

Описание на проект

Практическо въведение в машинното самообучение с Python

Зимен семестър 2025/2026

1. Участници

Име и фамилия:Галя Додова.....

Факултетен номер:45616.....

Програма: Бакалавър

2. Тип на проекта

Тип В – Състезание по регресия

Проектът е базиран на участие в състезание за машинно самообучение с регресионна задача.

3. Название на проекта (предметна област)

Predicting Wild Blueberry Yield Using Regression Models

(Прогнозиране на добива от диви боровинки чрез регресионни модели)

4. Кратко описание на задачата

Целта на проекта е да се изгради модел за **регресия**, който предсказва **добива (yield) на диви боровинки**, използвайки различни **екологични, климатични и биологични характеристики**.

Задачата е част от състезание от серията *Your Goal*, в което участниците разработват и сравняват регресионни модели върху реални данни.

5. Описание на данните

Данните представляват табличен набор от наблюдения, като всяко наблюдение описва условията за растеж на диви боровинки и съответния получен добив.

Характеристики на данните:

- климатични показатели (напр. температури, валежи);
- показатели, свързани с опрашването;
- характеристики на почвата и околната среда;
- целева променлива: **yield (добив)** – непрекъсната числова стойност.

Първоначален анализ на данните (EDA):

В рамките на проекта ще бъде извършен:

- анализ на липсващи стойности;
- статистическо описание на променливите;
- визуализации:
 - хистограми и boxplot-и на основните променливи;
 - корелационна матрица;
 - scatter plot-и между ключови признаци и целевата променлива.

EDA ще помогне за:

- разбиране на разпределенията;
 - откриване на аномалии;
 - избор на подходящи модели.
-

6. Алгоритми и модели

В проекта ще бъдат използвани и сравнени няколко регресионни подхода, включително:

Базови модели:

- Linear Regression
- Ridge и Lasso Regression

По-сложни модели:

- Random Forest Regressor
- Gradient Boosting (XGBoost / LightGBM)
- Neural Networks (MLP)

Оценка на моделите:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)

- Cross-validation

Целта е да се анализира влиянието на различните характеристики и да се постигне възможно най-добра точност при предсказване на добива.

7. Библиотеки и технологии

Проектът ще бъде реализиран изцяло на **Python**, като ще се използват следните библиотеки и инструменти:

- **NumPy** – числени изчисления
 - **Pandas** – обработка на таблични данни
 - **Matplotlib / Seaborn** – визуализация на данни
 - **Scikit-learn** – регресионни модели и оценка
 - **XGBoost / LightGBM** – градиентно буустинг модели
 - **PyTorch / TensorFlow** (по избор) – невронни мрежи
 - **Jupyter Notebook** – разработка и документация
-

8. Очаквани резултати

Очаква се проектът да:

- демонстрира пълен процес по машинно самообучение;
- сравни различни регресионни модели;
- постигне конкурентен резултат в рамките на състезанието;
- даде интерпретация на факторите, влияещи върху добива на диви боровинки.