

| | |
|-------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Код та назва дисципліни українською мовою/ Назва дисципліни англійською мовою | 2-ф11-6 Дослідницький аналіз даних / Exploratory Data Analysis (EDA) |
| Рекомендується для галузі знань | A – Освіта, E – Природничі науки, математика та статистика, F – Інформаційні технології, G – Інженерія, виробництво та будівництво |
| Кафедра | Статистики й теорії ймовірностей |
| П.І.П. НПП | Бондаренко Яна Сергіївна |
| Рівень ВО | Перший (бакалаврський) |
| КУРС, семестр (в якому буде викладатись) | 1 курс, парний семестр |
| Мова викладання | Українська |
| Пререквізити (передумови вивчення дисципліни) | «Вища математика («Теорія ймовірностей та математична статистика») або «Теорія ймовірностей та математична статистика» або «Теорія ймовірностей» або «Теорія ймовірностей, ймовірнісні процеси та математична статистика» |
| Чому це цікаво/треба вивчати | Дослідницький аналіз даних (Exploratory Data Analysis, EDA) – це процес дослідження даних з метою виявлення важливих характеристик, закономірностей та аномалій перед глибшим аналізом чи моделюванням. EDA допомагає зрозуміти, як дані розподілені, чи містять вони пропущені значення, аномалії або кореляції між змінними. |
| Перелік тем з дисципліни | <p>Розділ 1. Збір даних</p> <p>Тема 1. Джерела даних: відкриті набори, веб-скрапінг, API, бази даних</p> <p>Тема 2. Структуровані та неструктуровані дані</p> <p>Тема 3. Етичні аспекти збору даних</p> <p>Розділ 2. Очищення та попередня обробка даних</p> <p>Тема 1. Виявлення та заповнення пропущених значень</p> <p>Тема 2. Робота з аномаліями та викидами</p> <p>Тема 3. Перетворення форматів даних</p> <p>Тема 4. Кодування категоріальних змінних</p> <p>Розділ 3. Описова статистика</p> <p>Тема 1. Міри центральної тенденції: середнє, медіана, мода</p> <p>Тема 2. Міри розмаху та варіації: мінімум, максимум, дисперсія, стандартне відхилення</p> <p>Тема 3. Розподіл значень: квартилі та IQR</p> <p>Розділ 4. Одновимірний аналіз</p> <p>Тема 1. Розподіл змінних</p> <p>Тема 2. Виявлення закономірностей у змінних</p> <p>Тема 3. Візуалізація розподілів змінних</p> <p>Розділ 5. Двовимірний аналіз</p> <p>Тема 1. Кореляційний аналіз</p> <p>Тема 2. Аналіз залежностей між змінними</p> <p>Тема 3. Візуалізація взаємозв'язків</p> <p>Розділ 6. Багатовимірний аналіз</p> <p>Тема 1. Зменшення розмірності даних: PCA</p> <p>Тема 2. Кластеризація: K-Means, DBSCAN</p> <p>Тема 3. Виявлення прихованих структур у даних</p> <p>Розділ 7. Генерація ознак</p> <p>Тема 1. Створення нових змінних. Комбінація змінних, логарифмічні перетворення</p> <p>Тема 2. Автоматичний вибір найкращих ознак. Методи SelectKBest, Recursive Feature Elimination</p> <p>Тема 3. Збалансування категоріальних класів. Oversampling, undersampling.</p> |

| | |
|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <p>Розділ 8. Візуалізація даних Тема 1. Графічне представлення даних Тема 2. Використання бібліотек Matplotlib, Seaborn, Plotly Тема 3. Інтерактивні графіки та дашборди</p> |
| Як можна користуватися набутими знаннями і уміннями | Опанування дослідницького аналізу даних відкриває широкі кар'єрні перспективи у різних галузях, зокрема у сфері аналітики даних (Data Analyst, Business Analyst, Machine Learning Engineer, Financial Analyst, Marketing Data Analyst, Data Scientist). Володіння навичками збору, очищення, аналізу та візуалізації даних є важливим і затребуваним у бізнесі, фінансах, медицині, маркетингу та ІТ. |
| Очікувані результати навчання | Після опанування дисципліни студенти зможуть самостійно проводити дослідницький аналіз даних, використовувати методи візуалізації, працювати з великими масивами інформації та отримувати корисні нові знання для ухвалення рішень. |
| Інформаційне забезпечення | Відкриті набори даних Kaggle та UCI Machine Learning Repository, основна мова для реалізації EDA – Python Робоча програма навчальної дисципліни, інтернет-ресурси |
| Види навчальних занять | Лекції, практичні заняття |
| Вид семестрового контролю | Диференційований залік |
| Максимальна кількість здобувачів на семестр | без обмежень |

Декан факультету _____ *Олександр ХАМІНІЧ*