Heart Disease Health Indicators

Abstract

The goal of this project Use survey responses from the (BRFSS) Behavioral Risk Factor Surveillance System for predicting heart disease risk , preventive health practices and risk behaviors that are linked to chronic diseases in the adult population. Heart Disease is among the most prevalent chronic diseases in the United States, impacting millions of Americans each year and exerting a significant financial burden on the economy. In the United States alone, heart disease claims roughly 647,000 lives each year — making it the leading cause of death .

Design

The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Use survey responses from the (BRFSS))Behavioral Risk Factor Surveillance System for predicting heart disease risk , preventive health practices and risk behaviors that are linked to chronic diseases s in the adult population.

Data

This dataset contains 253,680 survey responses from BRFSS .Contains 22 colons (features) is (High Blood Pressure, High Cholesterol, Cholesterol Check, Smoker, Stroke, Body Mass Indicator (BMI), Diabetes, Phys Activity, Fruits, Any Health care, No Docbc Cost, General Health, Mental Health , Diff Walk ,Sex, Education, Income).

https://github.com/maryam0003/Data

Algoritm

Model

- LogisticRegression
- KNeighborsClassifier
- RandomForestClassifier
- XGBClassifier

- AdaBoostClassifier
- GradientBoostingClassifie

Model Evaluation and Selection

- The entire training dataset of 253,680 records was split into 80/20 train
- Random Forest Classifier gave me the best results and high accuracy

Modeling

KNeighborsClassifier

Logistic Regression

RandomForestClassifier

GradientBoostingClassifier

XGBClassifier

AdaBoostClassifier

Accuracy: 88.2522%
Log Loss: 4.0575
Score : 0.88046

Accuracy: 89.5423%
Log Loss: 3.6119
Score : 0.89423

Accuracy: 90.2850%
Log Loss: 3.3554
Score : 0.8879
.

Accuracy: 89.5149%
Log Loss: 3.6214
Score : 0.8930

Accuracy: 88.0601%
Log Loss: 4.1239
Score : 0.8937

Accuracy: 89.5194%
Log Loss: 3.6198
Score : 0.8921

Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting
- Tableau for interactive visualizations

Communication
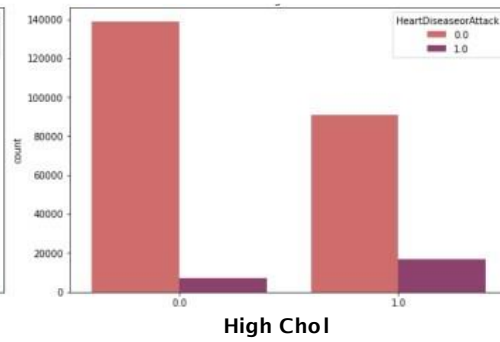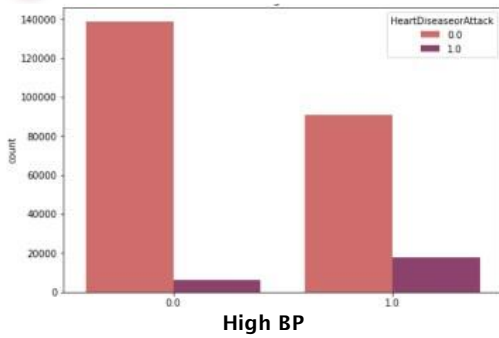
From the graph, it is clear to us from the sample of people that the rate of heart disease increases in people who suffer from high blood pressure and high cholesterol

From the graph, the incidence of heart disease increases among smokers, in addition that the majority of people with strokes have heart disease.

From the graph heart disease is higher in men who smoke

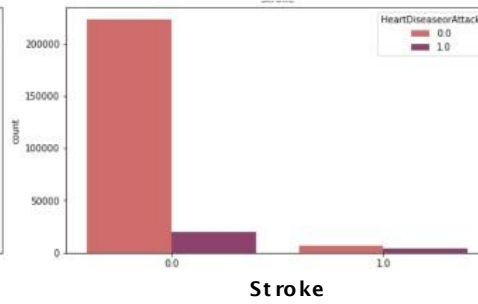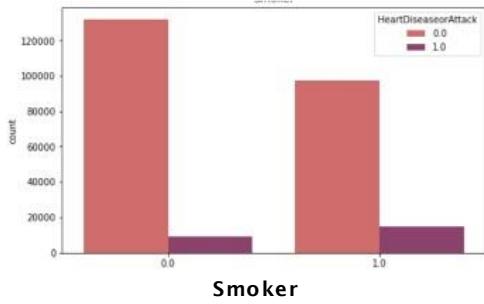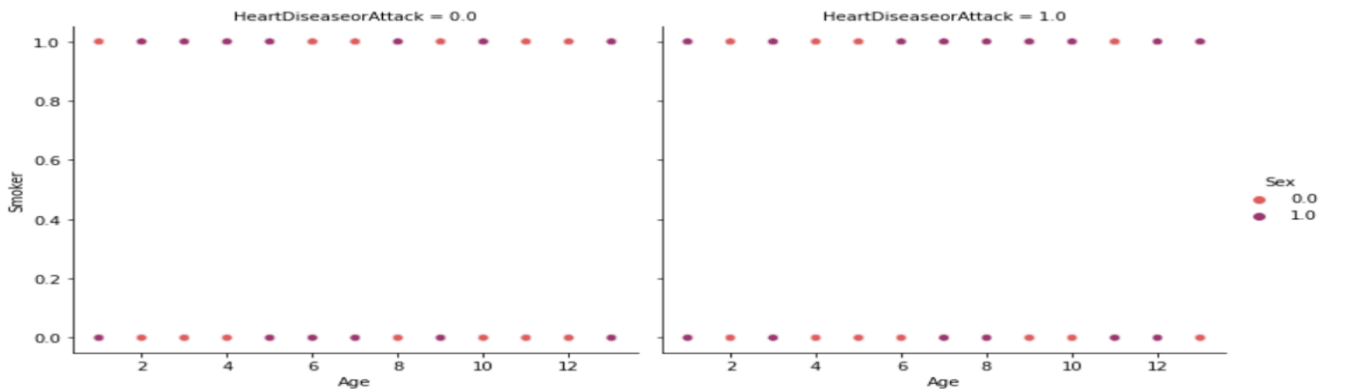|	HeartDiseaseorAttack	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	AnyHealthcare	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income
HeartDiseaseorAttack	1.00	0.20	0.18	0.05	0.03	0.10	0.20	0.17	-0.07	-0.00	-0.02	-0.04	0.03	0.02	0.24	0.05	0.17	0.20	0.09	0.23	-0.08	-0.11
HighBP	0.20	1.00	0.28	0.12	0.19	0.07	0.12	0.26	-0.10	-0.01	-0.04	-0.02	0.06	-0.00	0.26	0.03	0.14	0.21	0.05	0.34	-0.10	-0.13
HighChol	0.18	0.28	1.00	0.10	0.08	0.07	0.09	0.20	-0.06	-0.02	-0.02	-0.02	0.06	-0.00	0.18	0.05	0.11	0.13	0.02	0.26	-0.04	-0.05
CholCheck	0.05	0.12	0.10	1.00	0.05	-0.00	0.03	0.08	0.01	0.02	-0.00	-0.02	0.12	0.05	0.07	0.00	0.04	0.05	-0.03	0.10	-0.01	-0.00
BMI	0.03	0.19	0.08	0.05	1.00	-0.02	0.01	0.21	-0.12	-0.06	-0.04	-0.06	-0.00	0.04	0.20	0.06	0.09	0.18	0.03	-0.05	-0.06	-0.06
Smoker	0.10	0.07	0.07	-0.00	-0.02	1.00	0.05	0.04	-0.06	-0.06	-0.01	0.09	-0.01	0.03	0.13	0.07	0.10	0.10	0.10	0.10	-0.13	-0.09
Stroke	0.20	0.12	0.09	0.03	0.01	0.05	1.00	0.10	-0.05	-0.00	-0.03	-0.02	0.02	0.03	0.17	0.06	0.14	0.17	0.00	0.13	-0.06	-0.11
Diabetes	0.17	0.26	0.20	0.08	0.21	0.04	0.10	1.00	-0.09	-0.02	-0.04	-0.07	0.03	0.02	0.28	0.05	0.15	0.20	0.03	0.19	-0.10	-0.14
PhysActivity	-0.07	-0.10	-0.06	0.01	-0.12	-0.06	-0.05	-0.09	1.00	0.12	0.13	0.03	0.02	-0.04	-0.23	-0.10	-0.19	-0.23	0.03	-0.09	0.16	0.15
Fruits	-0.00	-0.01	-0.02	0.02	-0.06	-0.06	-0.00	-0.02	0.12	1.00	0.24	-0.03	0.02	-0.03	-0.06	-0.05	-0.02	-0.02	-0.09	0.08	0.08	0.04
Veggies	-0.02	-0.04	-0.02	-0.00	-0.04	-0.01	-0.03	-0.04	0.13	0.24	1.00	0.03	0.02	-0.01	-0.08	-0.03	-0.04	-0.06	-0.07	-0.00	0.12	0.11
HvyAlcoholConsump	-0.04	-0.02	-0.02	-0.02	-0.06	0.09	-0.02	-0.07	0.03	-0.03	0.03	1.00	-0.00	-0.00	-0.06	0.01	-0.04	-0.05	0.01	-0.04	0.04	0.08
AnyHealthcare	0.03	0.06	0.06	0.12	-0.00	-0.01	0.02	0.03	0.02	0.02	0.02	-0.00	1.00	-0.23	0.14	-0.02	0.01	0.02	-0.02	0.15	0.11	0.14
NoDocbcCost	0.02	-0.00	-0.00	-0.05	0.04	0.03	0.03	0.02	-0.04	-0.03	-0.01	-0.00	-0.23	1.00	0.14	0.18	0.13	0.10	-0.05	-0.13	-0.08	-0.18
GenHlth	0.24	0.26	0.18	0.07	0.20	0.13	0.17	0.28	-0.23	-0.06	-0.08	-0.06	0.14	0.14	1.00	0.28	0.51	0.44	-0.01	0.15	-0.23	-0.32
MentHlth	0.05	0.03	0.05	0.00	0.06	0.07	0.06	0.05	-0.10	-0.05	-0.03	0.01	-0.04	0.18	0.28	1.00	0.33	0.21	-0.08	-0.11	-0.07	-0.17
PhysHlth	0.17	0.14	0.11	0.04	0.09	0.10	0.14	0.15	-0.19	-0.02	-0.04	-0.04	0.01	0.13	0.51	0.33	1.00	0.46	-0.04	0.10	-0.12	-0.23
DiffWalk	0.20	0.21	0.13	0.05	0.18	0.10	0.17	0.20	-0.23	-0.02	-0.06	-0.05	0.02	0.10	0.44	0.21	0.46	1.00	-0.07	0.21	-0.16	-0.29
Sex	0.09	0.05	0.02	-0.03	0.03	0.10	0.00	0.03	0.03	-0.09	-0.07	0.01	-0.02	-0.05	-0.01	-0.08	-0.04	-0.07	1.00	-0.03	0.01	0.13
Age	0.23	0.34	0.26	0.10	-0.05	0.10	0.13	0.19	-0.09	0.08	-0.00	-0.04	0.15	-0.13	0.15	-0.11	0.10	0.21	-0.03	1.00	-0.09	-0.11
Education	-0.08	-0.10	-0.04	-0.01	-0.06	-0.13	-0.06	-0.10	0.16	0.08	0.12	0.04	0.11	-0.08	-0.23	-0.07	-0.12	-0.16	0.01	-0.09	1.00	0.41
Income	-0.11	-0.13	-0.05	-0.00	-0.06	-0.09	-0.11	-0.14	0.15	0.04	0.11	0.08	0.14	-0.18	-0.32	-0.17	-0.23	-0.29	0.13	-0.11	0.41	1.00