Case study: DESIGN AND ANALYSIS OF OPTIMIZED CLASSIFIER FOR PREDICTION OF HEART DISEASES.

Scenario: Heart diseases remain one of the leading causes of mortality worldwide. Early and accurate prediction of heart disease can significantly improve patient outcomes by enabling timely interventions. Traditional diagnostic methods, while effective, often require extensive medical expertise and can be time-consuming. Thus, developing an optimized machine learning classifier to predict heart disease using readily available patient data is crucial.

Ask:

Introduction: Welcome to the DESIGN AND ANALYSIS OF OPTIMIZED CLASSIFIER FOR PREDICTION OF HEART DISEASES. . In order to answer the business questions, follow the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path. Heart disease remains the leading cause of death, such that nearly one-third of all deaths worldwide are estimated to be caused by heart-related conditions. Coronary heart disease (CHD) /Acute coronary syndrome (ACS). The mortality rate is expanding due to obesity, cholesterol, high blood pressure and usage of tobacco among the people. An accurate prediction of heart disease is necessary for the early stage of treatment and overcoming the mortality rate. Machine learning (ML) can be used to assist clinical decision-making. We developed a ML model for the prediction of 1-year mortality after heart transplantation (HT) in adults with congenital heart disease. This study proposes a machine learning approach to Prediction of heart disease more accurately. Moreover, Accuracy, F1-Measure, Precision, and Recall are used to measure the performance of machine learning models. Different types of supervised machine learning model can be applied on data set. We wanted to design and develop machine learning model for Prediction of heart disease. In addition, a user-friendly web app and a user-friendly mobile app are built based on the most accurate model.

Keywords: Heart disease, supervised machine learning, Accuracy, F1-Measure, Precision, and Recall, Machine Learning, Prediction Model, Healthcare, Data Analysis.

Prepare:

- **Overview**

Source: UCI Machine Learning Repository

Number of Instances: 303

Number of Attributes: 14 (including the target variable)

Target Variable: Presence of heart disease (binary classification: 0 = absence, 1 = presence)

- **Attributes**

Age: Age of the patient in years.

Sex: Sex of the patient (1 = male, 0 = female).

Chest Pain Type (cp):

0: Typical angina

1: Atypical angina

2: Non-anginal pain

3: Asymptomatic

Resting Blood Pressure (trestbps): Resting blood pressure in mm Hg.

Serum Cholesterol (chol): Serum cholesterol in mg/dl.

Fasting Blood Sugar (fbs): Fasting blood sugar > 120 mg/dl (1 = true, 0 = false).

Resting Electrocardiographic Results (restecg):

0: Normal

1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

2: Showing probable or definite left ventricular hypertrophy by Estes' criteria.

Maximum Heart Rate Achieved (thalach): Maximum heart rate achieved during the test.

Exercise Induced Angina (exang): Exercise-induced angina (1 = yes, 0 = no).

Oldpeak (ST depression induced by exercise relative to rest): ST depression induced by exercise relative to rest.

Slope of the Peak Exercise ST Segment (slope):

0: Upsloping

1: Flat

2: Downsloping

Number of Major Vessels (ca) Colored by Fluoroscopy: Number of major vessels (0-3) colored by fluoroscopy.

Thalassemia (thal):

0: Normal

1: Fixed defect

2: Reversible defect

Target Variable (num): Diagnosis of heart disease (0 = no disease, 1 = disease).

Process: The dataset is often used for training and testing machine learning models to predict the presence of heart disease in patients. Common steps include:

1. **Data Preprocessing**: Handling missing values, encoding categorical variables, and scaling numerical features.
2. **Feature Selection**: Identifying the most relevant features that contribute to the prediction of heart disease.
3. **Model Training**: Implementing various machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, SVM, and Neural Networks.
4. **Model Evaluation**: Using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to evaluate the performance of the models.
5. **Hyperparameter Tuning**: Optimizing the parameters of the models to improve their predictive performance.

Analyze:

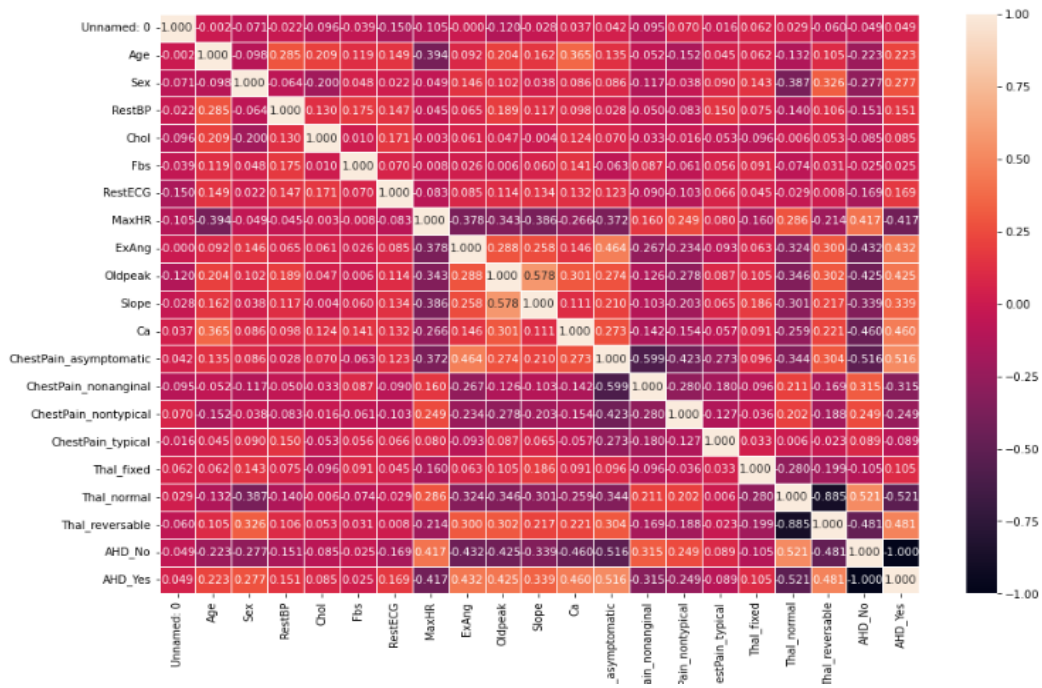
- Missing Values: The dataset may have missing values that need to be handled appropriately.
- Imbalanced Classes: The target variable may be imbalanced, with fewer instances of heart disease compared to no heart disease, which could affect model performance.
- Feature Scaling: Numerical features may require scaling for certain machine learning algorithms.
- Handling imbalanced datasets if the number of patients with heart disease is significantly lower than those without.
- Ensuring the model is generalizable and performs well on unseen data.
- Addressing any ethical concerns regarding patient data privacy and the implications of predictive diagnostics.

By understanding and processing this dataset effectively, one can develop robust models to predict heart disease, which can have significant implications for healthcare and patient management.

Share: Creating a heatmap is a great way to visualize the correlation between different features in the Heart Disease dataset. It helps in understanding which features are strongly correlated with each other and with the target variable. Let's generate a heatmap using Python's seaborn and pandas libraries.

In [33]:
```
cor_mat=data_final.corr()
fig,ax=plt.subplots(figsize=(15,10))
sns.heatmap(cor_mat,annot=True,linewidths=0.5,fmt=".3f")
```

Out[33]: <AxesSubplot:>



Act: The Case study demonstrates the potential of machine learning in enhancing heart disease diagnosis. Continuous evaluation and collaboration with healthcare professionals are essential for effective integration into clinical practice and addressing real-world challenges.