Minor Project - Lung Cancer Prediction Project

Problem Statement

The objective of this project is to develop a machine learning model to predict the likelihood of lung cancer based on a set of given features. Lung cancer is one of the leading causes of cancer-related deaths globally, and early detection is crucial for effective treatment. By leveraging machine learning techniques, we aim to build a predictive model that can assist in the early diagnosis of lung cancer, potentially saving lives and improving patient outcomes.

What We Are Trying to Solve

The primary goal is to predict whether a person has lung cancer based on various attributes related to their lifestyle, medical history, and symptoms. The prediction model will use the provided dataset to learn patterns and relationships between these attributes and the presence of lung cancer. The outcomes of this project can be used to raise awareness and encourage timely medical consultations for those at higher risk.

Dataset Information

The dataset consists of 309 entries with 16 columns. Each row represents an individual, and each column represents a specific feature or attribute of the individual.

Column Information

- GENDER: Gender of the individual (M: Male, F: Female)
- AGE: Age of the individual
- SMOKING: Smoking status (YES = 2, NO = 1)
- YELLOW_FINGERS: Presence of yellow fingers, a symptom often associated with smoking (YES = 2, NO = 1)
- ANXIETY: Presence of anxiety (YES = 2, NO = 1)
- PEER_PRESSURE: Influence of peer pressure (YES = 2, NO = 1)
- CHRONIC_DISEASE: Presence of chronic disease (YES = 2, NO = 1)
- FATIGUE: Presence of fatigue (YES = 2, NO = 1)
- ALLERGY: Presence of allergy (YES = 2, NO = 1)
- WHEEZING: Presence of wheezing (YES = 2, NO = 1)
- ALCOHOL_CONSUMING: Alcohol consumption status (YES = 2, NO = 1)
- COUGHING: Presence of coughing (YES = 2, NO = 1)
- SHORTNESS_OF_BREATH: Presence of shortness of breath (YES = 2, NO = 1)
- SWALLOWING_DIFFICULTY: Presence of difficulty in swallowing (YES = 2, NO = 1)

- ❖ CHEST_PAIN: Presence of chest pain (YES = 2, NO = 1)
- ❖ LUNG_CANCER: Presence of lung cancer (YES, NO)

Model Comparison Report

Create a report stating the performance of multiple models on this data and suggest the best model for production.

Report on Challenges faced

Create a report which should include challenges you faced on data and what technique used with proper reason.

Note:-All above tasks have been created on a single jupyter notebook and share the same while final submission of project to the following email_ID: **krutanic@gmail.com**