
PUSULA
TALENT
ACADEMY
BAŞLIYOR!
Pusula ile Geleceğine Yön Ver

Data Science Intern Case Study

Overview:

You will work on a physical medicine & rehabilitation dataset consisting of 2235 observations and 13 features. The goal is to conduct in-depth EDA and make the data ready for potential predictive modeling.

Target Definition: The target variable for this case is **TedaviSuresi**.

You are **not required to build a model**; your task is to make the data **model-ready** around this target (clean, consistent, and analyzable).

Columns of Dataset

HastaNo	Anonymized patient ID.
Yas	Age
Cinsiyet	Gender
KanGrubu	Blood type
Uyruk	Nationality
KronikHastalik	Chronic conditions (comma-separated list)
Bolum	Department/Clinic
Alerji	Allergies (may be single or comma-separated)
Tanilar	Diagnoses
TedaviAdi	Treatment name.
TedaviSuresi	Treatment duration in sessions
UygulamaYerleri	Application sites
UygulamaSuresi	Application duration

Tasks:

1. **Exploratory Data Analysis (EDA):**
    - In this phase, you are expected to thoroughly explore the dataset using Python. Utilize libraries such as Pandas, Matplotlib, and Seaborn to understand the overall structure of the data, identify the types of variables, and detect any anomalies or missing data. Use data visualization techniques (e.g., histograms, scatter plots, heatmaps) to uncover patterns and relationships within the data.

2. **Data Pre-Processing:**
    - Based on the findings from your EDA, clean and preprocess the data to make it ready for modeling. This includes handling missing values, encoding categorical variables, normalizing or standardizing numerical features, and addressing any data quality issues. For example, you can use methods like SimpleImputer or KNNImputer to handle missing data, and OneHotEncoder or LabelEncoder for categorical data.

Optional (Nice to Have):

- **Documentation:**
  - Prepare a document summarizing your findings from the EDA and the steps you took during data preprocessing. If you create this document, it should be included in your GitHub repository along with your code.
- **Pipeline Level Code:**
  - If possible, organize your code at the pipeline level, ensuring that it is modular and reusable.
- **Different Approach:**
  - If possible, you can use an approach other than the examples given.

Submission Requirements:

- **GitHub Repository:**
  - Your code, along with a README.md file, should be uploaded to a GitHub repository. The README should include an overview of the project, instructions on how to run the code, and any other relevant information. Don't forget to write your name, surname and email at the top of the README file.
  - If you have prepared a document summarizing your findings, include it in the repository. Don't forget to write your name, surname and email at the top of the document file.
  - Everyone should create their own GitHub repository. When the project is complete, don't forget to send your GitHub project link to the mail address you received.
  - Your GitHub repo name should be: (Pusula_Name_Surname)

NOTE:

- Those who do not comply with the Submission Requirements will not be considered for evaluation.
- You must submit your project until **06.09.2025 23:59**