

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інженерії програмного забезпечення (№ 603)

ЗАТВЕРДЖУЮ

Гарант освітньої програми/



І.Б. Туркін

(ініціали та прізвище)

« 30 » 08 _____ 2021 р.

**СИЛАБУС ВИБІРКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Основи BIG Data

(назва навчальної дисципліни)

Галузь знань: 12 «Інформаційні технології»

(шифр і найменування галузі знань)

Спеціальність: 121 «Інженерія програмного забезпечення»

(код та найменування спеціальності)

Освітня програма: «Інженерія програмного забезпечення»

(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2021 рік

Розробник: Куреннов С.С., д.т.н., доцент
(прізвище та ініціали, посада, науковий ступінь і вчене звання)


(підпис)

Силабус навчальної дисципліни розглянуто на засіданні кафедри інженерії програмного забезпечення (№ 603)

(назва кафедри)

Протокол № 2 від « 31 » серпня 2021 р.

Завідувач кафедри, д.т.н., професор
(науковий ступінь і вчене звання)


(підпис)

І. Б. Туркін
(ініціали та прізвище)

Погоджено з представником здобувачів освіти:
Представник студентського самоврядування


(підпис)

Колодій Д.В.
(ініціали та прізвище)

1. Загальна інформація про викладача



Куреннов Сергій Сергійович, доктор технічних наук, доцент, професор кафедри вищої математики і системного аналізу (405). Викладає лекційні курси *«Вища математика»*, *«Теорія ймовірності і математична статистика»*, *«Математичні методи логістики і управління запасами»*, *«Основи Data Science»*, *«Інтелектуальний аналіз даних»*, *«Теорія пружності»* (для аспірантів).

Напрями наукових досліджень: аналітичні і числові методи дослідження напруженого стану шаруватих конструкцій, математичне моделювання і числові методи, точні і евристичні методи оптимізації, статистичний аналіз даних.

Є автором або співавтором навчальних посібників *«Логістика та керування запасами»*, *«Основи теорії напружено-деформованого стану»* (2018 р.), *«Основи лінійного та цілочислового програмування»* (2019 р.), *«Рівняння математичної фізики»* (2020 р.).

Контактні дані: тел. 050 259 64 21, kurennov.ss@gmail.com

Профіль у Scopus:

<https://www.scopus.com/authid/detail.uri?authorId=11839716000>

2. Опис навчальної дисципліни

Семестр, в якому викладається дисципліна – 8

Обсяг дисципліни: 5 кредитів ЄКТС/ 150 годин, у тому числі аудиторних – 64 год., самостійної роботи здобувачів – 86 год.

Форма здобуття освіти – денна/дистанційна

Дисципліна – вибіркова

Види навчальної діяльності – лекції, лабораторні роботи

Види контролю – залік/диф. залік

Мова викладання – українська

Пререквізити – вища математика, теорія ймовірності і математична статистика.

Кореквізити – Реляційні бази даних, програмування мовою Python, оптимізація та просування сайтів, алгоритми підтримки рішень.

3. Мета та завдання навчальної дисципліни

Мета вивчення: ознайомлення студентів із сучасними технологіями аналізу даних, роботи з великими масивами даних та технологіями пошуку закономірностей у даних.

Завдання: опанування теоретичними навичками аналізу даних, зокрема кластеризації даних, побудови регресійних моделей, аналізу часових рядів, тощо.

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

Загальні компетентності:

- ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК02. Здатність застосовувати знання у практичних ситуаціях.
- ЗК05. Здатність вчитися і оволодівати сучасними знаннями.
- ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

різних джерел.

- ЗК10. Здатність діяти соціально відповідально та свідомо.

Фахові компетентності:

- ФК02. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

- ФК03. Здатність розробляти архітектури, модулі та компоненти програмних систем.

- ФК07. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

- ФК08. Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення.

- ФК10. Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.

- ФК11. Здатність реалізовувати фази та ітерації життєвого циклу програмних систем та інформаційних технологій на основі відповідних моделей і підходів розробки програмного забезпечення.

- ФК13. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.

Програмні результати навчання:

- ПРН01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

- ПРН02. Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.

- ПРН05. Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.
- ПРН06. Уміння вибирати та використовувати відповідну задачі методологію створення програмного забезпечення.
- ПРН07. Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.
- ПРН09. Знати та вміти використовувати методи та засоби збору, формулювання та аналізу вимог до програмного забезпечення.
- ПРН10. Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування.
- ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.
- ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.

Міждисциплінарні зв'язки: дисципліні передують курси: «Вища математика», «Алгоритми та структури даних», «Бази даних», «Системи штучного інтелекту»; дисципліна потрібна для подальшого вивчення курсів «Дипломне проектування».

4. Зміст навчальної дисципліни

Модуль 1

Змістовий модуль 1. Регресійний аналіз

Тема 1. Вступ до навчальної дисципліни «Основи Big Data».

Анотація. Основні задачі *Big Data*, методи і області застосування. Класифікація даних. Етапи процесу аналізу даних. Збір даних. Попередня обробка даних. Статистичні викиди. Неповні дані. Приведення даних до безрозмірної форми, центрування і нормалізація даних. Візуалізація даних. Гістограми, поля розсіювання, теплові карти.

Форма занять – лекції, лабораторні заняття, самостійна робота.

Аудиторне навантаження – лекції 4 години, лаб. – 4 години.

Теми лабораторних занять – візуалізація даних.

Самостійна робота – 8 годин. Теми самостійної роботи: Пошук даних. Попередня обробка даних. Нормалізація даних, пошук статистичних викидів.

Тема 2. Елементи теорії ймовірності і математичної статистики

Анотація. Ймовірність і частота. Основні відомості про ймовірність. Формула Байеса. Наївний баєсівський класифікатор спаму. Варіаційний ряд і його характеристики. Основні види статистичних розподілів. Кореляція і регресія. Парадокси і похибки кореляції. Лінійна регресія. Лінійна множинна регресійна модель. Нелінійна регресія. Перевірка статистичних гіпотез. Перевірка адекватності моделі і значущості коефіцієнтів моделі.

Форма занять – лекції, лабораторні заняття, самостійна робота.

Аудиторне навантаження – лекції 12 годин, лаб. – 12 годин.

Теми лабораторних занять – Лінійна регресія. Остаточна дисперсія. Лінійна множинна регресійна модель. Нелінійна регресія. Статистичні гіпотези. Метод бутстрап.

Самостійна робота – 20 годин. Теми самостійної роботи: Залежні і незалежні події. Повна ймовірність, умовна ймовірність. Основні статистичні розподіли. Вибірки, репрезентативність. Статистичні гіпотези і критерії. Матрична форма лінійна множинної регресії.

Тема 3. Метод головних компонент.

Анотація. Матриця повороту системи координат. Власні вектори. Метод головних компонент – ортогональне перетворення системи координат у просторі даних і введення прихованих змінних. Оцінка кількості необхідних головних компонент за різними методами та зменшення розмірності моделі.

Форма занять – лекції, лабораторні заняття, самостійна робота.

Аудиторне навантаження – лекції 4 години, лаб. – 4 години.

Теми лабораторних занять – Ортогональні перетворення. Головні компоненти. Метод кам'яної осипі.

Самостійна робота – 12 годин. Ортогональні перетворення. Аналіз змісту прихованих змінних. Методи зменшення розмірності моделі.

Модульний контроль

МОДУЛЬ 2

Змістовий модуль 1. Кластерний аналіз і глибоке навчання

Тема 4. Кластерний аналіз.

Анотація. Задачі кластерного аналізу. Основні поняття кластеризації. Агломеративні та дивізійні методи. Метрики для елементів та кластерів. Ієрархічна класифікація. Дендрограма. Оцінювання якості кластеризації. Суміш Гуса. EM - алгоритм. Метод k -середніх і його варіанти. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.

Форма занять – лекції, лабораторні заняття, самостійна робота.

Аудиторне навантаження – лекції 6 годин, лаб. – 8 годин.

Теми лабораторних занять – Види кластеризації даних. Відстань і метрика простору даних. Метод k -середніх.

Самостійна робота – 20 годин. Умови застосування кластерного аналізу. Метрика і відстань. Варіанти методу k -середніх.

Тема 5. Нейронні мережі. Deep learning.

Анотація. Персептрон. Порогові функції (функції активації). Багатошарові персептрони. Навчання. Елементи теорії оптимізації функцій багатьох змінних. Метод градієнтного спуску. Метод зворотного поширення помилки.

Форма занять – лекції, лабораторні заняття, самостійна робота.

Аудиторне навантаження – лекції 6 годин, лаб. – 4 години.

Теми лабораторних занять – Пошук екстремуму функції багатьох змінних. Градієнтний спуск. Метод зворотного поширення помилки.

Самостійна робота – 22 години. Методи пошуку екстремуму функцій багатьох змінних. Згорткові нейронні мережі. Аналіз текстових даних.

Модульний контроль

Структура навчальної дисципліни

Назви змістового модуля і тем	Кількість годин				
	усього	у тому числі			
		л	п	лаб	с.р.
1	2	3	4	5	6
Модуль 1					
Змістовий модуль 1. Регресійний аналіз					
Тема 1. Вступ «Основи Big Data».	16	4		4	8
Тема 2. Елементи теорії ймовірності і математичної статистики	44	12		12	20
Тема 3. Метод головних компонент	20	4		4	12
Модульний контроль	2				2
Разом за змістовим модулем 1	82	20		20	42
Модуль 2					
Змістовий модуль 1. Кластерний аналіз і глибоке навчання					
Тема 4. Кластерний аналіз	34	6		8	20
Тема 5. Глибоке навчання	32	6		4	22
Модульний контроль	2				2
Разом за змістовим модулем 1	68	12		12	44
Усього годин	150	32		32	86

5. Індивідуальні завдання

Не передбачено.

6. Методи навчання

- За джерелами придбання знань – словесні: лекція (вступна, традиційна, проблемна, з помилками), бесіда (евристична), диспут, дискусія, робота з друкованими та інтернет-джерелами; наочні: ілюстрація, спостереження; практичні: лабораторна робота, курсовий проект.

- За характером пізнавальної діяльності тих, хто навчається – інформаційно-репродуктивний, репродуктивний, проблемне викладання, частково-пошуковий.
- За логікою пізнання – індуктивний, дедуктивний, аналогій, вивідних знань.
- Методи перевірки й оцінки знань, умінь, навичок: спостереження, усне опитування, контрольні роботи, програмований контроль, тестування (традиційне та машинне).

7. Методи контролю

- Опитування.
- Модульні контрольні роботи.
- Форма підсумкового контролю успішності навчання: іспит (письмово)

8. Критерії оцінювання та розподіл балів, які отримують студенти

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях			
Робота на практичних заняттях			
Виконання і захист лабораторних (практичних) робіт	5...10	3	15...30
Модульний контроль	12...25	1	12...25
Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях			
Робота на практичних заняттях			
Виконання і захист лабораторних (практичних) робіт	9...10	2	18...20
Модульний контроль	15...25	1	15...25
Виконання і захист РГР (РР, РК)			
Усього за семестр			60...100

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування й за наявності допуску до іспиту. Під час складання семестрового іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту/заліку складається з двох теоретичних питань (кожне питання 35 балів) та одного практичного питання (питання оцінюється в 30 балів).

Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Здати тестування. розробляти й обґрунтовувати структуру і мету аналізу даних; обґрунтовувати вибір методу аналізу, перевіряти дані на повноту і якість, застосовувати методи аналізу даних і оцінювати результати дослідження, вміти подати результати дослідження у зручній формі.

Добре (75-89). Твердо знати мінімум. Досконало знати розробляти й обґрунтовувати структуру і мету аналізу даних; обґрунтовувати вибір методу аналізу, перевіряти дані на повноту і якість, застосовувати методи аналізу даних і оцінювати результати дослідження. Вміти оцінювати якість обраної моделі і вміти обрати шлях розвитку дослідження. Вміти подати результати дослідження у зручній формі.

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

9. Політика навчального курсу

Всі навчальні матеріали курсу розташовано у системі дистанційного навчання «Ментор». У разі відсутності здобувача вищої освіти на лекції або на лабораторній роботі він має можливість вивчити матеріал та виконати лабораторну роботу і надіслати її викладачу у будь-який зручний для себе час.

13. Методичне забезпечення

1. Розроблений лекційний курс та комплекс презентацій
2. Розроблені питання для модульних контрольних робіт
3. Розроблені питання для підсумкового контролю успішності навчання
4. Лабораторні роботи
5. Дібрані матеріали для самостійної роботи студентів

<https://mentor.khai.edu/course/view.php?id=4867>

11. Рекомендована література

Базова:

1. *Бахрушин В.Є.* Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
2. *Шумейко А. А.* Интеллектуальный анализ данных (Введение в Data Mining) : учеб.пособ. / А. А. Шумейко, С. Л. Сотник. – Днепропетровск: Белая Е. А., 2012. – 212 с.
3. *Олійник А. О.* Интеллектуальный анализ даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.

Допоміжна:

4. *Мюллер, Джон Пол, Массарон, Лука.* Python и наука о данных для чайников, 2-е изд.: Пер. с англ. — СПб.: ООО “Диалектика”, 2020. — 512 с.
5. *Силен Деви, Мейсман Арно, Али Мохамед* Основы Data-Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.
6. *Бин Анналин, Су Кеннет* Теоретический минимум по Big Data. Все что нужно знать о больших данных. – СПб.: Питер, 2019. – 208 с.
7. *Грас Д.* Data Science. Наука о данных с нуля: Пер. с англ. -2-е изд., перераб. и доп. -СПб.: БХВ-Петербург, 2021. -416 с.: ил.

12. Інформаційні ресурси

<https://library.khai.edu/>