

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО”  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ**

**ЗАТВЕРДЖУЮ**

Заступник директора

з науково-педагогічної роботи

ІПСА КПІ ім. Ігоря Сікорського

Романенко В.Д.

«25» червня 2021 р.

## **ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ**

### **ПРОГРАМА навчальної дисципліни**

**першого (бакалаврського) рівня вищої освіти**

**ступеня «Бакалавр»**

**спеціальність 122 «Комп'ютерні науки та інформаційні технології»**

**освітньої програми «Системи і методи штучного інтелекту»,**

**спеціальність 124 «Системний аналіз» освітніх програм «Системний аналіз і управління», «Системний аналіз фінансового ринку»**

Ухвалено методичною комісією

ІПСА КПІ ім. Ігоря Сікорського

Протокол від 25.06. 2021 р. № 9

Голова методичної комісії

Бідюк П.І.

«25» червня 2021 р.

РОЗРОБНИК ПРОГРАМИ:

Доцент, д.т.н., доц. Недашківська Надія Іванівна

\_\_\_\_\_  
(підпис)

Програму затверджено на засіданні кафедри  
математичних методів системного аналізу  
Протокол від «24» червня 2021 року № 9.

В.о. завідувача кафедри

\_\_\_\_\_  
(підпис) доц. Тимощук О.Л.

«24» червня 2021 р.

## Вступ

Програму навчальної дисципліни «Інтелектуальний аналіз даних» складено відповідно до освітніх програм «Системи і методи штучного інтелекту» першого (бакалаврського) рівня вищої освіти спеціальності 122 «Комп'ютерні науки та інформаційні технології».

Навчальна дисципліна «Інтелектуальний аналіз даних» належить до циклу професійної та практичної підготовки бакалаврів.

Статус навчальної дисципліни: **вибіркова**.

Обсяг навчальної дисципліни: 4,5 кредитів ЄКТС.

Міждисциплінарні зв'язки: Знання, набуті при вивченні цієї дисципліни, використовуються при опануванні дисциплін «Інтелектуальний аналіз великих сховищ даних», в дипломному проектуванні, у практичній самостійній роботі випускника в галузі інтелектуального аналізу даних під час аналізу великих і надвеликих баз даних та масивів тексту, при побудові прогнозів на основі статистичних даних та оцінок експертів, при розробці корпоративних інформаційно-аналітичних систем в державних і приватних управлінських структурах.

При вивченні дисципліни використовуються знання дисциплін «Теорія ймовірностей та математична статистика», «Математичний аналіз», «Дискретна математика (розділ «Теорія графів»)), «Бази даних та інформаційні системи», «Методи оптимізації».

## 1. Мета та завдання навчальної дисципліни

### 1.1. Мета навчальної дисципліни.

Метою навчальної дисципліни є формування у студентів здатностей:

- аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень,
- пошуку шаблонів у великих і надвеликих базах даних та у масивах тексту,
- побудови прогнозів з використанням сучасних методів і алгоритмів машинного навчання та інформаційних технологій,
- застосування сучасних методів машинного навчання,
- використання програмного забезпечення для машинного навчання в практичній роботі.

### 1.2. Основні завдання навчальної дисципліни.

Згідно з вимогами освітньо-професійної програми студенти після засвоєння навчальної дисципліни мають продемонструвати такі результати навчання:

#### **знання:**

сучасних методів і алгоритмів теорії машинного навчання, методики застосування цих методів, алгоритмів для побудови прогнозів на основі статистичних даних та оцінок експертів, системного вирішення практичних задач аналізу і пошуку шаблонів у великих і надвеликих базах даних, методики розв'язання практичних задач класифікації, кластеризації, пошуку асоціативних правил, прогнозування з використанням методів машинного навчання та інформаційних технологій, знання інструментів машинного навчання.

**уміння:**

застосовувати сучасні методи і алгоритми теорії машинного навчання з метою підтримки прийняття рішень, будувати прогнози на основі статистичних даних та оцінок експертів, розв'язувати практичні задачі класифікації, кластеризації, пошуку асоціативних правил, прогнозування з використанням методів машинного навчання та інформаційних технологій, системно вирішувати практичні задачі аналізу і пошуку шаблонів у великих і надвеликих базах даних, використовувати програмне забезпечення для машинного навчання в практичній роботі.

**досвід:**

теоретичний та практичний досвід аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень, побудови прогнозів, використання програмного забезпечення для машинного навчання в практичній роботі.

## **2. Зміст навчальної дисципліни**

### **Розділ 1. Вступ до інтелектуального аналізу даних (ІАД) та машинного навчання**

Тема 1. Загальні відомості про ІАД. Задачі ІАД та МН: класифікація, регресія, машинний переклад, пошук асоціативних правил, структурний вивід, синтез і вибірка, оцінка функції ймовірності або функції щільності ймовірності. Огляд методів ІАД та машинного навчання. Досвід в задачах МН.

Тема 2. Поняття перенавчання (overfitting) моделі МН. Компроміс між систематичною помилкою і дисперсією. Крива перевірки. Оцінювання якості алгоритмів. Порівняння алгоритмів та вибір гіперпараметрів.

Тема 3. Теорема Байеса. Максимальна апостеріорна гіпотеза. Метод максимальної правдоподібності.

Тема 4. Задача лінійної регресії. Міра якості в задачі лінійної регресії. Функція помилки в задачі регресії та її обґрунтування за допомогою байесівського підходу. Регуляризація зі зниженням ваги та її обґрунтування. Гребнева регресія.

Тема 5. Практичне застосування ІАД.

### **Розділ 2. Методи та алгоритми класифікації**

Тема 1. Побудова математичних функцій класифікації. Метод опорних векторів: лінійний та нелінійний випадки. Обґрунтування методу опорних векторів. Вибір параметру розмиття смуги. Функція ядра, властивості, типи ядер. Переваги і недоліки методу опорних векторів. Реалізація методу в scikit-learn python.

Тема 2. Дерева рішень. Алгоритм розбиття, його властивості. Критерії вибору змінної розбиття: ентропійний, Джині. Алгоритми ID3, C4.5 вибору змінної розбиття, їх властивості. Міри ефективності дерев рішень. Проблема зупинки побудови дерева.

Тема 3. Дерева рішень в scikit-learn python. Алгоритм розбиття CART для класифікації та регресії. Регуляризація дерев рішень. Приклади. Переваги і недоліки дерев рішень. Алгоритм покриття. 1-R алгоритм.

Тема 4. Байесівські методи класифікації. Оптимальний байесівський класифікатор. Оцінювання апіорних імовірностей та функцій правдоподібності за вибіркою. Наївний байесівський класифікатор.

Тема 5. Байєсівські методи класифікації. Задача розділу суміші. Алгоритм Expectation-Maximization (ЕМ) та його модифікації.

Тема 6. Оцінювання ефективності алгоритмів класифікації.

Тема 7. Основи теорії штучних нейронних мереж. Класичний і сучасні перцептрони. Реалізація багатопарового перцептрону в scikit-learn python.

Тема 8. Моделі і алгоритми навчання нейронних мереж: на основі коригування помилок, з використанням пам'яті, на основі правила Хебба, моделі конкурентного навчання, модель Больцмана.

Тема 9. Метод стохастичного градієнтного спуску.

Тема 10. Системи з нечіткою логікою. Нечітко-нейронні системи. Налаштування параметрів нечітко-нейронних систем.

### **Розділ 3. Методи та алгоритми кластеризації**

Тема 1. Ієрархічна кластеризація: агломеративний алгоритм найближчого сусіда, дівізімний алгоритм. Методи сусідства. Поняття дендрограми.

Тема 2. Алгоритм k-середніх, нечітких k-середніх, g-середніх та EM.

Тема 3. Методи кластеризації на основі теорії графів. Алгоритми побудови мінімального покриваючого дерева. Алгоритм Борувки.

Тема 4. Алгоритм Форел та його модифікації.

Тема 5. Метод самоорганізуючих карт Кохонена. Інтерпретація карт.

Тема 6. Аналіз результатів кластеризації. Розв'язання практичних задач.

### **Розділ 4. Ансамблі моделей ІАД**

Тема 1. Види ансамблів. Композиції дерев рішень та випадковий ліс. Беггінг.

Тема 2. Композиції простих алгоритмів. Поняття бустингу. Алгоритм AdaBoost. Обґрунтування бустингу.

Тема 3. Градієнтний бустинг.

Тема 4. Методи розрахунку коефіцієнтів відносних важливостей (ваг) моделей в ансамблі.

### **Розділ 5. Методи та алгоритми побудови асоціативних правил. Секвенційний аналіз**

Тема 1. Загальні відомості. Показники корисності асоціативних правил.

Тема 2. Алгоритми Apriori, Eclat та FP-росту.

Тема 3. Шаблони послідовностей. Алгоритм AprioriAll.

Тема 4. Розв'язання практичних задач пошуку асоціативних правил.

## **3. Заплановані види навчальної діяльності та методи навчання**

Заплановані види навчальних занять: лекції та лабораторні роботи.

Мета лабораторних робіт – закріплення теоретичних положень навчальної дисципліни, отримання практичних навичок використання методів інтелектуального аналізу даних і машинного навчання з метою підтримки прийняття рішень. В результаті виконання робіт студенти повинні вміти розв'язувати практичні задачі класифікації, кластеризації, пошуку асоціативних правил з використанням сучасного

програмного забезпечення Python, системно вирішувати практичні задачі аналізу і пошуку шаблонів у великих і надвеликих базах даних, будувати прогнози на основі статистичних даних та оцінок експертів, використовувати сучасне програмне забезпечення для інтелектуального аналізу даних та машинного навчання в практичній роботі.

№ з/п	Назва лабораторної роботи
1	Отримання навичок роботи в середовищі Python
2	Класифікація та регресія за допомогою бібліотеки Scikit-Learn Python
3	Кластеризація засобами бібліотеки Scikit-Learn Python
4	Побудова ансамблів моделей класифікації та регресії засобами бібліотеки Scikit-Learn Python
5	Побудова асоціативних правил за алгоритмами Apriori та FP-росту

Використовуються електронні презентації для лекційних занять.

#### **4. Оцінювання результатів навчання**

Семестрова атестація проводиться у виді заліку. Для оцінювання результатів навчання застосовується 100-бальна рейтингова система і університетська шкала оцінювання.

Для діагностики успішності навчання використовуються екзаменаційні білети з теоретичною частиною та практичними завданнями.

#### **5. Рекомендована література**

1. Н.І. Недашківська. Конспект лекцій з кредитного модуля «Інтелектуальний аналіз даних», 2020.
2. Н.І. Недашківська. Методичні вказівки до виконання лабораторних робіт з кредитного модуля «Інтелектуальний аналіз даних», 2020.
3. Гудфеллоу Я., Бенджио И., Курвилль А.. Глубокое обучение. пер. с англ. А. А. Слинкина. 2-е изд., испр. М.: ДМК Пресс, 2018, 652 с.
4. Николенко С., Кадурын А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018, 480 с.
5. Себастьян Рашка, Вахид Мирджалили. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2, 3-е изд.: Пер. с англ. - СПб. : ООО "Диалектика", 2020. - 848 с.
6. Плас Дж.В. Python для сложных задач. Наука о данных и машинное обучение. СПб.: Питер, 2018. 576 с.
7. Уэс Маккинли. Python и анализ данных / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 482 с.

8. Матеріали щодо інтелектуального аналізу даних на сайті <http://www.machinelearning.ru> - Режим доступу: <http://www.machinelearning.ru>. 2020.
9. Матеріали щодо машинного навчання на сайті <http://www.machinelearning.ru> - Режим доступу: <http://www.machinelearning.ru>. 2020.
10. Матеріали щодо прикладних систем аналізу даних на сайті <http://www.machinelearning.ru> - Режим доступу: <http://www.machinelearning.ru>. 2020.
11. Матеріали щодо інтелектуального аналізу даних на сайті <https://basegroup.ru> - Режим доступу: <https://basegroup.ru>. 2020.
12. Матеріали щодо машинного навчання на сайті <https://basegroup.ru> - Режим доступу: <https://basegroup.ru>. 2020.
13. Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение. СПб.: Питер, 2017. 336 с.
14. Мюллер А., Гидо С. Введение в машинное обучение с помощью Python. М.: O'Reilly Media, 2017. 392 с.
15. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2017. 336 с.
16. Замятин А.В. Введение в интеллектуальный анализ данных. Учебное пособие. Томск. 2016. 120 с.
17. А.А.Барсегян, М.С.Куприянов, И.И.Холод, М.Д.Тесс, С.И.Елизаров. Анализ данных и процессов: учеб.пособие. 3-е изд., перераб.и доп. СПб.:БХВ-Петербург, 2009. 512 с.
18. Чубукова И.А. Data Mining. БИНОМ. Лаборатория знаний, Интернет-университет информационных технологий. ИНТУИТ.ру. 2008.
19. Scikit-Learn Documentation. <https://scikit-learn.org/>.