

Detect Lung Cancer using patient diagnosis data.

Objective

Build a system that can predict the survival of a patient given details of the patient. Explore the data to understand the features and figure out an approach.

Dataset

This dataset contains data about lung cancer Mortality and is a comprehensive collection of patient information, specifically focused on individuals diagnosed with cancer.

Description of columns:

- id: A unique identifier for each patient in the dataset.
- age: The age of the patient at the time of diagnosis.
- gender: The gender of the patient (e.g., male, female).
- country: The country or region where the patient resides.
- diagnosis_date: The date on which the patient was diagnosed with lung cancer.
- cancer_stage: The stage of lung cancer at the time of diagnosis (e.g., Stage I, Stage II, Stage III, Stage IV).
- family_history: Indicates whether there is a family history of cancer (e.g., yes, no).
- smoking_status: The smoking status of the patient (e.g., current smoker, former smoker, never smoked, passive smoker).
- bmi: The Body Mass Index of the patient at the time of diagnosis.
- cholesterol_level: The cholesterol level of the patient (value).
- hypertension: Indicates whether the patient has hypertension (high blood pressure) (e.g., yes, no).
- asthma: Indicates whether the patient has asthma (e.g., yes, no).
- cirrhosis: Indicates whether the patient has cirrhosis of the liver (e.g., yes, no).
- other_cancer: Indicates whether the patient has had any other type of cancer in addition to the primary diagnosis (e.g., yes, no).
- treatment_type: The type of treatment the patient received (e.g., surgery, chemotherapy, radiation, combined).
- end_treatment_date: The date on which the patient completed their cancer treatment or died.
- survived: Indicates whether the patient survived (e.g., yes, no).