Topic Idea 1: <Heart disease Analysis & Prediction>

Qi Li

1. Problem Statement

Heart disease is a serious medical emergency. Thus, the topic of our project is Heart disease Analysis and Prediction. Our goal is to build a model that can predict the probability of heart disease according to the combination of physical characteristics that describe the disease. Heart disease is the leading cause of death worldwide. We realized that if we can predict heart diseases as accurately as possible, it may be helpful for the treatment of heart disease.

2. Significance of the Problem

Heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the whole world,accounting for one-third of deaths in 2019. What's worse, the death toll continues to rise. In the US, about 647,000 Americans die from heart disease every year, which is 1 in every 4 deaths. One person dies every 36 seconds in the United States from cardiovascular disease. If we can extract the relevant indicators of human bodies' physical examination and then analyze the impact of different human body characteristics on heart disease through data mining, it will play a vital role in the prevention and treatment of heart diseases.

3. Potential Datasets

Heart Disease Prediction
Link: https://data.world/informatics-edu/heart-disease-prediction

HEART DISEASE DATASET (COMPREHENSIVE)
Link: https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive

An integrated polygenic tool substantially enhances coronary artery disease prediction
Link: https://zenodo.org/record/4421038#.YK6_OGYzYq0

Heart Attack Prediction
Link: https://www.kaggle.com/imnikhilanand/heart-attack-prediction

Heart Disease Data Set
Link: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

data : https://www.kaggle.com/imnikhilanand/heart-attack-prediction
heart : https://www.kaggle.com/ronitf/heart-disease-uci
Heart_Disease_Prediction : https://data.world/informatics-edu/heart-disease-prediction?oauthError=github

Topic Idea 2: <What makes you become a famous vlogger?>

Yuxi Shen

1. Problem Statement

The topic of this project is “What makes you become a famous vlogger?”.
There are thousands of outstanding vloggers on Youtube sharing their daily life, and their popularity can be seen from the thousands of thumb ups and millions of followers they get. To accurately predict the key characteristics which should be included in a short video, I hope to use various dataset resources to analyze how each vlogger with a different personality attracts millions of fans through their videos. To do so, I would like to find some datasets containing the information of time length, language speaking, objects presented, and find the correlation between these components.

2. Significance of the Problem

Youtube is a popular video-sharing social website. Individuals can upload their videos on it, and the videos would be “rated” by the thumbs up and the number of collections they own. By analyzing the successful videos, all the industries could benefit from it. For example, the advertising industry can understand what should be in short videos to attract consumers. The make-up industry can go to find the one with attractive facial images to better sell their products, and so on.

3. Potential Datasets

Kaggle :
Trending Youtube Video Statistic
Link : https://www.kaggle.com/datasnaek/youtube-new

Youtube :
Segment-rated frame-level features dataset
Frame-level features dataset
Video-level features dataset
Link : https://research.google.com/youtube8m/download.html

Youtube :
Classification in video segments
Object detection in video segments
Link : https://research.google.com/youtube-bb/download.html

Most viewed YOuTube channel owners of all time as of February 2021, by views.
Link : https://www.statista.com/statistics/373753/most-viewed-youtubers-all-time/

Topic Idea 3: <Prediction of COVID cases on college campuses in the United States>

Ruoxuan Wang

1. Problem Statement

The safe reopening of college campuses amidst the coronavirus pandemic has been a challenging problem for the past year. With several claims that college campuses are a “hot spot” for COVID cases, many universities have tried to limit the number of students on campus or close campus all entirely. The purpose of this study is to see whether or not the reopening of college campuses did increase the spread of coronavirus in the region.

2. Significance of the Problem

The results from this study could be used to determine whether or not reopening college campuses did increase the spread of the coronavirus. If it is found that certain colleges were able to effectively mitigate the spread of coronavirus to the local community, understanding the restrictions they put in place could be important and provide a standard for college reopenings should another pandemic ever occur. These mitigation methods could also be used in reference to allow for a safe reopening of the country.

3. Potential Datasets

COVID cases for college campuses + United States by state
https://github.com/nytimes/covid-19-data

COVID cases for United States by state
https://covidtracking.com/data