

Чернівецький національний університет імені Юрія Федьковича

Факультет математики та інформатики

Кафедра математичного аналізу

СИЛАБУС

навчальної дисципліни

«Python для Data Science»

Рівень вищої освіти

перший (бакалаврський)

Мова навчання

українська

Розробники: доц. кафедри математичного аналізу д. ф.-м. н. Нестеренко В.В.

Профайл викладача <http://math.chnu.edu.ua/>

Контактний тел. +380 50 283 0718

E-mail: v.nesterenko@chnu.edu.ua

Опис курсу

Як зрозуміти, чи ви достатньо сильні для проходження цього курсу? Просто дайте собі відповіді на такі питання: чи цікавить вас робота з даними (таблицями даних), маніпуляції з ними та їх візуалізація? Чи пам'ятаєте ви зі школи/університету/роботи суть понять «медіана», «мода», «розподіл ймовірностей»? Якщо відповіді позитивні, то у вас є хороші шанси на успішне проходження курсу.

Цей курс буде корисним:

- Студентам, які хочуть розпочати кар'єру у Data science;
- Фахівцям з Data science, які ще не освоїли можливості роботи з даними у Python;
- Фахівцям, які працюють з даними (аналітикам, бізнес-аналітикам, дослідникам), які хочуть освоїти нові гнучкі інструменти для роботи з даними.

Після курсу ви зможете:

- Розуміти основні алгоритми машинного навчання та практично їх використовувати для задач класифікації та регресії
- Аналізувати дані та будувати моделі машинного навчання за допомогою мови Python
- Робити якісні та інтерактивні візуалізації даних

Загальна інформація

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	-	-	3	90	30	15	-	-	45	-	залік

Програма курсу:

Модуль 1. Введення в машинне навчання та Data Science

1. Визначення машинного навчання та науки про дані, їх завдання
2. Інструментарій для аналізу даних: мови програмування Python
3. Типи машинного навчання
4. Складові частини моделі машинного навчання
5. Базові концепти машинного навчання

Модуль 2. Основи мови програмування Python для Data science

1. IDE для зручної роботи: Jupyter, Spyder
2. Семантика мови
3. Типи даних
4. Структури даних
5. Оператори контролю виконання (розгалуження та циклів)
6. Векторизація в Python (numpy)

Модуль 3. Функціональне програмування в Python

1. Comprehension, Generators, Iterators в Python
2. Функції у Python
3. Вбудовані функції Python

Модуль 4. Дослідження та підготовка даних до аналізу

1. Numpy
2. Pandas

Модуль 5. Візуалізація даних у Python

1. Matplotlib
2. Seaborn

Модуль 6. Основні поняття статистичного аналізу

1. Scipy.stats
2. Описові статистики
3. Залежності між змінними

Модуль 7. Типові проблеми у підготовці даних до аналізу

1. Перетворення факторних змінних
2. Боротьба з пробілами в даних
3. Проблема мультиколінеарності
4. Шкалювання даних
5. Вибір інформативних змінних для моделі
6. Типовий алгоритм підготовки даних до моделювання

Модуль 8. Базові регресійні моделі

1. Лінійна регресія
2. Поліноміальна регресія
3. Метод найближчих сусідів
4. Метрики оцінки якості регресії
5. Тюнінг регресійних моделей

Модуль 9. Базові моделі класифікації

1. Логістична регресія
2. Метод опорних векторів
3. Наївний Байєсовський класифікатор

4. Метрики оцінки якості класифікації
5. Тюнінг параметрів моделей класифікації

Модуль 10. Базові моделі кластеризації та зниження розмірності

1. Метод до-середніх
2. Ієрархічний кластерний аналіз
3. DBSCAN
4. Аналіз
5. Кореспонденс-аналіз

Модуль 11. Дерева рішень та ансамблі

1. Базові деревоподібні моделі
2. Random forest
3. Xgboost, lightgbm
4. Тюнінг гіперпараметрів в ансамблевих моделях

Модуль 12. Методи покращення якості моделей машинного навчання

1. Аугментація
2. Feature engineering
3. Cross-validation
4. Боротьба з перенавчанням
5. Боротьба з дисбалансом даних
6. Регулювання моделі

Модуль 13. Введення в архітектуру нейронних мереж. Нейронні мережі для класифікації зображень

1. Keras
2. tensorflow
3. Персептрон
4. CNN

Модуль 14. Прогнозування часових рядів

1. AR, MA, ARMA
2. ARIMA
3. VAR
4. Lstm

Модуль 15. Документація проектів Data Science, Data Science in prod

1. Шляхи організації продакшна для data science-проектів
2. Специфіка документації data science проектів
3. Markdown, Latex

Модуль 16. Практикум шляхом поліпшення якості моделей

1. Презентація та обговорення курсових проектів
2. Підбиття підсумків курсу