Problem Statement

The United States Health Department aims to reduce the annual budget allocated to colorectal cancer-related services by 10% in 2025, while maintaining the quality of care provided to patients. The project is set to be completed by the end of 2025, with periodic reviews to ensure progress towards the 10% budget reduction target. What is the projected average cost of patient care in the United States for the upcoming year?

Context:

Colorectal cancer is the second leading cause of cancer deaths, projected to result in 52,900 fatalities in the U.S. by 2025. Survival rates indicate the percentage of individuals alive five years post-diagnosis but don't predict individual lifespans. They reflect treatment success likelihood. The United States Health Department offers various services to patients, but recent challenges from COVID-19 and drug overdose crises have strained resource allocation. In 2025, they aim to cut the budget for colon cancer services by 10% without compromising care quality. Previous survival predictions by health professionals lacked accuracy. Can machine learning identify high-risk patients and estimate next year's average patient care costs?

https://www.cancer.org/cancer/types/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html

Criteria for Success:

By processing available data and developing a predictive model, health professionals can make informed decisions to optimize resource allocation. This objective is attainable given the advancements in machine learning and the availability of data. By identifying patients at risk, the Health Department intends to reallocate necessary health services and reduce unnecessary services by providers. This approach will enable the Health Department to decrease healthcare claim costs by 10% compared to last year, by the end of 2025. Additionally, along with improving survival rates, the Health Department will be able to prepare a budget proposal based on the average cost of patient care.

Scope of solution space:

The model prediction will be evaluated with professional opinions to address patient needs individually. Analysis will be limited to the USA since the data is global.

Constraints within solution

- Since this dataset is global, its application to specific countries or regions may be constrained.
- The available dataset includes a relatively small amount of pathological data. The data analysis will be based on this dataset, which has limited pathological data and more epidemiological data.

Stakeholders to provide key insight:

- State Health department officials
- State Budget committee members
- Data Analysis team

Key Data Sources:

The dataset offers real-world data on colorectal cancer cases from various countries, covering patient demographics, lifestyle risks, medical history, cancer stage, treatments, survival rates, and healthcare costs. It reflects global trends in colorectal cancer incidence, mortality, and prevention.

https://www.kaggle.com/datasets/ankushpanday2/colorectal-cancer-global-dataset-and-predictions/code