення кількості даних та проблеми швидкого збільшення кількості даних в сучасному світі. Професії, що працюють з даними. Відкриті дані — це цінний ресурс, який допомагає посилити цифрову та "реальну" економіку країни. Засоби вилучення, перетворення і завантаження даних. Поняття ETL, pipeline, datalake, data warehouse.	7	2		2		3
<i>Тема 2.</i> Формати зберігання даних. Джерела даних: таблиці, файли, бази даних, web-сервіси. Зчитування та запис файлових даних. Типи даних за шкалами вимірювання.	8	2		2		4
<i>Тема 3.</i> Візуалізація даних як етап аналізу даних	7	2		2		3
<i>Тема 4.</i> Коваріація та кореляція даних. Кореляційний аналіз кількісних ознак. Побудова рекомендаційних систем.	8	2		2		4
<i>Тема 5.</i> Статистичний аналіз даних. Обчислення описової	7	2		2		3

статистики даних. Статистичні моменти даних. Роботи з великою кількістю даних (оновлення статистики даних) Нормування та стандартизація даних						
<i>Тема 6.</i> Розподіли числових даних. Нормальний розподіл даних. Види розподілів. Перевірка гіпотез. Види гіпотез. P-value. Параметричні та непараметричні тести.	8	2		2		4
<i>Тема 7.</i> Дослідницький аналіз даних. Етапи дослідницького аналізу даних.	7	2		2		3
<i>Тема 8.</i> Створення даних для бізнес аналітики (BI). Business intelligence – інструменти, технології і процеси, які використовуються для перетворення даних в інформацію. Збільшення цінності даних. Проблеми імплементації BI. Засоби BI – Tableau для створення бізнес аналітики.	8	2		2		4
<i>Разом – МОДУЛЬ 1</i>	<b>60</b>	16		16		28
<b>Змістовний модуль 2 Зменшення кількості даних та робота з базами даних</b>						
Тема 9. Підготовка даних до класифікації. Редукція та класифікація даних. Метод головних компонент.	7	2		2		3
Тема 10. Побудова дерев класифікації. Random Forest. Порівняння методів класифікації.	8	2		2		4
Тема 11. Доступ до баз даних за допомогою	7	2		2		3

модулів python. API Python для роботи з базами даних MySQL, SQLite, Принципи побудови DataLake						
Тема 12. Поняття ETL. Принципи побудови сховищ даних DataWareHouse.	<b>8</b>	2		2		4
Тема 13. Робота з нереляційними базами даних. JSON-формат та особливості його використання. Створення та парсинг формату. Типи даних json та масиви. Особливості зчитування json у python. Підготовка даних до кластеризації.	<b>7</b>	2		2		3
Тема 14. Графові бази даних. Графи та структури даних на графах. Представлення та візуалізація даних за допомогою графів	<b>8</b>	2		2		4
Тема 15. Особливості роботи з сирими raw даними. Чисельні системи представлення даних та частота дискретизації даних	<b>7</b>	2		2		3
Тема 16. Pipeline - трубопровід для опрацювання даних. Схеми побудови трубопроводів. Зберігання та оброблення даних в розподілених файлових системах. Особливості роботи з BigData.	<b>8</b>	2		2		4
<i>Разом – МОДУЛЬ 2</i>	<b>60</b>	16		16		28
<b>Усього годин</b>	<b>120</b>	<b>32</b>		<b>32</b>		<b>56</b>

### ***5. Темі семінарських занять***

Семінарські заняття в курсі не передбачені.

### ***6. Темі практичних занять***

Практичні заняття в курсі не передбачені.



### **7. Теми лабораторних занять**

№ з/п	Назва теми	Кількість годин
1	Вступне заняття. Інструкція з техніки безпеки. Методи бібліотеки pandas для роботи з даними	2
2	Вивчення методів бібліотеки pandas (python) для роботи з даними: сортування, групування, фільтрація, об'єднання даних. Основи ВІ аналітики.	2
3	Підготовка даних до візуалізації. Візуалізація даних.	
4	Коваріація та кореляція даних.	2
5	Статистичний аналіз даних	2
6	Захисне заняття	
7	Дослідницький аналіз даних.	2
8	Побудова дашбордів за допомогою Tableau	2
9	Підготовка даних до класифікації та ML	2
10	Редукція та класифікація даних	2
11	Доступ до баз даних за допомогою модулів python	2
12	Моделювання побудови сховищ даних Data WareHouse	2
13	JSON-формат та особливості його використання..	
14	Графи та структури даних на графах.	
15	Особливості роботи з сирими raw даними.	2
16	Підсумкове заняття	2
	<b>Разом</b>	<b>32</b>

### **8. Самостійна робота**

№ з/п	Назва теми	Кількість Годин
1	Опрацювання лекційного матеріалу	16
2	Підготовка до лабораторних робіт, оформлення звітів	16
3	Бібліотеки python для аналізу та візуалізації даних. Бібліотека pandas для інженерії даних. Робота з категорійними даними та даними часу і дати	2
4	Ознайомлення з базовими концепціями обробки даних, які дозволять правильно структурувати дані для подальшого їхнього опрацювання	2
5	ВІ аналіз	2
6	Редукція та класифікація даних	4
7	API Python для роботи з базами даних MySQL, SQLite,	4
8	Підходи до зберігання, представлення та обробки інформації в сучасних інформаційних системах	4
9	Методи для роботи з потоками даних	3
10	Побудова pipeline для аналізу даних	3
	<b>Разом</b>	<b>56</b>

### **9.Індивідуальні завдання**

#### **10. Методи навчання**

Навчальні заняття проводяться у формі лекційних та лабораторних робіт. Лекція – основна форма проведення навчальних занять, призначена для засвоєння теоретичного матеріалу. Лабораторні заняття проводяться у навчальному комп'ютерному класі.

Самостійна робота студента є основним засобом оволодіння навчальним матеріалом у час, вільний від обов'язкових навчальних занять. Самостійна робота студентів складається з таких видів робіт:

- підготовка до лекцій та лабораторних робіт,
- оформлення звітів про виконані лабораторні роботи,
- самостійне опрацювання окремих тем навчальної дисципліни,
- підготовка до модульних контрольних робіт.

### ***11. Методи контролю***

Оцінка якості засвоєння навчальної дисципліни включає поточний контроль успішності та складання підсумкового заліку.

Для поточного контролю засвоєння студентами навчального матеріалу передбачається виконання 14 обов'язкових лабораторних робіт, усного опитування по темі та захист лабораторних робіт.

Іспит проводиться в письмово-усній формі.

### ***12. Розподіл балів, які отримують студенти (для іспиту)***

Контроль засвоєння матеріалу включає поточний контроль, лабораторні роботи та оцінка відповідей (50 балів) та іспит (50 балів). Сумарна оцінка виставляється за 100-бальною шкалою.

При оцінюванні лабораторної роботи враховується підготовка до виконання лабораторної роботи, хід виконання лабораторної роботи, оформлення звіту, отримані результати та захист звіту про виконану лабораторну роботу.

Контроль засвоєння матеріалу включає поточний контроль (контрольну роботу на 20 балів), лабораторні роботи та оцінка відповідей (50 балів) та підсумковий контроль (30 балів) - разом за семестр 100 балів. Залік виставляється за 100-бальною шкалою, на основі балів, набраних протягом семестру.

<b>Поточне тестування та самостійна робота</b>																<b>Іспит</b>	<b>Сума</b>
<b>Змістовий модуль 1</b>								<b>Змістовий модуль 2</b>								<b>50</b>	<b>100</b>
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16		
3	3	3	3	3	3	4	4	3	3	3	3	3	3	4	4		

#### **Шкала оцінювання: Університету, національна та ECTS**

Сума балів за всі види навчальної діяльності	Оцінка за ECTS	Оцінка за національною шкалою		
			іспит	залік
90-100	<b>A</b>	<b>5</b>	відмінно	
81-89	<b>B</b>		добре	
71-80	<b>C</b>	<b>4</b>		
61-70	<b>D</b>			
51-60	<b>E</b>		<b>3</b>	задовільно

21–50	<b>FX</b>	<b>2</b>	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-20	<b>F</b>	<b>2</b>	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

## **14. РЕКОМЕНДОВАНА ЛІТЕРАТУРА**

### **Основна**

1. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с
2. Литвин В. В. Методи та засоби інженерії даних та знань / В. В. Литвин // навчальний посібник з грифом МОНУ. — Львів : «Магнолія-2006», 2012. — 241 с.
3. Han, Jiawei. Data mining : concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei. – 3rd ed. ISBN 978-0-12-381479-1 Chapter 3. Data preprocessing
4. Michael R. Brzustowicz Data Science with Java Practical Method for scientists and engineers /Michael R. Brzustowicz. – O'REILLY, 2017. – 311p.
5. Бахрушин В.С. Методи аналізу даних : навчальний посібник для студентів /В.С. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
6. Методичні вказівки до виконання лабораторних робіт з дисципліни “Інженерія прикладних інтелектуально-орієнтованих програмних продуктів” для студентів спеціальностей 121 “Інженерія програмного забезпечення” та 122 “Комп’ютерні науки та інформаційні технології” (всіх форм навчання) / В.М. Льовкін. – Запоріжжя : ЗНТУ, 2016. – 80 с.

### **Допоміжна**

7. Extract, transform, and load (ETL) // Електронний ресурс. Режим доступу: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>
8. Edward L.Robinson Data Analysis for Scientists and Engineers // Pricenton University Press, 2016, - P.408, ISBN 9781400883066

## **15. Інформаційні ресурси**

1. Kaggle - це платформа для змагань з Data Science.[Електронне джерело] <https://www.kaggle.com/>

2. Український цент суспільних даних .[Електронне джерело] <https://socialdata.org.ua/>
3. Відеоматеріали про роботу інженерів даних .[Електронне джерело] [https://www.youtube.com/watch?v=QB6qw4pSJPQ&ab\\_channel=FacultyofAppliedSciencesUCU](https://www.youtube.com/watch?v=QB6qw4pSJPQ&ab_channel=FacultyofAppliedSciencesUCU)
4. Професії дата інженерів у фірмі EPAM .[Електронне джерело] <https://training.epam.ua/News/Items/305?lang=ua>
5. DataScientist vs DataEngineer .[Електронне джерело] <https://uk.education-wiki.com/3868210-data-science-vs-data-engineering--which-one-is-more-useful>