

# СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

## 1. Загальна інформація про навчальну дисципліну

<b>Повна назва навчальної дисципліни</b>	Обробка та аналіз великих даних
<b>Повна офіційна назва закладу вищої освіти</b>	Сумський державний університет
<b>Повна назва структурного підрозділу</b>	Факультет електроніки та інформаційних технологій. Кафедра інформаційних технологій
<b>Розробник(и)</b>	Кузнецов Едуард Геннадійович
<b>Рівень вищої освіти</b>	Другий рівень вищої освіти, НРК – 7 рівень, QF-LLL – 7 рівень, FQ-EHEA – другий цикл
<b>Семестр вивчення навчальної дисципліни</b>	16 тижнів протягом 2-го семестру
<b>Обсяг навчальної дисципліни</b>	Обсяг становить 5 кред. ЄКТС, 150 год. Для денної форми навчання 32 год. становить контактна робота з викладачем (16 год. лекцій, 16 год. лабораторних занять), 118 год. становить самостійна робота.
<b>Мова викладання</b>	Українська

## 2. Місце навчальної дисципліни в освітній програмі

<b>Статус дисципліни</b>	Вибіркова навчальна дисципліна для освітніх програм "Інформаційні технології проектування", "Інформаційні технології проектування"
<b>Передумови для вивчення дисципліни</b>	Знання алгоритмізації та програмування
<b>Додаткові умови</b>	Додаткові умови відсутні
<b>Обмеження</b>	Обмеження відсутні

## 3. Мета навчальної дисципліни

Метою дисципліни є вивчення методів машинної (комп'ютерної) підготовки та обробки розрізнених і неупорядкованих даних задля пошуку та виявлення латентних закономірностей їх вмісту, походження, цінності.

## 4. Зміст навчальної дисципліни

Тема 1 Вступ. Типи даних. Структуровані дані, неструктуровані дані. Дані природньою мовою, машинні дані. Графові (мережні) дані, аудио, відео й графіка, потокові дані  
Вступ. Типи даних. Структуровані дані, неструктуровані дані. Дані природньою мовою, машинні дані. Графові (мережні) дані, аудио, відео й графіка, потокові дані

<p>Тема 2 Область визначення великих даних і data science. Розподілені файлові системи. Інфраструктура розподіленого програмування. Інфраструктура інтеграції даних. Інфраструктури машинного навчання бази даних NoSQL. Інструменти планування</p> <p>Область визначення великих даних і data science. Розподілені файлові системи. Інфраструктура розподіленого програмування. Інфраструктура інтеграції даних. Інфраструктури машинного навчання бази даних NoSQL. Інструменти планування. Інструменти порівняльного аналізу. Розгортання системи. Програмування служб. Безпека. Приклад використання Hadoop</p>
<p>Тема 3 Визначення цілей дослідження й створення проектного завдання. Збір даних. Очищення, інтеграція й перетворення даних. Дослідницький аналіз даних. Побудова моделей. Представлення результатів і побудова додатків на їхній основі</p> <p>Визначення цілей дослідження й створення проектного завдання. Збір даних. Очищення, інтеграція й перетворення даних. Дослідницький аналіз даних. Побудова моделей. Представлення результатів і побудова додатків на їхній основі</p>
<p>Тема 4 Застосування машинного навчання в процесі data science. Інструменти Python, використовувані в машинному навчанні. Створення нових показників і вибір моделі, тренування моделі, перевірка адекватності моделі, прогнозування нових спостережень</p> <p>Застосування машинного навчання в процесі data science. Інструменти Python, використовувані в машинному навчанні. Створення нових показників і вибір моделі, тренування моделі, перевірка адекватності моделі, прогнозування нових спостережень. Типи машинного навчання</p>
<p>Тема 5 Проблеми при роботі з великими об'ємами даних і загальні методи обробки великих об'ємів даних. Правильний вибір алгоритму, структури даних, інструментів</p> <p>Проблеми при роботі з великими об'ємами даних і загальні методи обробки великих об'ємів даних. Правильний вибір алгоритму, структури даних, інструментів</p>
<p>Тема 6 Прогнозування шкідливих адрес URL. Побудова рекомендаційної системи</p> <p>Прогнозування шкідливих адрес URL. Побудова рекомендаційної системи</p>

## 5. Очікувані результати навчання навчальної дисципліни

Після успішного вивчення навчальної дисципліни здобувач вищої освіти зможе:

РН1	Здатність використовувати усталені наукові концепції та сучасні принципи створення і застосування технологій штучного інтелекту
РН2	Уміння розробляти, досліджувати та використовувати математичні методи і алгоритми обробки, аналізу та візуалізації даних
РН3	Знання принципів та інструментальних засобів підготовки та обробки розрізаних і неупорядкованих даних. Здатність до розробки багатовимірних моделей даних, концепцій організації та методик побудови сховищ даних
РН4	Демонструвати та використовувати знання сучасних математичних методів, алгоритмів обробки даних, методів оптимізації на основі технології штучного інтелекту

PH5	Формулювати та вирішувати дослідницьке завдання, для його вирішення збирати, оброблювати та систематизувати інформацію та формулювати висновки
PH6	Знати види даних, методи їх збору, очищення, інтеграції та перетворення. Вміти обирати алгоритми, структури і моделі даних, інструменти їх використання та обробки з метою автоматизації пошуку закономірностей, притаманних групі даних

## 7. Роль освітнього компонента у формуванні соціальних навичок

Загальні компетентності та соціальні навички, формування яких забезпечує навчальна дисципліна:

CH1	Здатність вчитися і оволодівати сучасними знаннями.
CH2	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
CH3	Здатність планувати та управляти часом.
CH4	Здатність застосовувати знання у практичних ситуаціях.

## 8. Види навчальних занять

<b>Тема 1. Вступ. Типи даних. Структуровані дані, неструктуровані дані. Дані природньою мовою, машинні дані. Графові (мережні) дані, аудіо, відео й графіка, потокові дані</b>	
Лк1 "Вступ. Типи даних. Структуровані дані, неструктуровані дані"	Дані природньою мовою, машинні дані. Графові (мережні) дані, аудіо, відео й графіка, потокові дані
Лб1 "Типи даних"	Визначення типів. Приведення до вигляду, зручному для машинної обробки
<b>Тема 2. Область визначення великих даних і data science. Розподілені файлові системи. Інфраструктура розподіленого програмування. Інфраструктура інтеграції даних. Інфраструктури машинного навчання бази даних NoSQL. Інструменти планування</b>	
Лк2 "Область визначення великих даних і data science. Розподілені файлові системи"	Інфраструктура розподіленого програмування. Інфраструктура інтеграції даних
Лк3 "Інфраструктури машинного навчання бази даних NoSQL. Інструменти планування"	Інструменти порівняльного аналізу. Розгортання системи
Лк4 "Програмування служб. Безпека"	Приклад використання Hadoop
Лб2 "Область визначення великих даних і data science. Розподілені файлові системи"	Інфраструктура розподіленого програмування. Інфраструктура інтеграції даних

Лб3 "Інфраструктури машинного навчання бази даних NoSQL. Інструменти планування" Інструменти порівняльного аналізу. Розгортання системи
Лб4 "Програмування служб. Безпека" Приклад використання Hadoop
<b>Тема 3. Визначення цілей дослідження й створення проектного завдання. Збір даних. Очищення, інтеграція й перетворення даних. Дослідницький аналіз даних. Побудова моделей. Представлення результатів і побудова додатків на їхній основі</b>
Лк5 "Визначення цілей дослідження й створення проектного завдання. Збір даних" Очищення, інтеграція й перетворення даних. Дослідницький аналіз даних. Побудова моделей. Представлення результатів і побудова додатків на їхній основі
Лб5 "Процес DataScience" Процес DataScience
<b>Тема 4. Застосування машинного навчання в процесі data science. Інструменти Python, використовувані в машинному навчанні. Створення нових показників і вибір моделі, тренування моделі, перевірка адекватності моделі, прогнозування нових спостережень</b>
Лк6 "Застосування машинного навчання в процесі data science" Інструменти Python, використовувані в машинному навчанні
Лб6 "Застосування машинного навчання в процесі data science" Інструменти Python, використовувані в машинному навчанні
<b>Тема 5. Проблеми при роботі з великими об'ємами даних і загальні методи обробки великих об'ємів даних. Правильний вибір алгоритму, структури даних, інструментів</b>
Лк7 "Проблеми при роботі з великими об'ємами даних і загальні методи обробки великих об'ємів даних" Правильний вибір алгоритму, структури даних, інструментів
Лб7 "Проблеми при роботі з великими об'ємами даних і загальні методи обробки великих об'ємів даних" Правильний вибір алгоритму, структури даних, інструментів
<b>Тема 6. Прогнозування шкідливих адрес URL. Побудова рекомендаційної системи</b>
Лк8 "Прогнозування шкідливих адрес URL" Побудова рекомендаційної системи для бази даних
Лб8 "Прогнозування шкідливих адрес URL" Побудова рекомендаційної системи для бази даних

## 9. Стратегія викладання та навчання

### 9.1 Методи викладання та навчання

Дисципліна передбачає навчання через:

МН1	Лекційне навчання
МН2	Практикоорієнтоване навчання
МН3	Самостійне навчання

Лекції надають студентам теоретичні матеріали з тем дисципліни, що є основою для самостійного навчання здобувачів вищої освіти. Лекції доповнюються лабораторними заняттями, що надають студентам можливість застосовувати теоретичні знання на практичних прикладах. Самостійному навчанню сприятиме підготовка до лекцій, лабораторних занять

Необхідним елементом успішного засвоєння матеріалу навчальної дисципліни є самостійна робота студентів з вітчизняною та закордонною спеціальною літературою, спеціальними засобами оцінки та аналізу ризиків, періодичними виданнями. Основні види самостійної роботи, які запропоновані студентам: передбачає самостійне опрацювання матеріалу лекції; підготовку до практичних занять згідно методичних вказівок до практичних занять та організації самостійної роботи. Презентації використовуються для представлення певних досягнень: демонстрації та обговорення результатів виконання лабораторних робіт та теоретичної підготовки до їх виконання, демонстрації та обговорення результатів самостійної підготовки з окремих питань, що сприяє успішному засвоєнню

### 9.2 Види навчальної діяльності

НД1	Інтерактивні лекції
НД2	Підготовка до лекцій
НД3	Підготовка до лабораторних робіт
НД4	Виконання та презентація результатів лабораторної роботи
НД5	Робота з підручниками та релевантними інформаційними джерелами

## 10. Методи та критерії оцінювання

### 10.1. Критерії оцінювання

Визначення	Чотирибальна національна шкала оцінювання	Рейтингова бальна шкала оцінювання
Відмінне виконання лише з незначною кількістю помилок	5 (відмінно)	$90 \leq RD \leq 100$
Вище середнього рівня з кількома помилками	4 (добре)	$82 \leq RD < 89$
Загалом правильна робота з певною кількістю помилок	4 (добре)	$74 \leq RD < 81$
Непогано, але зі значною кількістю недоліків	3 (задовільно)	$64 \leq RD < 73$

Виконання задовольняє мінімальним критеріям	3 (задовільно)	$60 \leq RD < 63$
Можливе повторне складання	2 (незадовільно)	$35 \leq RD < 59$
Необхідний повторний курс з навчальної дисципліни	2 (незадовільно)	$0 \leq RD < 34$

### 10.2 Методи поточного формативного оцінювання

	Характеристика	Дедлайн, тижні	Зворотний зв'язок
МФО1 Настанови викладача в процесі виконання лабораторних завдань	Настанови викладача в процесі виконання лабораторних завдань	1 тиждень	Спілкування з викладачем під час занять
МФО2 Обговорення та самокорекція виконаної роботи студентами	Обговорення результатів виконання лабораторних робіт, розбір помилок	1 тиждень	Консультації
МФО3 Надання зворотного зв'язку про результати перевірки навчальних досягнень здобувачів за матеріалом, що вивчається	Розбір типових помилок при виконанні лабораторних робіт., рекомендації по їх виправленню. Рекомендації до повторного перегляду окремих розділів навчальних матеріалів	Протягом семестру	Колективне спілкування з викладачем та одногрупниками

### 10.3 Методи підсумкового сумативного оцінювання

	Характеристика	Дедлайн, тижні	Зворотний зв'язок
МСО1 Підготовка та захист звітів з виконання лабораторних робіт	Виконання завдання лабораторної роботи, формулювання висновків за результатами виконання, підготовка звіту, захист результатів та висновків роботи	1 тиждень	Освітня платформа МІКС
МСО2 Проміжний та фінальний модульні контролі	Модульний контроль у формі відповідей на теоретичні питання та виконання практичного завдання	1 тиждень	Освітня платформа МІКС

## Контрольні заходи:

		Максимальна кількість балів	Мінімальна кількість балів	Можливість перескладання з метою підвищення оцінки
<b>2 семестр</b>		<b>100 балів</b>		
МСО1. Підготовка та захист звітів з виконання лабораторних робіт		<b>80</b>		
	8x10	80	Не передбачено	Ні
МСО2. Проміжний та фінальний модульні контролю		<b>20</b>		
	2x10	20	Не передбачено	Ні

Лабораторні завдання, в яких виявлені ознаками академічного плагіату, будуть відхилені з можливістю повторного подання роботи на перевірку. Студенту може бути виставлений залік з дисципліни, якщо він виконав 60% від обсягу завдань та склав модульний контроль

## 11. Ресурсне забезпечення навчальної дисципліни

### 11.1 Засоби навчання

ЗН1	Мультимедіа, відео- і звуковідтворювальна, проєкційна апаратура (відеокамери, проєктори, екрани, смартошки тощо)
ЗН2	Програмне забезпечення: MS Office (Word, Excell), Python 3.x
ЗН3	Комп'ютери, комп'ютерні системи та мережи

### 11.2 Інформаційне та навчально-методичне забезпечення

<b>Основна література</b>	
1	Олещенко Л.М. Технології оброблення великих даних. Конспект лекцій. Київ: вид-во КПІ ім. І.Сикорського, 2021. - 221 с. - <a href="https://ela.kpi.ua/handle/123456789/42206">https://ela.kpi.ua/handle/123456789/42206</a>
2	Таран В.І., Гордієнко Ю.Г., Стіренко С.Г. Технології BigData. Практикум. Київ: вид-во КПІ ім. І.Сикорського, 2021. - 56 с. - <a href="https://ela.kpi.ua/handle/123456789/50014">https://ela.kpi.ua/handle/123456789/50014</a>
3	Олещенко Л.М. Технології оброблення великих даних. Комп'ютерний практикум. Київ: вид-во КПІ ім. І.Сикорського, 2021. - 85 с. - <a href="https://ela.kpi.ua/handle/123456789/45890">https://ela.kpi.ua/handle/123456789/45890</a>
<b>Допоміжна література</b>	
1	Peng, R.D. R Programming for Data Science [Електронний ресурс] / R.D. Peng. — Lean Publishing. — 140 p.

2	Intelligence Science and Big Data Engineering [Електронний ресурс] : 7th International Conference, IScIDE 2017, Dalian, China, September 22-23, 2017, Proceedings / edited by Yi Sun, Huchuan Lu, Lihe Zhang, Jian Yang, Hua Huang. — 1st ed. 2017. — Cham
<b>Інформаційні ресурси в Інтернеті</b>	
1	Наука про дані ( <a href="https://www.coursera.org/specializations/jhu-data-science">https://www.coursera.org/specializations/jhu-data-science</a> )