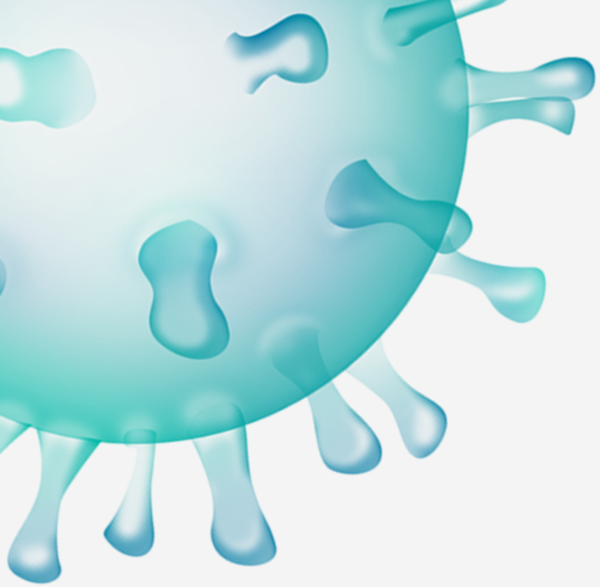
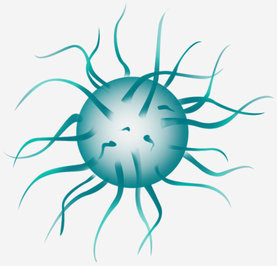
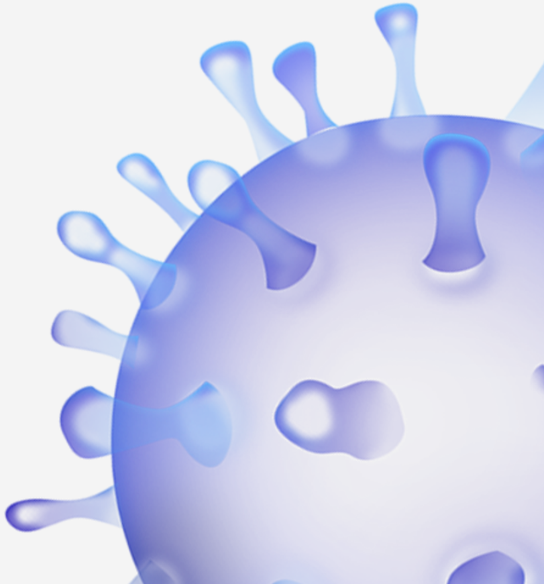



CANCER PREDICTIVE MODEL

By Haya Hadaya





Introduction to the Subject Area and Problem Statement

Subject Area Overview:

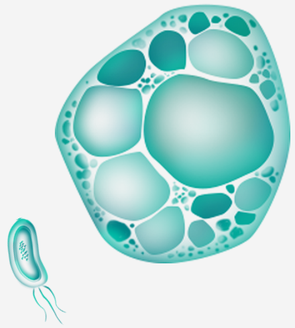
- Cancer is a prevalent global health concern, affecting millions worldwide.It is a disease in which some of the body’s cells grow uncontrollably and spread to other parts of the body

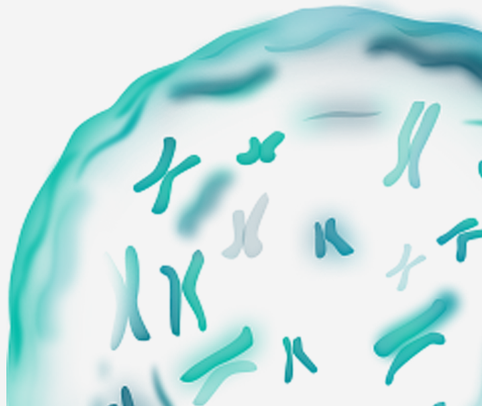
- Despite advancements, early detection remains critical for better patient outcomes

The Problem Statement

- The problem statement for cancer prediction involves addressing the challenge of early and accurate diagnosis, crucial for improving patient outcomes and reducing healthcare costs.

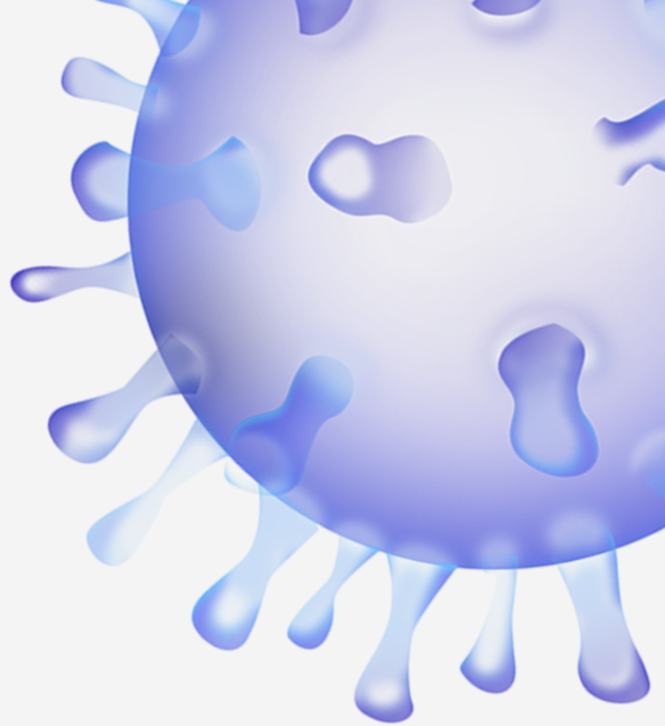
- The users who experience these problems are healthcare professionals such as doctors and genetic counsellors, patients, radiologists and healthcare systems.

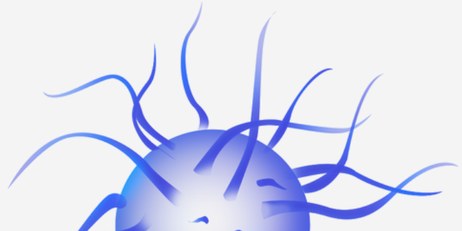




Machine learning and Cancer prediction

Machine learning is a powerful tool for predicting cancer by analyzing medical history, genetics, and lifestyle. It examines key factors like age, chronic diseases, obesity, and overall health to assess the risk of developing cancer. By recognizing patterns in various data points, these models can suggest early screening or lifestyle changes, offering personalized risk assessments for patients





Data Science-Driven Steps for Developing the Model

01 Data Collection and Cleaning
02 Exploratory Data Analysis
03 Feature Selection
04 Feature Engineering
05 Model Selection
06 Model Training
07 Model Evaluation
08 Hyperparameter Tuning
09 Validation and Testing

Early Diagnosis Impact: Machine Learning's Role in Healthcare Transformation

01 Early detection through machine learning aids in effective cancer treatments and improves patient outcomes by mitigating the rapid and painful progression of the disease.

02 Shifting focus to early intervention strategies significantly reduces healthcare costs, diverting resources from expensive late-stage treatments to cost-effective preventive measures.

03 Early detection can facilitate the development of targeted therapies, leading to more personalized treatment plans

04 Providing individuals with information about their cancer risk empowers them to make informed decisions about their health and lifestyle choices

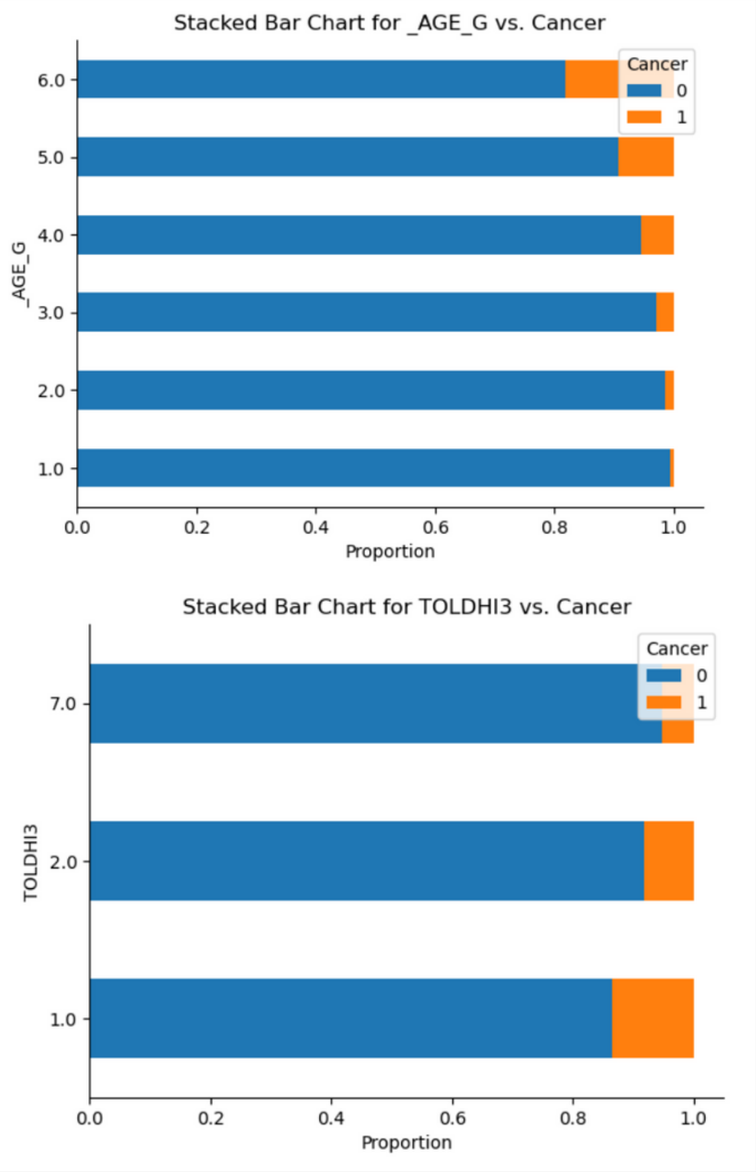
ANALYSIS

Cancer Prediction Dataset

- **The dataset has 438,693 rows and 303 columns , and it was condensed down to just 18 columns.**
- **All columns are categorical and there are some missing values handled by imputation and removal**
- **No duplicated rows or columns in the original dataset.**
- **The selected features are:**

- RFHLTH (Adult Health)
- SEXVAR (Gender)
- TOLDHI3 (High Cholesterol)
- CHECKUP1 (Routine Checkup Visits)
- CHCCOPD3 (Chronic Disease)
- HAVARTH5 (Arthritis Related Conditions)
- DIABETE4 (Diabetes)
- _RFHYPE6 (High Blood Pressure)
- _PHYS14D (Physical Health Status)
- _AGE_G (Age Category)
- _IMPRACE (Race/Ethnicity)
- _RFBMI5 (Overweight/Obesity)
- _SMOKER3 (Smoking Status)
- _RFDRHV7 (Heavy Drinking)
- _TOTINDA (Physical Activity)
- CHCKDNY2 (Kidney Disease)
- _FRTLT1A (Fruit Consumption)
- Cancer (Cancer Risk)

Top Key Findings from EDA



- **There is a higher association between older age groups and cancer presence.**
- **A notable proportion of cancer patients have reported high cholesterol levels and high blood pressure.**
- **Females constitute a higher number within the group of patients diagnosed with cancer.**