

Міністерство освіти і науки України

Національний аерокосмічний університет ім. М.Є. Жуковського
«Харківський авіаційний інститут»

Кафедра «Вищої математики та системного аналізу» (№ 405)

ЗАТВЕРДЖУЮ

Керівник проектної групи


(підпис)

О.Г. Ніколаєв
(ініціали та прізвище)

« 01 » _____ 09 _____ 2020 р.

**РОБОЧА ПРОГРАМА ВИБІРКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Прикладний аналіз даних на PYTHON

(назва навчальної дисципліни)

Галузь знань: 12 «Інформаційні технології»
(шифр і назва галузі знань)

Спеціальність: 124 «Системний аналіз»
(шифр і назва спеціальності)

Освітня програма: Системний аналіз і управління
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший(бакалаврський)

Харків 2020 рік

Робоча програма «Прикладний аналіз даних на PYTHON»

(назва дисципліни)

для студентів за спеціальністю 124 «Системний аналіз»
освітньою програмою «Системний аналіз і управління»

«20» червня 2020 р., - 9 с.

Розробник: Макарічев В.О., доцент каф. 405, к. ф.-м. н.


(підпис)

Робочу програму розглянуто на засіданні кафедри вищої математики та систем-
ного аналізу

(назва кафедри)

Протокол № 11 від «26» 06 2020 р.

Завідувач кафедри д.ф.-м.н., професор

(науковий ступінь та вчене звання)


(підпис)

О.Г. Ніколаєв

(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліна
		Денна форма навчання
Кількість кредитів –5	Галузь знань 12 «Інформаційні технології» (шифр та найменування) Спеціальність 124 «Системний аналіз» (код та найменування) Освітня програма «Системний аналіз і управління» (найменування) Рівень вищої освіти: перший (бакалаврський)	Цикл професійної підготовки
Кількість модулів – 2		Навчальний рік
Кількість змістовних модулів – 2		2020/2021
Індивідуальне науково-дослідне завдання <u>розрахункова робота</u> (назва)		Семестр
		5 - й
Загальна кількість годин – денна – 64/150		Лекції
		32 год.
Кількість тижневих годин для денної форми навчання: аудиторних – 4 год. самост. робота – 5,4 год.		Практичні
		32 год.
		Лабораторні
		-
		Самостійна робота
		86 год.
		Індивідуальна робота
		-
		Вид контролю
		залік

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить:

для денної форми навчання – 64/86.

2. Мета та завдання навчальної дисципліни

Мета: формування навичок застосування мови програмування Python та відповідних програмних інструментів для створення програм для застосування у професійній діяльності.

Завдання: набуття навичок самостійної розробки програм на мові програмування Python, для наукових розрахунків, обробки даних, візуалізації результатів у вигляді графіків та ін.

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких **компетентностей**:

- здатність до абстрактного мислення, аналізу та синтезу (ЗК 1);
- здатність застосовувати знання у практичних ситуаціях (ЗК 2);
- здатність планувати і управляти часом (ЗК 3);
- здатність знати та розуміти предметну область і професійну діяльність (ЗК 4);
- здатність до пошуку, оброблення та аналізу інформації з різних джерел (ЗК 7);
- здатність оцінювати та забезпечувати якість виконуваних робіт (ЗК 14);
- здатність формалізувати проблеми, описані природною мовою, у тому числі за допомогою математичних методів, застосовувати загальні підходи до математичного моделювання конкретних процесів (ФК 2);
- здатність використовувати сучасні інформаційні технології для комп'ютерної реалізації математичних моделей та прогнозування поведінки конкретних систем, а саме: об'єктно-орієнтований підхід при проектуванні складних систем різної природи, прикладні математичні пакети, застосування баз даних і знань, використання моделей алгоритмічних обчислень, оцінювання їх ефективності та складності для адекватного моделювання предметних областей (ФК 8);
- здатність представляти математичні аргументи і висновки з них з ясністю і точністю і в таких формах, які підходять для аудиторії як усно так і в письмовій формі (ФК 10).

Програмні результати навчання:

- володіти сучасними методами розробки програм і програмних комплексів, та прийняття оптимальних рішень щодо складу програмного забезпечення, алгоритмів, процедур і операцій (ПРН 8);
- вміти створювати ефективні алгоритми для обчислювальних задач системного аналізу та систем підтримки прийняття рішень (ПРН 9);

Міждисциплінарні зв'язки: програмування та алгоритмічні мови, програмування на Python.

3. Програма навчальної дисципліни

Модуль 1

Змістовий модуль 1. Засоби обробки природної мови

Тема 1. Засоби обробки природної мови (NLP)

Постановка задач обробки природної мови. NLP-бібліотеки TextBlob, NLTK, Textatistic і spaCy. Розбиття текстів на слова та речення. Емоційний окрас висловів та його аналіз. Засоби перевірки. Визначення мови тексту і переклад. Видалення ігнорованих слів з тексту. Визначення удобочитання тексту за допомогою Textatistic.

Тема 2. Засоби глибокого аналізу даних Twitter

Twitter API. Використання Twitter Search API. Twitter Streaming API. Виконання аналізу емоційного забарвлення твітів. Twitter Trends API. Засоби folium і OpenStreetMap. Способи зберігання твітів.

Модульний контроль

Модуль 2

Змістовий модуль 2. Засоби машинного навчання

Тема 3. Засоби машинного навчання: класифікація, регресія і кластеризація

Використання scikit-learn. Візуалізація і дослідження даних засобами Seaborn і Matplotlib. Машинне навчання з учителем: методи k-найближчих сусідів і лінійної регресії. Множинна класифікація. Поділ набору даних. Метод k-прохідної перехресної перевірки. Вимірювання ефективності моделі. Множинна лінійна регресія. Просторова згортка з використанням PCA і t-SNE. Машинне навчання без учителя кластеризацією методом k-середніх.

Тема 4. Засоби глибокого навчання

Концепція нейронних мереж. Бібліотека Keras: рівні, функції активації, функції втрат і оптимізатори. Згорткова нейронна мережа Keras (CNN). Рекуррентна нейронна мережа Keras (RNN). TensorBoard: візуалізація ходу глибокого навчання мереж. Попередньо навчені нейронні мережі, що поставляються з Keras.

Модульний контроль

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі										
		л	п	лаб	інд	с.р.						
1	2	3	4	5	6	7	8	9	10	11	12	13
Семестр 4												
Модуль 1												
Змістовий модуль 1. Засоби обробки природної мови												
Тема 1. Засоби обробки природної мови (NLP)	36	8	8	–	–	20	–	–	–	–	–	–
Тема 2. Засоби глибокого аналізу даних Twitter	31	8	6	–	–	17	–	–	–	–	–	–
Модульний контроль	2	–	2	–	–	–	–	–	–	–	–	–
Разом за змістовим модулем 1	69	16	16	–	–	37	–	–	–	–	–	–

1	2	3	4	5	6	7	8	9	10	11	12	13
Модуль 2												
Змістовий модуль 2. Засоби машинного навчання												
Тема 3. Засоби машинного навчання: класифікація, регресія і кластеризація	41	8	6	–	–	27	–	–	–	–	–	–
Тема 4. Засоби глибокого навчання	38	8	8	–	–	22	–	–	–	–	–	–
Індивідуальне завдання	–	–	–	–	–	–	–	–	–	–	–	–
Модульний контроль	2	–	2	–	–	–	–	–	–	–	–	–
Разом за змістовим модулем 2	81	16	16	–	–	49	–	–	–	–	–	–
Семестровий контроль: іспит (проводиться у разі відмови студента від балів поточного тестування та за наявності допуску)												
Усього годин	150	32	32	–	–	86	–	–	–	–	–	–

5. Теми семінарських занять

№ з/п	Назва теми	Кількість годин
1		

6. Теми практичних занять

№ з/п	Назва теми	Кільк. год.
1	2	3
1 - 4	Використання засобів обробки природної мови (NLP) (Тема 1).	8
5 - 7	Використання засобів глибокого аналізу даних Twitter (Тема 2).	6
8.	Модульний контроль	2
9 - 11.	Використання засобів машинного навчання (класифікація, регресія і кластеризація) (Тема 3).	6
12 - 15.	Використання засобів глибокого навчання (Тема 4).	8
16.	Модульний контроль	2
	Разом	32

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1		

8. Самостійна робота

№ з/п	Назва теми	Кільк. годин
1.	Бібліотеки Natural Language Toolkit (NLTK), Gensim, polyglot та CoreNLP: огляд, особливості використання (Тема 1).	20
2.	Практичні кейси використання засобів глибокого аналізу даних Twitter. (Тема 2).	17
3.	Бібліотеки Theano, TensorFlow, PyTorch: огляд та практичні кейси використання. (Тема 3).	27
4.	Бібліотеки NeuroLab, Scikit-Neural Network, Lasagne: огляд та практичні кейси використання. (Тема 4).	22
	Разом	86

9. Індивідуальні завдання

№ з/п	Назва теми
1.	Виконання розрахункової роботи на тему «Застосування засобів Python для прикладного аналізу даних»

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами, опублікованими кафедрою (методичні посібники), проведення олімпіад.

11. Методи контролю

Проведення поточного контролю, письмового модульного контролю, фінальний контроль у вигляді іспиту.

12. Критерії оцінювання та розподіл балів, які отримують студенти

12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0...1	8	0...8
Виконання та захист практичних робіт	0...2	7	0...14
Самостійна робота	0...5	2	0...10
Модульний контроль	0...20	1	0...20
Змістовний модуль 2			
Робота на лекціях	0...1	8	0...8
Виконання та захист практичних робіт	0...2	7	0...14
Самостійна робота	0...10	2	0...20
Модульний контроль	0...20	1	0...20
Виконання та захист РР	0...20	1	0...20
Всього за семестр (*)			0...134

(*) Якщо кількість модульних балів у студента перевищує 100, то в якості підсу-

мкової оцінки виставляється 100 балів.

Семестровий контроль (залік) проводиться у разі відмови студента від балів поточного тестування та за наявності допуску до заліку. При складанні семестрового іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з 2 теоретичних та 3 практичних запитань, за кожне з яких можна отримати максимум 20 балів.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

- базові можливості з обробки природньої мови;
- засоби Python, призначені для класифікації, дослідження регресії та кластеризації;
- базові засоби Python для роботи зі штучними нейронними мережами.

Необхідний обсяг умінь для одержання позитивної оцінки:

- застосовувати аналітичний апарат та навички розробки програмного забезпечення в навчальному процесі і науково-дослідницькій діяльності;
- визначати межу можливих застосувань існуючих засобів;
- досліджувати питання коректності постановки задач і існування розв'язків.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити індивідуальне завдання та здати тестування. Мати уявлення про базові можливості Python щодо обробки природньої мови, класифікації, дослідження регресії та кластеризації, роботи зі штучними нейронними мережами.

Добре (75-89). Твердо знати мінімум, захистити індивідуальне завдання, здати тестування та поза аудиторну самостійну роботу. Вміти розробляти програмне забезпечення з використанням бібліотек Python для обробки природньої мови, класифікації, дослідження регресії та кластеризації, роботи зі штучними нейронними мережами.

Відмінно (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та вміти застосовувати їх.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

13. Методичне забезпечення

Електронний ресурс, на якому розміщено навчально-методичний комплекс дисципліни: [//library.khai.edu](http://library.khai.edu). Комплекс включає в себе такі обов'язкові складові:

- робоча програма дисципліни;
- конспект лекцій;
- методичні вказівки та рекомендації для виконання розрахункових та практичних робіт, а також рекомендації для самостійної підготовки;
- тематики індивідуальних завдань;
- приклади розв'язування типових задач чи виконання типових завдань;
- питання для контрольних заходів.

14. Рекомендована література

Базова

1. Основи програмування. Python [Електронний ресурс]: підручник для студ. спеціальності 122 "Комп'ютерні науки", спеціалізації "Інформаційні технології в біології та медицині" / А. В. Яковенко. – Київ : КПІ ім. Ігоря Сікорського, 2018. – 195 с.
2. S. Raschka, V. Mirjalili. Python Machine Learning. – Packt Publishing, 2019. – 770 p.
3. F. Chollet. Deep Learning with Python. – Manning Publications, 2017. – 384 p.
4. Документація Python: <https://www.python.org/doc/>
5. Документація TextBlob: https://textblob.readthedocs.io/en/dev/api_reference.html
6. Документація Twitter API: <https://developer.twitter.com/en/docs/tools-and-libraries>
7. Документація Keras: <https://keras.io/api/>

Допоміжна

1. M. Lutz. Learning Python, 5th ed. – O'Reilly Media, 2013. – 1648 p.
2. P. Deitel, H. Deitel. Python for Programmers: with Big Data and Artificial Intelligence Case Studies. – Pearson, 2019. – 640 p.
3. Документація з офіційних сайтів бібліотек Python.

15. Інформаційні ресурси

Сайт бібліотеки Національного аерокосмічного університету ім. М.Є. Жуковського «Харківський авіаційний інститут» <https://library.khai.edu>.