Cancer Susceptibility

Creating a lung cancer risk predictor.

Selected Topic & Reason for Topic

Topic:

- Susceptibility to lung cancer based on lifestyle and demographic parameters such as age, gender, alcohol use, genetic risk, and smoking.

Reason for Topic:

- Being able to define a patient's risk level of getting lung cancer can help them make changes in lifestyle choices to avoid contraction.

Description of Data Source

This dataset was sources from Kaggle, a community form of datasets. It shows the demographic and lifestyle data of 1000 lung cancer patients.

Questions to Answer

- Which individual aspects of a patient's lifestyle would make them most susceptible to lung cancer?

- Which combination of a patient's lifestyle would make them most susceptible to lung cancer?

- Which aspects of a patient's lifestyle would make them least susceptible to lung cancer?

Description of Data Exploration Phase

- After defining the questions and topic, we searched for relevant datasets on Google Dataset Search and Kaggle.

- We were specifically looking for data on cancer patients and their lifestyle choices leading up to them contracting cancer.

- We also searched for data on cancer patient demographics.

Description of Data Analysis Phase

- The dataset was prepared for the machine learning model by cleaning, removing unnecessary information, and converting all data to numerical values.

- Statistical information was extracted using python and Excel.

- The dataset was visualized using python to find and display trends.

- The dataset was put through a supervised machine learning algorithm to create a predictive model.

Dashboard