HEART ATTACK PREDICTION

AGENDA

- Data description
- EDA
- Choose model
- Conclusion

INTRODUCTION

Heart attack prediction involves using risk factors and predictive models to identify individuals at high risk of experiencing a heart attack

.Machine learning and statistical analyses are key in analyzing large datasets to predict cardiovascular events, allowing healthcare providers to personalize care and mitigate heart attack risks efficiently

DATA DESCRIPTION

Age: Age of the patient.

Sex: Sex of the patient.

exang: Exercise-induced angina (1 = yes; 0 = no).

Angina = Chest pain ca: Number of major vessels (0-3). The variable "ca" with values ranging from 0 to 3 represents the number of major blood vessels (coronary arteries) that are colored or visualized during the fluoroscopy procedure. Each value corresponds to a different degree of vessel involvement: ca = 0: Indicates that there is no significant coloration of the vessels, suggesting a potential absence of major vessel blockages. ca = 1: Implies that one major vessel is colored, indicating a partial blockage or disease in one of the coronary arteries. ca = 2: Suggests that two major vessels are colored, signifying a more extensive blockage or disease involving multiple coronary arteries. ca = 3: Indicates that three major vessels are colored, suggesting a severe blockage or disease affecting most of the coronary arteries.

cp: Chest pain type. Typical Angina (Value 1): Substernal discomfort or pain, often squeezing or heaviness, triggered by exertion, indicative of coronary artery disease. Atypical Angina (Value 2): Non-classic symptoms, variations in pain characteristics, may not be clearly associated with physical exertion. Non-Anginal Pain (Value 3): Sharp or stabbing pain, not typical of angina, may have non-cardiac causes like musculoskeletal or gastrointestinal issues. Asymptomatic (Value 4): Absence of chest pain or discomfort, patient reports no symptoms related to chest pain, requiring further assessment of other risk factors.

trtbps: Resting blood pressure (in mm Hg).

chol: Cholesterol in mg/dl fetched via BMI sensor.

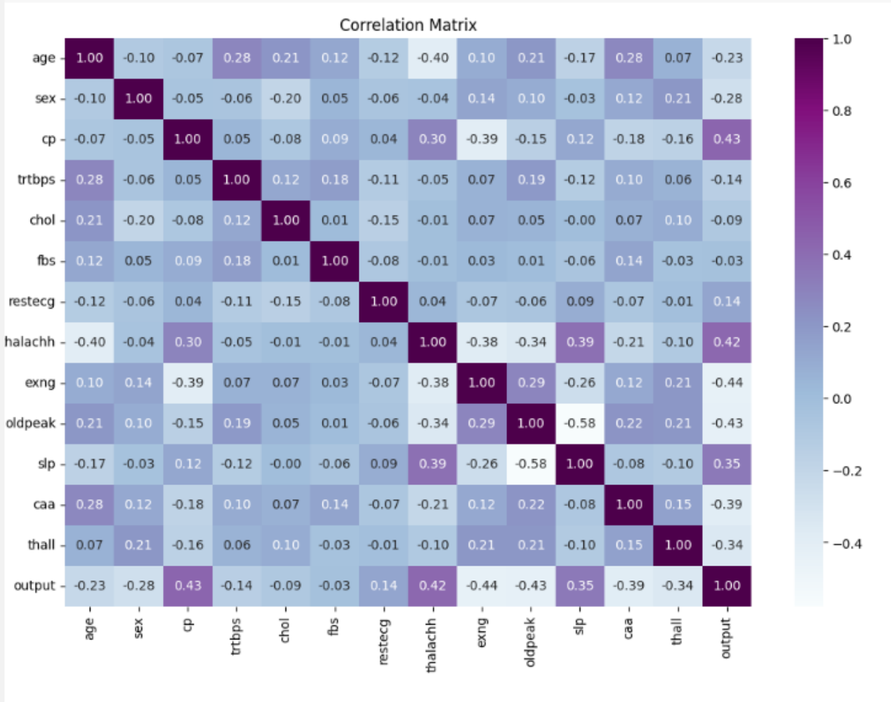
fbs: Fasting blood sugar > 120 mg/dl (1 = true; 0 = false).

rest_ecg: Resting electrocardiographic results. Value 0: Normal Value 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) Value 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria

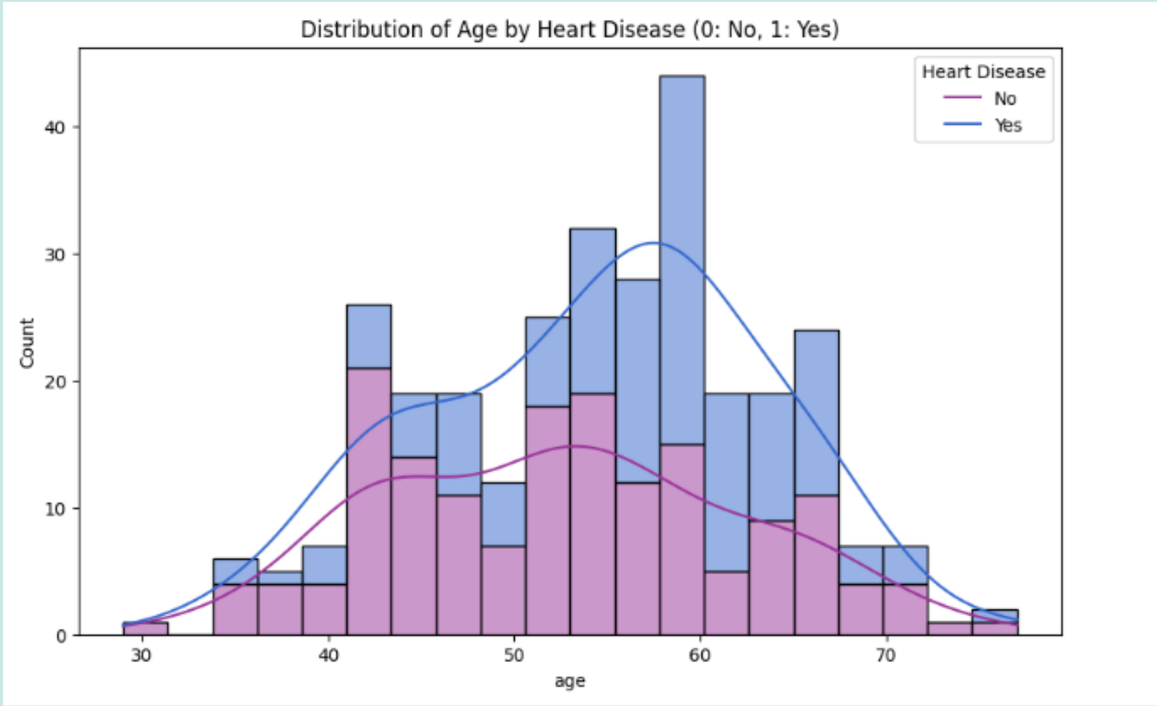
thalach: Maximum heart rate achieved.

target: 0 = Less chance of a heart attack, 1 = More chance of a heart attack.

EDA

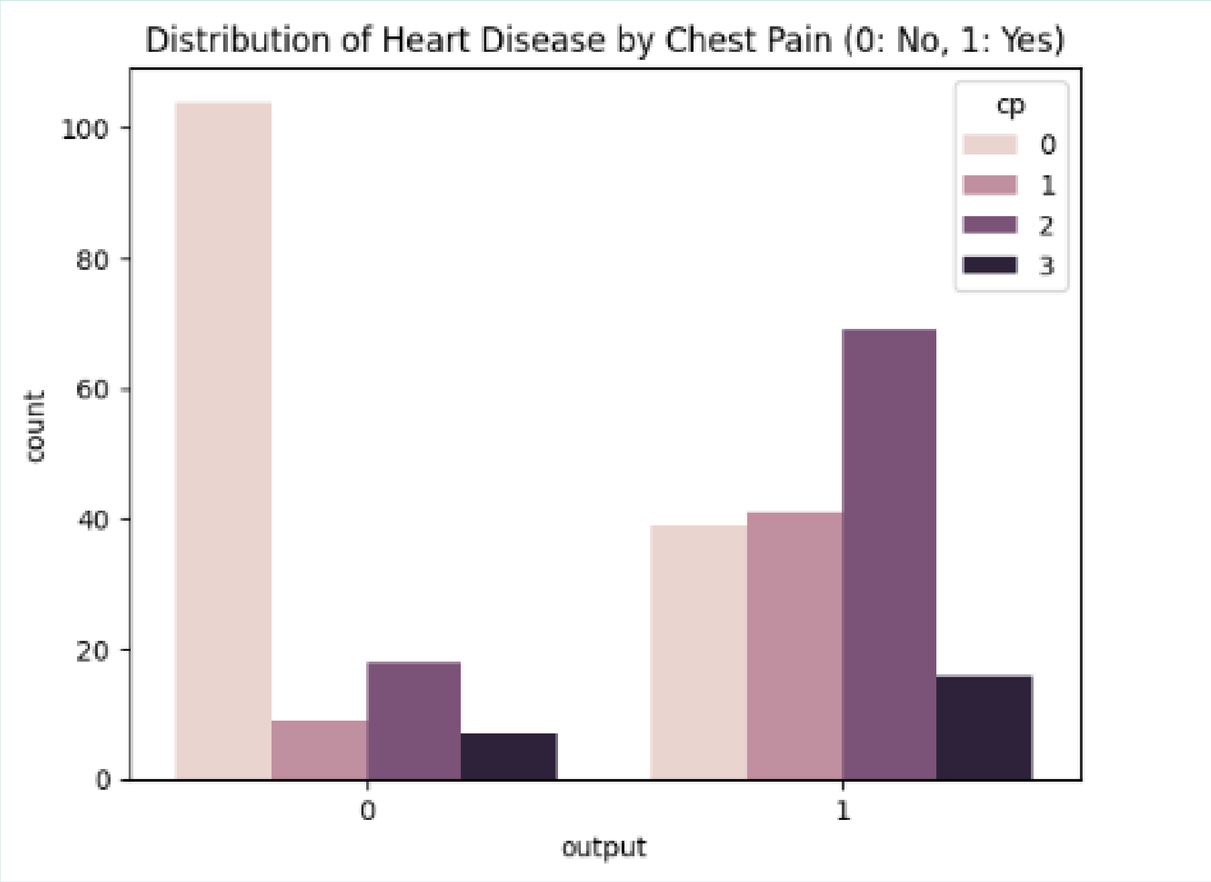


Heart rate and chest pain are highly correlated, with changes in heart rate often associated with the presence and severity of chest pain. This correlation is crucial to cardiovascular health, indicating conditions like heart disease or angina.



From the histogram, we can observe that the majority of the individuals with heart disease (blue bars) are in the age range of roughly 50 to 65 years old. In contrast, the individuals without heart disease (blue bars) have a more dispersed age distribution but seem to peak in the younger age groups, around 40-45 years old.

EDA



It can be seen that at a cp level of 0, the risk of heart disease is very low. However, when reaching level 1 or 2, the risk increases very high

CHOOSING MODEL

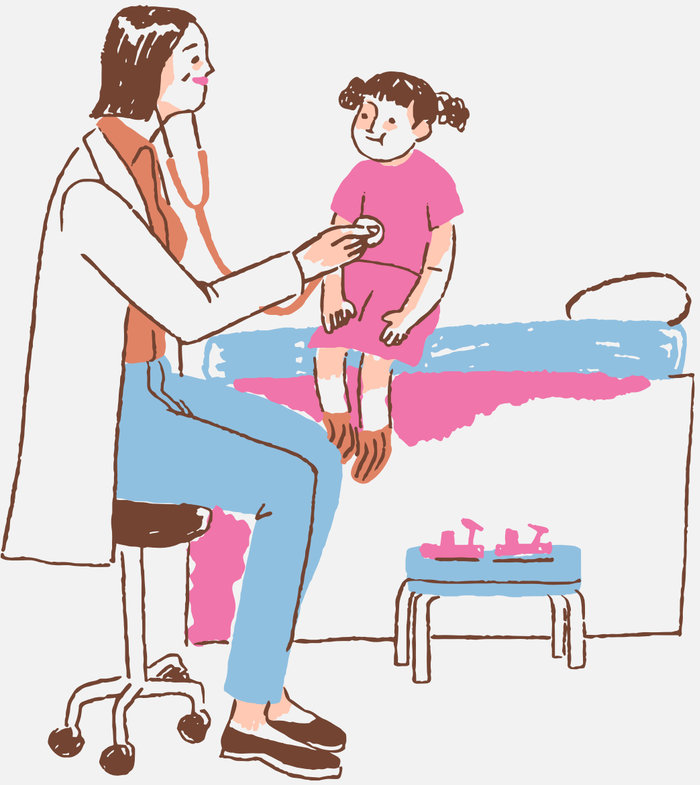
Chọn mô hình: sử dụng các mô hình như sau

- KNN
- RandomForestClassifier
- DecisionTreeClassifier
- Logistic Regression

Model	Accuracy
KNN	86.89%
Decision Tree	77.05%
Random Forest	83.61%
Logistic Regression	81.97%

TRAIN-TEST SPLIT

```
X = df.drop('output', axis=1)
y = df['output']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```



CONCLUSION

Based on the analysis and prediction model developed, it can be concluded that the risk of heart disease can be effectively predicted using the selected features and machine learning algorithms. The model shows promising accuracy and can be used as a valuable tool for early detection and prevention of heart disease. Early identification of individuals at risk can lead to timely interventions and improved outcomes in managing heart disease.