Using a dataset from Kaggle, I created a machine learning model that will be able to find out whether someone has or is likely to develop a heart disease depending on his/her characteristics.

The dataset comes from the Behavioural Risk Factor Surveillance System, a division of the CDC (Centres for Disease Control and Prevention, (USA). It had already been cleaned and the most important variables had been kept.

I first performed EDA to have a look at the data and see if I could find any interesting insight right away. I then proceeded to creating the model using Scikit learn and classification.

You can find the list of characteristics below :

* **Response Variable / Dependent Variable (y)**

Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI).

**Independent Variables: (X)**

* **High Blood Pressure**

Adults who have been told they have high blood pressure by a doctor, nurse, or other health professional.

* **High Cholesterol**

Have you EVER been told by a doctor, nurse or other health professional that your blood cholesterol is high?

Cholesterol check within past five years.

* **Smoking**

Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes].

* **Other Chronic Health Conditions**

(Ever told) you had a stroke.

(Ever told) you have diabetes (If "Yes" and respondent is female, ask "Was this only when you were pregnant?". If Respondent says pre-diabetes or borderline diabetes, use response code 4.).

* **Diet**

Consume Fruit 1 or more times per day.

Consume Vegetables 1 or more times per day.

* **Alcohol Consumption**

Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week).

* **Health Care**

Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare, or Indian Health Service?

Was there a time in the past 12 months when you needed to see a doctor but could not because of cost?

* **Health General and Mental Health**

Would you say that in general your health is: --> GENHLTH

Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good? --> MENTHLTH

Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good? --> PHYSHLTH

Do you have serious difficulty walking or climbing stairs? --> DIFFWALK

* **Demographics**

Indicate sex of respondent.

Fourteen-level age category.

What is the highest grade or year of school you completed?

Is your annual household income from all sources: (If respondent refuses at any income level, code "Refused.").

* **BMI**

Body Mass Index (BMI).

* **Physical Activity**

Adults who reported doing physical activity or exercise during the past 30 days other than their regular job.