

І СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

1. Загальна інформація про навчальну дисципліну	
Повна назва навчальної дисципліни	Пайтон та аналіз даних в задачах науки про дані
Повна офіційна назва закладу вищої освіти	Сумський державний університет
Повна назва структурного підрозділу	Факультет електроніки та інформаційних технологій, кафедра прикладної математики та моделювання складних систем
Розробник(и)	Лисенко Олександр Володимирович, д-р фіз.-мат. наук, професор
Рівень вищої освіти	другий рівень вищої освіти, НРК – 7 рівень, QF-LLL – 7 рівень, FQ-EHEA – другий цикл
Семестр вивчення навчальної дисципліни	16 тижнів протягом 1-го семестру
Обсяг навчальної дисципліни	Обсяг дисципліни становить 5 кред. ЄКТС, 150 год., з яких 64 год. становить контактна робота з викладачем (32 год. лекцій, 32 практичних занять)
Мова(и) викладання	Українська
2. Місце навчальної дисципліни в освітній програмі	
Статус дисципліни	Обов'язкова навчальна дисципліна для освітньої програми «Наука про дані та моделювання складних систем»
Передумови для вивчення дисципліни	Необхідні знання з: математичного аналізу, лінійної алгебри, теорії ймовірностей та математичної статистики
Додаткові умови	Відсутні
Обмеження	Обмеження відсутні

3. Мета навчальної дисципліни

Метою навчальної дисципліни є досягнення студентами сучасного фундаментального мислення та системи спеціальних знань у галузі Data Science та здатності їх використовувати для аналізу даних різного типу

4. Зміст навчальної дисципліни

Тема 1 Огляд задач аналізу даних в Data Science, основи Python

Огляд задач машинного навчання. Задачі класифікації та регресії, пошук прихованих структур. Базові типи та конструкції у Python. Організація коду. Функції. Класи та об'єкти. Бібліотеки NumPy, Pandas, Matplotlib, Seaborn. Знайомство Anaconda Navigator, Jupyter Notebook, хмари Google Colab та Kaggle.

Тема 2 Навчання на розмічених даних з реалізацією на мові Python

Лінійні моделі в задачах регресії та класифікації. Усунення перенавчання та оцінювання якості. Дерева рішень, байєсівська класифікація та регресія на Python. Композиції алгоритмів. Метричні алгоритми та SVM на Python.

Тема 3 Пошук структури даних з використанням спеціалізованих пакетів Python

Задача та методи кластеризації. Приклади задач кластеризації. Методи зниження розмірності. Методи відбору ознак. Матричні розкладання. Візуалізація та пошук аномалій. Тематичне моделювання.

Тема 4 Побудова висновків за даними

Довірчі інтервали. Перевірка гіпотез. A/B-тестування. Параметричні та непараметричні критерії. Пошук залежностей в даних. Кореляції. Множинна перевірка. Регресія.

Тема 5 Прикладні задачі аналізу даних та особливості їх реалізації на Python

Прогнозування часових рядів. Аналіз поведінки користувачів. Класифікація зображень. Розпізнавання облич та об'єктів. Аналіз текстів. Рекомендації та ранжування.

5. Очікувані результати навчання навчальної дисципліни

Після успішного вивчення навчальної дисципліни здобувач вищої освіти зможе:

PH1	Використовувати мову програмування Python для аналізу даних різного типу з використанням його основних бібліотек, наприклад, NumPy, SciPy, Matplotlib, Pandas, Seaborn, Scikit-learn.
PH2	Застосовувати на практиці алгоритми класифікації и регресії: лінійні моделі, дерева рішень. Підвищувати якість окремих алгоритмів, використовуючи техніку побудови композицій, в тому числі створювати випадкові ліси, застосовувати метод градієнтного бустінгу. Підбирати параметри моделі, обчислювати метрики якості.
PH3	Використовувати алгоритми кластеризації даних, матричні розкладання, розв'язувати задачі тематичного моделювання, знижувати

	розмірність даних, шукати аномалії та візуалізувати багатовимірні дані.
РН4	Організовувати експерименти, виконувати А/В-тестування, використовувати універсальні методи оцінки параметрів та перевірки гіпотез, знаходити кореляції та причинно-наслідкові зв'язки.
РН5	Зводити задачу замовника до формальної постановки задачі машинного навчання, перевіряти якість побудованої моделі на історичних даних і в онлайн-експерименті, використовувати схеми вирішення основних типів прикладних задач
6. Роль навчальної дисципліни у досягненні програмних результатів	
Програмні результати, досягнення яких забезпечує навчальна дисципліна:	
ПРН1	Знати сучасні мови програмування та основні напрями задач, в яких вони застосовуються найбільш ефективно. Бути здатним до самостійного оволодіння мовою програмування, найбільш пристосованою для певної задачі та її надбудовами.
ПРН9	Застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.
ПРН17	Володіти математичними методами первинної обробки великих наборів даних. Вміти обирати до застосовування оптимальні методи для конкретної задачі побудови моделі поведінки складної системи за існуючим набором даних.
ПРН18	Уміти розробляти та використовувати на практиці алгоритми, пов'язані з спрощенням даних, що описують поведінку системи, класифікацією даних за певними ознаками без навчання та за попередньої наявності класів даних.
ПРН19	Уміти оцінити на адекватність результат обробки великих масивів даних.
ПРН20	Уміти застосовувати набуті знання про дані на практиці у сферах економіки, соціології, медицини і т. ін.
7. Види навчальних занять та навчальної діяльності	
7.1 Види навчальних занять	
<p>Видами навчальних занять при вивченні дисципліни є лекції (on-line лекції) (Л) та практичні заняття (ПР):</p> <p>Тема 1 Огляд задач аналізу даних в Data Science, основи Python Л 1 Огляд задач аналізу даних в Data Science, основи Python</p> <p>ПР 1 Основи Python</p> <p>Тема 2 Навчання на розмічених даних з реалізацією на мові Python Л 2 Лінійні моделі в задачах регресії та класифікації. Усунення перенавчання та оцінювання якості.</p>	

Л 3 Дерева рішень. Композиції алгоритмів.
Л 4 Байєсівська класифікація та регресія.
Л 5 Метричні алгоритми та SVM на Python.

ПР 2 Лінійні моделі в задачах регресії та класифікації. Усунення перенавчання та оцінювання якості.

ПР 3 Дерева рішень. Композиції алгоритмів.
ПР 4 Байєсівська класифікація та регресія.
ПР 5 Метричні алгоритми та SVM на Python.

Тема 3 Пошук структури даних з використанням спеціалізованих пакетів Python

Л 6 Задача та методи кластеризації. Приклади задач кластеризації.
Л 7 Методи зниження розмірності. Методи відбору ознак.
Л 8 Матричні розкладання. Візуалізація та пошук аномалій.
Л 9 Тематичне моделювання.

ПР 6 Задача та методи кластеризації. Приклади задач кластеризації.
ПР 7 Методи зниження розмірності. Методи відбору ознак.
ПР 8 Матричні розкладання. Візуалізація та пошук аномалій.
ПР 9 Тематичне моделювання.

Тема 4 Побудова висновків за даними

Л 10 Довірчі інтервали. Перевірка гіпотез. A/B-тестування.
Л 11 Параметричні та непараметричні критерії.
Л 12 Пошук залежностей в даних. Кореляції.
Л 13 Множинна перевірка. Регресія.

ПР 10 Довірчі інтервали. Перевірка гіпотез. A/B-тестування.
ПР 11 Параметричні та непараметричні критерії.
ПР 12 Пошук залежностей в даних. Кореляції.
ПР 13 Множинна перевірка. Регресія.

Тема 5 Прикладні задачі аналізу даних та особливості їх реалізації на Python

Л 14 Прогнозування часових рядів. Аналіз поведінки користувачів.
Л 15 Класифікація зображень. Розпізнавання облич та об'єктів.
Л 16 Аналіз текстів.
Л 17 Аналіз текстів. Рекомендації та ранжування.

ПР 14 Прогнозування часових рядів. Аналіз поведінки користувачів.
ПР 15 Класифікація зображень. Розпізнавання облич та об'єктів.
ПР 16 Аналіз текстів.
ПР 17 Аналіз текстів. Рекомендації та ранжування.

7.2 Види навчальної діяльності

НД 1. Вивчення теоретичного матеріалу за лекціями та спеціальною літературою.
НД 2. Виконання практичних завдань.

- НД 3. Виконання ситуативних вправ
 НД 4. Участь у лекціях-дискусіях
 НД 5. Підготовка звітів за результатами практичних завдань
 НД 6. Самонавчання

8. Методи викладання, навчання

Дисципліна передбачає навчання через:

- МН 1. Проблемні-лекції, лекції-візуалізації (on-line лекції)
 МН 2. Практико-орієнтоване навчання.
 МН 3. Проблемно-пошуковий метод
 МН 4. Творчий метод

Лекції надають студентам теоретичний матеріал, що є основою для самостійного навчання здобувачів вищої освіти. Лекції доповнюються практичними заняттями, що надають студентам можливість застосовувати теоретичні знання на конкретних прикладах.

Формування загальних компетентностей (soft skills):

- ЗК1. Здатність учитися, здобувати нові знання, уміння, у тому числі в галузі, відмінній від професійної
 ЗК2. Здатність гнучко адаптуватися до різних професійних ситуацій, проявляти творчий підхід, ініціативу.
 ЗК3. Здатність критично оцінювати й переосмислювати накопичений досвід (власний і чужий), аналізувати свою професійну й соціальну діяльність.
 ЗК5. Здатність ефективно використовувати комп'ютерні та інформаційні технології в професійній діяльності
 ЗК6. Здатність здійснювати виробничу чи прикладну діяльність у міжнародному середовищі.

9. Методи та критерії оцінювання

9.1. Критерії оцінювання

Оцінювання знань студентів здійснюється за 100- бальною шкалою, яка переводиться відповідно у національну шкалу («відмінно», «добре», «задовільно», «незадовільно») та шкалу європейської кредитно-трансферної системи (ЄКТС–А, В, С, D, E, FX, F) відповідно до таблиці:

Сума балів за шкалою університету*	Оцінка ЄКТС	Оцінка за національною шкалою, 4 – бальна шкала	Визначення
90 – 100	A	Відмінно	Відмінне виконання лише з незначною кількістю помилок
82 – 89	B	Добре	Вище середнього рівня з кількома помилками
74 – 81	C	Добре	Загалом правильна робота з певною кількістю помилок
64 – 73	D	Задовільно	Непогано, але зі значною кількістю недоліків

60 – 63	E	Задовільно	Виконання задовольняє мінімальні критерії
35 – 59	FX	Незадовільно	Можливе повторне складання
0– 34	F	Незадовільно	Необхідний повторний курс з навчальної дисципліни

Студент, який протягом навчального періоду виконав всі заплановані види навчальної роботи та за наслідками модульних атестацій набрав необхідну, яка відповідає позитивній оцінці, кількість рейтингових балів (не менше 60), отримує семестрову оцінку у відповідності до набраних рейтингових балів.

Студент, який протягом поточної роботи не набрав кількість рейтингових балів, що відповідає позитивній оцінці, але не менше 35, зобов'язаний перескладати захід підсумкового семестрового контролю, який може складати два рази: один роз викладачеві, другий – комісії.

Студент, який за наслідками модульних атестацій не набрав мінімально необхідної кількості рейтингових балів (не менше 35) не допускається до повторного складання підсумкового семестрового контролю і отримує оцінку «неприйнятно» (за шкалою ECTS – «F»).

Заохочувальні рейтингові бали нараховуються за публікацію та виступ, що пов'язаний дисципліною «Пайтон та аналіз даних в задачах науки про дані», на конференції, за активну роботу на лабораторних та лекційних заняттях (виступ перед аудиторією з доповіддю, створення додаткових комп'ютерних програм).

9.2 Методи поточного формативного оцінювання

За дисципліною передбачені такі методи поточного формативного оцінювання:

- 1 Опитування та усні коментарі викладача за його результатами
- 2 Наставови викладача в процесі виконання практичних завдань
- 3 Самооцінювання виконання практичної роботи
- 4 Обговорення та взаємооцінювання студентами виконаних практичних завдань.

9.3 Методи підсумкового сумативного оцінювання

Оцінка студента формується таким чином:

1. Виконання та захист практичних робіт (ПР) – максимально 48 балів за всі практичні роботи (16 практичних робіт).
2. Підсумковий модульний контроль (ПМК) – максимально 44 бали (2 підсумкових контролю)
3. Тестове опитування під час лекцій (ТЛ) – максимально 8 балів за всі лекції (16 лекцій).

Методи контролю:

1. Перевірка виконаних практичних робіт
2. Перевірка модульного контрольного завдання
3. Перевірка тестових завдань з теоретичного матеріалу

10. Ресурсне забезпечення навчальної дисципліни

10.1 навчання	Засоби 1 Комп'ютери, комп'ютерні системи та мережи (ЗН1) 2 Мультимедіа, відео- і звуковідтворювальна, проєкційна апаратура (відеокамери, проєктори, екрани, смартдошки тощо) (ЗН2) 3 Прикладне програмне забезпечення (ЗН3) 4 Бібліотечні фонди (ЗН4)
10.2 навчально-методичне забезпечення	Основна література: 1 Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. - М.: ДМК Пресс, 2017. - 418 с. 2 Золотых Н. Ю. Машинное обучение и анализ данных / Н. Ю. Золотых. – 2018. – 693 с. 3 Вьюгин В. В. Математические основы теории машинного обучения и прогнозирования / В. В. Вьюгин. – М., 2013. – 387 с. Додаткова література: 4 Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с. 5 Мюллер А. Введение в машинное обучение с помощью Python : руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо. – М., 2018. – 480 с.