أكاديمية سدايا
SDAIA Academy

Project proposal:

Cardiovascular diseases are the leading cause of death worldwide. In this “Cardiovascular” dataset, we will use it to build classification models and try to analyze and collect insights from the data set and predict the likelihood of a person developing cardiovascular disease based on different criteria defined in this data set.

Dataset:

The data that will be used in this project is downloaded from Kaggle.com (https://www.kaggle.com/sulianova/cardiovascular-disease-dataset). The data is Cardiovascular Disease data.The dataset consists of over 70000 observations with 14 features.

1. **Age** | Objective Feature | age | int (days)
2. **Height** | Objective Feature | height | int (cm) |
3. **Weight** | Objective Feature | weight | float (kg) |
4. **Gender** | Objective Feature | gender | categorical code |
5. **Systolic blood pressure** | Examination Feature | ap_hi | int |
6. **Diastolic blood pressur**e | Examination Feature | ap_lo | int |
7. **Cholesterol** | Examination Feature | cholesterol | 1: normal, 2: above normal, 3: well above normal |
8. **Glucose** | Examination Feature | gluc | 1: normal, 2: above normal, 3: well above normal |
9. **Smoking** | Subjective Feature | smoke | binary |
10. **Alcohol intake** | Subjective Feature | alco | binary | 11.**Physical activity** | Subjective Feature | active | binary |
11. **Presence or absence of cardiovascular disease** | Target Variable | cardio | binary |

Tools:

- **Technologies:** Python, Jupyter Notebook , Sqlite.
  Libraries: Pandas, NumPy, Seaborn, Matplot, and sklearn, sqlalchemy .