



Data Science Intern Case Study

Overview:

You will be working on a dataset related to drug side effects. Your task is to perform an in-depth analysis and prepare the data for further predictive modeling.

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:

- **Code Standards:**
 - Write your code following the PEP-8 guidelines. Ensure your code is well-organized, easy to read, and adheres to best practices.
- **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.
- The code part should not take more than one day. You must submit your project until 22.09.2024.
- If you have any problems and questions during the submit phase of the project, you can contact us via e-mail.
 - ozlem.akcakoca@pusulayazilim.com.tr
 - abdurrahman.keskin@pusulayazilim.com.tr