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

Кафедра комп'ютерних наук та інформаційних технологій (№ 302)

ЗАТВЕРДЖУЮ

Голова НМК 2

_____ Дмитро КРИЦЬКИЙ
(підпис) (ініціали та прізвище)

«31» 08 2023 р.

СИЛАБУС ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Машинне навчання на мові Python
(назва навчальної дисципліни)

Галузі знань: 10 Природничі науки, 11 Математика та статистика, 12 Інформаційні технології, 15 Автоматизація та приладобудування, 16 Хімічна та біоінженерія, 17 Електроніка та телекомунікації, 19 Архітектура та будівництво, 27 Транспорт

Спеціальності: 101 Екологія, 103 Науки про Землю, 113 Прикладна математика, 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія, 124 Системний аналіз, 125 Кібербезпека, 126 Інформаційні системи та технології, 151 Автоматизація та комп'ютерно-інтегровані технології, 152 Метрологія та інформаційно-вимірювальна техніка, 153 Мікро- та наносистемна техніка, 163 Біомедична інженерія, 172 Телекомунікації та радіотехніка, 173 Авіоніка, 193 Геодезія та землеустрій, 272 Авіаційний транспорт

Освітні програми: усі освітні програми відповідних спеціальностей

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Силабус введено в дію з 01.09.2023 року

Харків – 2023 р.

Розробник: Олександр ПРОХОРОВ, професор, д.т.н., професор
(прізвище та ініціали, посада, науковий ступінь та вчене звання) (підпис)

Силабус навчальної дисципліни розглянуто на засіданні кафедри комп'ютерних наук та інформаційних технологій (№ 302)

Протокол № 659/09 від « 29 » 08 2023 р.

Завідувач кафедри д.т.н., проф. Олег ФЕДОРОВИЧ
(науковий ступінь та вчене звання) (підпис) (ініціали та прізвище)

1. Загальна інформація про викладача

Прохоров Олександр Валерійович, д.т.н., професор. З 2001 з року викладає в університеті, має багаторічний досвід професіональної розробки програмного забезпечення та роботи на керівних посадах в R&D компаніях, є головним розробником та керівником проектів у нафтогазовій промисловості та муніципальному управлінні, виключний досвід роботи із системами штучного інтелекту в управлінні різноманітними процесами, в даний час наступні такі:

- «Системи штучного інтелекту та машинне навчання»;
- «Інтелектуалізація прийняття рішень»;
- «Промислова автоматизація, вбудовані системи реального часу та Інтернету-речей»;
- «Інтегровані автоматизовані системи управління»;
- «Сучасний стан і тенденції розвитку компютерних наук»;
- «Машинне навчання на мові Python».

Напрями наукових досліджень: штучний інтелект, некласичні логіки, машинне навчання, мультиагентні системи, промислова автоматизація, Інтернет речей, імітаційне моделювання, Semantic Web, онтології, хмарні технології, технології AR/VR.

2. Опис навчальної дисципліни

Семестр, в якому викладається дисципліна – 8 семестр.

Обсяг дисципліни:

5 кредитів ЄКТС (150 годин), у тому числі аудиторних – 48 годин, самостійної роботи здобувачів – 102 години.

Форми здобуття освіти

Денна, дистанційна.

Дисципліна – вибіркова.

Види навчальної діяльності – лекції, лабораторні роботи, самостійна робота здобувача.

Види контролю – поточний, модульний та підсумковий (семестровий) контроль (іспит).

Мова викладання – українська.

Необхідні обов'язкові попередні дисципліни (пререквізити) – знання з комп'ютерних наук, інформаційних систем та технологій, комп'ютерної та програмної інженерії, комп'ютерних систем та мереж,

Необхідні обов'язкові супутні дисципліни (кореквізити) – немає.

3. Мета та завдання навчальної дисципліни

Мета

Курс фокусується на формуванні теоретичних знань та практичних навичок в області сучасних методів побудови та аналізу різноманітних моделей машинного навчання для вирішення прикладних задач у різних предметних галузях. Значну увагу приділено вивченню методів, моделей і задач машинного навчання, що базуються на деревах прийняття рішень та нейронних мережах. Разом з тим докладно представлене кейсорієнтовне розроблення моделей в Python за допомогою бібліотек машинного навчання Keras і TensorFlow. Пропонується поєднання теоретичних принципів машинного навчання з практичним підходом до написання коду для повного розуміння теорії машинного навчання і реалізація за допомогою Python та хмарних платформ.

Завдання

Набуття вмінь та навичок розв'язання задач з використанням методів та засобів машинного навчання на мові Python і створення інтелектуальних систем штучного інтелекту для прийняття рішень у різних предметних галузях.

Після опанування дисципліни здобувач набуде наступні **компетентності**:

– здатність роботи з даними в Python;

– розуміння основних алгоритмів машинного навчання, зокрема нейронні мережі, та практичне їх використання для задач класифікації та регресії;

– здатність аналізувати дані та будувати моделі машинного навчання за допомогою Python із застосуванням бібліотек Keras і TensorFlow;

– здатність робити якісні й інтерактивні візуалізації даних;

– здатність застосовувати отримані знання в практичних ситуаціях, застосовувати вивчені інструменти для Data Science.

Очікується, що після опанування дисципліни здобувач будуть досягнуті наступні **результати навчання** і він буде:

– знати основи роботи з даними на мові Python;

– мати уявлення про сучасний стан, досягнення та тренди у галузі штучного інтелекту та машинного навчання;

– знати процеси, які входять до сфери вирішення задач класифікації та регресії за допомогою методів машинного навчання;

– вміти будувати моделі машинного навчання за допомогою Python із застосуванням бібліотек Keras і TensorFlow;

– вміти робити якісні й інтерактивні візуалізації даних для вирішення аналітичних завдань;

– вільно спілкуватися з професійних питань усно і письмово державною мовою, включати знання спеціальної термінології та навички міжособистісного спілкування.

4. Зміст навчальної дисципліни

Модуль 1.

Змістовний модуль 1.

Тема 1. Вступ до навчальної дисципліни «Машинне навчання на мові Python», основні поняття

- *Форма занять: лекція, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Роль і значення штучного інтелекту та машинного навчання. Місце дисципліни в навчальному плані. Список рекомендованої літератури. Цифрова трансформація світу. Ускладнення систем та процесів. Делегування. Ключові сценарії та цілі використання технологій розумних просторів.

- *Обсяг самостійної роботи здобувачів: 4-6 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень.

Тема 2. Штучний інтелект: стан, проблеми та напрямки.

- *Форма занять: лекція, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Деякі напрямки AI. Бум AI та його передумови. Революція у штучному інтелекті. Машинний інтелект сьогодні. Машинний інтелект завтра. Проблема

даних. Ризики, пов'язані з людським фактором. Ризик у розвитку людської цивілізації. Основні напрямки вирішення завдань із застосуванням AI. Напрями досліджень AI. Технології AI, які змінюють світ. AI для проблем планетарного масштабу.

- *Обсяг самостійної роботи здобувачів: 8-10 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень.

Тема 3. Основні поняття та завдання Data Mining і Machine Learning.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Лабораторна робота: «Робота с даними у Machine Learning, обробка та візуалізація на Python».*
- *Лабораторна робота: «Дерева рішень за допомогою Scikit-Learn. Класифікація та регресія».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Інтелектуальний аналіз даних (Data Mining). Завдання Data Mining. Особливості систем Data Mining. Машинне навчання. Що потрібне для машинного навчання? Методи машинного навчання. Застосування машинного навчання. Дерева рішень. Переваги та недоліки дерев рішень. Алгоритми побудови дерев рішень: ID3, C4.5, CART. Залежність якості від глибини дерева. Композиція дерев. Bootstrap. Bagging. Випадковий ліс. Бустінг дерев. Дерева рішень для класифікації. Дерева рішень для регресії. Дерева рішень за допомогою Scikit-Learn

- *Обсяг самостійної роботи здобувачів: 14-16 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Тема 4. Нейромережі в задачах прийняття рішень.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Лабораторна робота: «Вступ до штучних нейронних мереж і глибокого навчання. Класифікація та регресія».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Деякі галузі застосування. Штучний нейрон. Хронологія нейромереж. Штучна нейронна мережа. Глибоке навчання. Навчання нейромережі. Нейромережі прямого поширення. Загальні функції активації. Різноманіття архітектур. Важливі тренди. Згорткові нейромережі CNN. Склад CNN. Операція

згортки. Операція pooling. Inception, ResNet. Різновиди згорткових мереж. Fully-convolutional networks. Deconvolution networks. Рекурентні нейромережі RNN. Алгоритм Backpropagation through time. Довга короткочасна пам'ять LSTM. Різновиди рекурентних мереж. Мультимодальне навчання. Знайомство з TensorFlow для глибокого навчання. API моделі TensorFlow. Короткий огляд екосистеми TensorFlow. Інструменти. Моделі та набори даних. Розгортання моделі. Основи тензорів. Створення тензора за допомогою tf.constant(). Створення тензора за допомогою tf.variable(). Створення тензора з існуючих функцій. Вибір даних у тензорі. Виконання операцій над тензорами. Маніпулювання формою тензора. Введення в регресію з TensorFlow: збір даних; перегляд даних; підготовка даних для моделі; створення, компіляція та навчання моделі; оцінка моделі; удосконалення моделі. Введення в класифікацію за допомогою TensorFlow. Двійкова класифікація. Перехід від бінарного класифікатора до мультикласового класифікатора: 10 модних класифікаторів. Класифікація за допомогою TensorFlow:збір даних; перегляд даних; підготовка даних для моделі; створення, компіляція та навчання моделі; оцінка моделі; керування навчанням за допомогою зворотних викликів; удосконалення моделі. Використання TensorBoard для візуалізації моделі

- *Обсяг самостійної роботи здобувачів: 14-16 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Модульний контроль 1

- *Форма занять: написання модульної роботи в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 2 години*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер.*
- *Обсяг самостійної роботи здобувачів – за необхідністю (підготовка до модульного контролю).*

Змістовний модуль 2.

Тема 5. Особливості вирішення завдань комп'ютерного зору.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Лабораторна робота: «Вирішення завдань комп'ютерного зору за допомогою нейронних мереж»*

- *Лабораторна робота: «Розпізнавання об'єктів та обличь з web-камери за допомогою попередньо навченої моделі».*
- *Лабораторна робота: «Розпізнавання об'єктів та обличь на робототехнічному макеті».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Вступ до комп'ютерного зору та згорткових нейронних мереж (CNN). Що таке згорткові нейронні мережі? Типова архітектура згорткових нейронних мереж. Кодування ConvNets: класифікація зображень: завантаження даних; перегляд даних; підготовка даних для моделі; створення, компіляція та навчання моделі; візуалізація результатів моделі; оцінка моделі; покращення моделі; збереження та завантаження. CNN для набору даних реального світу та доповнення зображень. Доповнення зображень за допомогою ImageDataGenerator. Доповнення зображення за допомогою шарів доповнення зображення Keras. Архітектури CNN і трансферне навчання. Швидка класифікація зображень на попередньо підготовленій моделі. Трансферне навчання та тонке налаштування на практиці. Класифікація зображень і навчання передачі за допомогою TensorFlow Hub.

- *Обсяг самостійної роботи здобувачів: 14-16 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Тема 6. Особливості вирішення завдань обробки тексту.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Лабораторна робота: «Вирішення завдань обробки природного тексту».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Вступ до рекурентних нейронних мереж (RNN). RNN на практиці, аналіз настроїв у фільмах: отримання даних; підготовка даних; побудова, компіляція та навчання моделі RNN; візуалізація результатів моделі. Вступ до довгострокових спогадів. LSTM на практиці, класифікація новин: отримання даних; підготовка даних; побудова та навчання моделі RNN; візуалізація результатів моделі; використання стекованих LSTM; використання рекуррентного модуля (GRU). Використання згорткових нейронних мереж для класифікації текстів.

- *Обсяг самостійної роботи здобувачів: 14-16 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Тема 7. Кластерізація даних на основі нейромереж.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 2 години.*
- *Лабораторна робота: «Навчання без вчителя. Кластерізація».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Що таке Unsupervised Learning? Набір даних для неконтрольованого навчання. Алгоритми неконтрольованого навчання. Алгоритм K-means. Застосування Unsupervised Learning.

- *Обсяг самостійної роботи здобувачів: 14-16 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Тема 8. Навчання з підкріпленням.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 2 години.*
- *Лабораторна робота: «Навчання з підкріпленням».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер, Google Colab.*

Вступ до Reinforcement Learning. Вступ да OpenAI gym. Проста жорстко закодована політика. Політики нейронних мереж. Градієнти політики. Марковський процес прийняття рішень. Q-Value. Q-Learning. Deep Q-Network.

- *Обсяг самостійної роботи здобувачів: 6-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вивчення термінів та визначень. Оформлення лабораторної роботи та підготовка до її здачі.

Модульний контроль 2

- *Форма занять: написання модульної роботи в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 2 години*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, браузер.*
- *Обсяг самостійної роботи здобувачів – за необхідністю (підготовка до модульного контролю).*

5. Індивідуальні завдання

Не передбачено навчальним планом

6. Методи навчання

Словесні, наочні, практичні. Проведення аудиторних занять, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами.

7. Методи контролю

Поточний контроль (теоретичне опитування й розв’язання практичних завдань), модульний контроль (тестування за розділами курсу) та підсумковий (семестровий) контроль (іспит).

8. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Виконання і захист лабораторних робіт	0…7	3	0…21
Модульний контроль	0…17	1	0…17
Змістовний модуль 2			
Виконання і захист лабораторних робіт	0…7	6	0…42
Модульний контроль	0…20	1	0…20
За семестр			**0…100**

Прийнята шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка для екзамену, *курсового проекту* (роботи), практики
90-100	відмінно
75-89	добре
60-74	задовільно
0-59	незадовільно з можливістю повторного складання

Іспит проводиться у вигляді письмової відповіді на білет, який складається з трьох питань: двох теоретичних (по 30 балів) та одного практичного (40 балів). Під час складання семестрового іспиту здобувач має можливість отримати максимум 100 балів.

Критерії оцінювання роботи здобувача протягом семестру

Задовільно (60-74) – мати мінімум знань і умінь для забезпечення програмних результатів навчання. Відпрацювати та захистити всі лабораторні роботи.

Добре (75-89) – знати основні теми дисципліни. Достатньо знати основні дії та процеси, вміти користуватися бібліотеками Python для роботи з даними та фреймворками машинного навчання. Відпрацювати та захистити всі лабораторні роботи.

Відмінно (90-100) – мати знання, що дозволять самостійно, вільно та обґрунтовано відповідати на будь які питання. Відпрацювати та захистити всі лабораторні роботи.

9. Політика навчального курсу

Відпрацювання пропущених занять відбувається відповідно до розкладу консультацій, за попереднім погодженням з викладачем. Питання, що стосуються академічної доброчесності, розглядає викладач або за процедурою, визначеною у Положенні про академічну доброчесність.

10. Інформаційні ресурси

1. Mining of Massive Datasets Stanford University. //The Stanford InfoLab. URL: http://i.stanford.edu/~ullman/mmds/book.pdf
2. UCI Knowledge Discovery in Databases Archive // http://kdd.ics.uci.edu/
3. Фреймворк TensorFlow // https://www.tensorflow.org/
4. Kaggle: Your Machine Learning and Data Science Community // https://www.kaggle.com
5. Датасети // https://www.kaggle.com/datasets
6. Датасети // https://datasetsearch.research.google.com/
7. Датасети // https://www.tensorflow.org/datasets/catalog/overview#all_datasets
8. Датасети // https://console.cloud.google.com/marketplace/browse

11. Рекомендована література

Базова

1. Жерон О. Прикладне машинне навчання за допомогою Scikit-Learn, Keras і TensorFlow концепції, інструменти і техніки для створення інтелектуальних систем 2-е вид., 2020 – 1000 с.
2. Alpaydin E. Introduction to Machine Learning // MIT Press, 2020 – 321 с.
3. Russell S.J., Norvig P. Artificial Intelligence: A Modern Approach // Prentice Hall, 4th US ed., 2020 – 1166 p.

Допоміжна

1. 1. Aggarwal C.C. Linear Algebra and Optimization for Machine Learning // Springer, 2020 – 517 с.
2. Trappenberg T., Fundamentals of Machine Learning // Oxford University Press, 2020 – 272 с.
3. Jung A. Machine Learning: The Basics (Machine Learning: Foundations, Methodologies, and Applications) // Springer, 2020 – 229 с.