Lung Cancer

NAME: Bhagawath Narayanan N

Email:bhagawath57@gmail.com

Institution:Rajalakshmi engineering college

internship ID: UMID25092559291

contact no:6374835866

Title: **Lung Cancer Prediction Using Machine Learning**

Description:

Lung cancer is one of the leading causes of death worldwide, and early detection plays a crucial role in improving patient survival rates. This project focuses on predicting the likelihood of lung cancer using machine learning techniques based on patient medical history, lifestyle habits, and demographic information.

The system performs complete data preprocessing including data cleaning, handling missing values, feature encoding, and scaling. Multiple classification models such as Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine are trained and evaluated to identify the most accurate predictor. Model performance is assessed using standard evaluation metrics including accuracy, precision, recall, and F1-score.

A preprocessing pipeline is implemented using ColumnTransformer and Pipeline to ensure efficient and scalable data transformation. The best-performing model is saved using joblib, allowing it to be reused for future predictions or deployment in real-world

applications. This project demonstrates the effective use of machine learning for medical risk prediction and decision support systems.

Tech Stack:

Programming Language:

- **Python**

Data Analysis & Processing:

- **Pandas** – Data manipulation and cleaning
- **NumPy** – Numerical computations

Data Visualization:

- **Matplotlib** – Graphs and plots for analysis

Machine Learning:

- **Scikit-learn**
  - Logistic Regression
  - Random Forest Classifier
  - Gradient Boosting Classifier
  - Support Vector Machine (SVM)
  - Pipeline & ColumnTransformer
  - OneHotEncoder & StandardScaler

Model Persistence:

- **Joblib** – Saving and loading trained models

Development Environment:

- **Jupyter Notebook / Google Colab**