FATIMAH OBAIDALLAH ALAHMADI

ABSTRACT:

The purpose of this project is to diagnose cardiovascular disease based on several features and symptoms given by the client I will use the features to determine if the disease exists or not in order to be able to warn the client and notify him either way. I worked with a dataset provided by the Kaggle website, dataset was collected at the moment of the medical examination. First, I started to explore the dataset, then I used clean data by dropping and creating some features. Then visualize the data to help identify the features that led diagnosis of cardiovascular disease by using Matplotlib and Seaborn tools, after that, I build Decision Tree, Bagging, Random Forest, logistic regression, and XGBoost model to predict clients from the test set that is most likely to diagnose cardiovascular disease.

The design:

This project helps to know the features that contribute to the client's cardiovascular disease. Knowing the features that cause the disease helps reduce the injury by treating these features.

Data Description:

- The dataset from medical examination which were collected at the moment of medical examination.
- Dataset consists of 70 000 records of patients data, 11 features + target link .

ALGORITHMS:

- clean data by dropping and creating some features.
- Then visualize the data to help identify the features that led diagnosis of cardiovascular disease by using Matplotlib and Seaborn tools,
- After that, I build Decision Tree, Bagging, Random Forest, logistic regression, and XGBoost model

Model	Accuracy
Decision Tree	64%
Bagging	68%
Random Forest	72%
Logistic Regression	72%
XGBoost	74%

The beast model was XGBoost.

TOOLS:

- Programming Language: Python
- Environment: Jupyter notebook
- I will use different types of Python libraries for data science :
    - NumPy
    - Seaborn
    - Pandas
    - Matplotlib
    - SciKit-Learn
    - Xgboost

COMMUNICATION:

Presentation that includes visuals for communicating the objectives and findings.