

Галузь знань: 01 Освіта / Педагогіка
Ступінь вищої освіти: Бакалавр
Спеціальність: 015 Професійна освіта (Комп'ютерні технології)
Освітня програма: Професійна освіта. Комп'ютерні технології
Кафедра інформатики і кібернетики

Навчальна дисципліна
«Аналіз даних та машинне навчання»

Семестр – 4

Форма контролю – залік

Кількість кредитів ЄКТС – 5

I. Основна мета засвоєння курсу набуття знань зі збору та аналізу даних, вибору оптимального рішення і оцінки його якості у цілях машинного навчання, ознайомлення із сучасними методами класифікації і регресії, пошуку структури в даних, проведення експериментів, побудови висновків; напрацювання вмінь користуватися сучасними аналітичними інструментами і адаптувати їх під особливості конкретних завдань, працювати з великим масивом даних.

II. Місце навчальної дисципліни в освітній програмі

Дисципліна «Аналіз даних та машинне навчання» дозволяє набуту студентам додаткових фахових компетентностей при опануванні циклу професійної підготовки.

III. Завдання дисципліни: ознайомити студентів із засобами систематизації, обробки і використання статистичної інформації для виявлення ознаки або ознак певної сукупності елементів; основними методами регресійного та кореляційного аналізу; засобами побудови емпіричних залежностей та емпіричних формул; оволодіння методами вирішення типових завдань машинного навчання і аналізу даних.

IV. Основні знання та уміння, яких набуває студент після опанування даної дисципліни

Основні знання:

- існуючих технологій підготовки даних до аналізу;
- основних методів пошуку в даних внутрішніх закономірностей, взаємозв'язків, тенденцій;
- тенденцій розробки універсальних програмних засобів і аналітичних платформ, призначених для аналітичної обробки даних, побудови прогнозів і аналітичних сценаріїв;
- основних методів консолідації, трансформації, візуалізації, оцінки якості, очищення та попередньої обробки даних;
- принципи побудови і структурну організацію сховищ даних;
- алгоритми пошуку асоціативних правил і кластерного аналізу;
- статистичні та машинні методи класифікації і регресії;
- можливості універсальних програмних засобів і аналітичних платформ, що застосовуються для аналізу даних;
- можливості використання засобів мов програмування до аналізу даних;
- проблемні питання впровадження аналітичних програмних продуктів і технологій в професійну діяльність організацій і установ;

Основні вміння:

- оволодіння практичними вміннями і навичками реалізації технологій аналітичної обробки даних, формування і перевірки гіпотез про їх природу і структуру, варіювання застосовуваними моделями;
- формування умінь і навичок застосування універсальних програмних пакетів і аналітичних платформ для аналізу даних.

- практично застосовувати методи консолідації, трансформації, візуалізації, оцінки якості, очищення та попередньої обробки даних для якісної підготовки даних до аналізу;
- застосовувати технології аналізу електронних масивів даних для вирішення конкретних практичних проблем;
- скористатися наявними можливостями універсальних програмних засобів, аналітичних платформ, засобів мов програмування для аналітичної обробки даних, побудови прогнозів і аналітичних сценаріїв.

V. Короткий зміст дисципліни

Тема 1. Поняття про аналіз даних. Визначення інтелектуального аналізу даних, історія та передумови виникнення. Аналіз даних і витяг знань з даних. Місце і роль аналізу даних в процесі прийняття рішень. Основні завдання аналізу даних, види моделей. Аналіз даних і статистичний аналіз. Сховища даних. OLAP. Аналіз даних як самостійна галузь індустрії програмного забезпечення.

Тема 2. Базова фундаментальна математика для аналізу даних. Функції та їх властивості. Межа і похідна. Екстремум. Похідна та її застосування. Лінійна алгебра. Векторні простори. Матриці. Оптимізація і матричні розкладання. Випадковість. Оптимізація: глобальна оптимізація і оптимізація негладкою функції.

Тема 3. Машинне навчання та лінійні моделі. Поняття про машинне навчання. Основні терміни у машинному навчанні. Основні способів вирішення завдань навчання на розмічених даних. Задачі регресії і класифікації. Прогнози за допомогою лінійних моделей. Проблема перенавчання. Вирішальні дерева і композиції алгоритмів

Тема 4. Пошук структури в даних. Структура даних, взаємозв'язки і закономірності. Методи зниження розмірності, засновані на матричних розкладаннях. Завдання тематичного моделювання, «навчання без учителя». Алгоритми кластеризації даних. Побудова матричних розкладань, пошук аномалій. Візуалізація багатовимірних даних.

Тема 5. Побудова висновків за даними. Збір даних. Шум. Ймовірність. Методи статистики. Перетворення даних у висновки. Організація експериментів, А / В-тестування, універсальні методи оцінки параметрів і перевірки гіпотез, кореляції та причинно-наслідкові зв'язки.

Тема 6. Прикладні завдання аналізу даних. Методи машинного навчання. Аналіз тексту і інформаційний пошук, колаборативна фільтрація і рекомендаційні системи. Бізнес-аналітика. Прогнозування часових рядів. Витягування ознак з різнорідних даних. Формалізація постановки задачі машинного навчання. Якість моделі. Плюси і мінуси алгоритмів машинного навчання.

VI. Назва кафедри та викладацький склад, який буде забезпечувати викладання курсу

Кафедра інформатики і кібернетики факультету інформатики, математики, економіки.

VII. Обсяги навчального навантаження та терміни викладання курсу

На вивчення дисципліни відводиться 150 годин (5 кредитів ЄКТС).

Дисципліна викладається у 4-му семестрі.

VIII. Основні інформаційні джерела до вивчення дисципліни

1. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
2. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с.
3. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.

IX. Система оцінювання:

Поточний контроль: оцінювання виконання завдань на лабораторних заняттях, оцінювання 2-х модульних тестових контрольних робіт, виконання творчих індивідуальних завдань.

Підсумковий контроль: залік у 4-му семестрі.