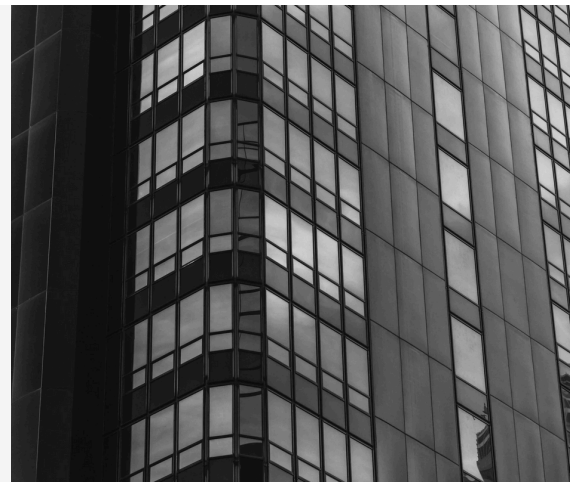


CodeAlpha Internship



Task Title: Disease Prediction from Medical Data

Task Description: Create a model to predict the likelihood of a disease based on medical data (e.g., symptoms, patient history). use datasets with labeled medical records and apply classification algorithms.

Disease Prediction from Medical Data

A "Disease Prediction from Medical Data" model typically involves the use of machine learning algorithms to predict the likelihood of a disease based on patient data.

Objective:

The goal is to develop a machine learning model that can predict the presence or risk of a specific disease based on a set of input medical data such as patient demographics, laboratory results, symptoms, and medical history.

Data Collection

The model uses a dataset containing patient records, including features such as:

- **Demographics:** Age, gender, ethnicity, etc.
- **Vital Signs:** Blood pressure, heart rate, body temperature, etc.
- **Laboratory Results:** Blood test results, urine analysis, cholesterol levels, etc.
- **Medical History:** Previous diagnoses, family medical history, medications, etc.
- **Symptoms:** Reported symptoms and their severity.

Feature Selection

Data Cleaning: Handle missing values, remove outliers, and correct any inconsistencies in the data.

Feature Engineering: Create new features, such as BMI from weight and height, or categorize continuous variables (e.g., age groups).

Normalization/Standardization: Scale features like glucose levels and BMI for uniformity.

Encoding Categorical Variables: Convert categorical data into numerical values (e.g., one-hot encoding for gender).

Model Training

Splitting Data: The dataset is divided into training and testing sets, often using an 80-20 split.

Training: The model is trained with Random Forest Classifier on a subset of the data, learning to identify patterns and relationships between features and outcomes.

Validation: The model's performance is tested on a separate dataset to ensure it generalizes well to unseen data. Metrics like AUC-ROC, precision, recall, and F1-score are commonly used for evaluation.

Model Evaluation

Test the Model: Evaluate the model on a separate test dataset to measure its accuracy and generalization.

Confusion Matrix: Generate a confusion matrix to see how well the model performs on each class.

Accuracy and Loss Plots: Plot the training and validation accuracy/loss over epochs to visualize the model's learning process.