



Heart Attack Prediction

BIAI PROJECT

JAKUB ŁAPAJ

ADRIAN PIRÓG

Agenda

Machine learning models introduction

Project presentation

Data

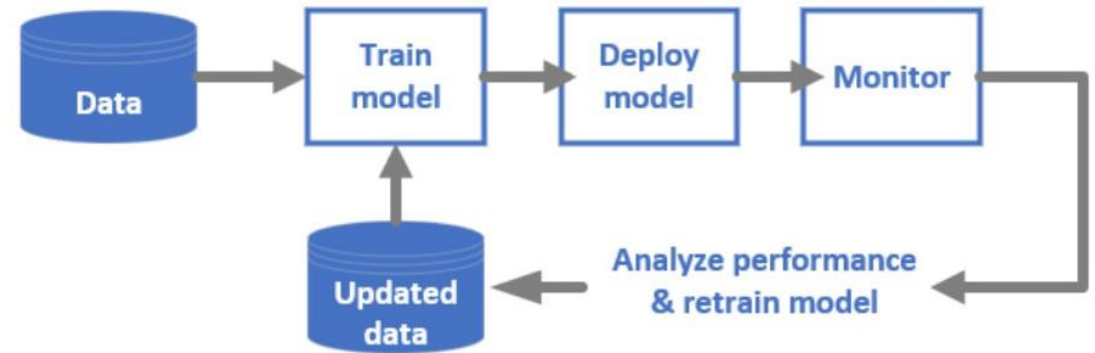
Models

Output analysis

Conclusions

Machine learning models

Machine learning models are algorithms that can learn patterns and make predictions or decisions based on data. They analyze and generalize from examples, rather than relying on explicit instructions. These models can be trained on labeled data, where the input features are associated with known outputs, or on unlabeled data, where they discover patterns and structures independently. Machine learning models can range from simple linear regression and decision trees to complex deep neural networks, each suited for different tasks and data types.



Project presentation

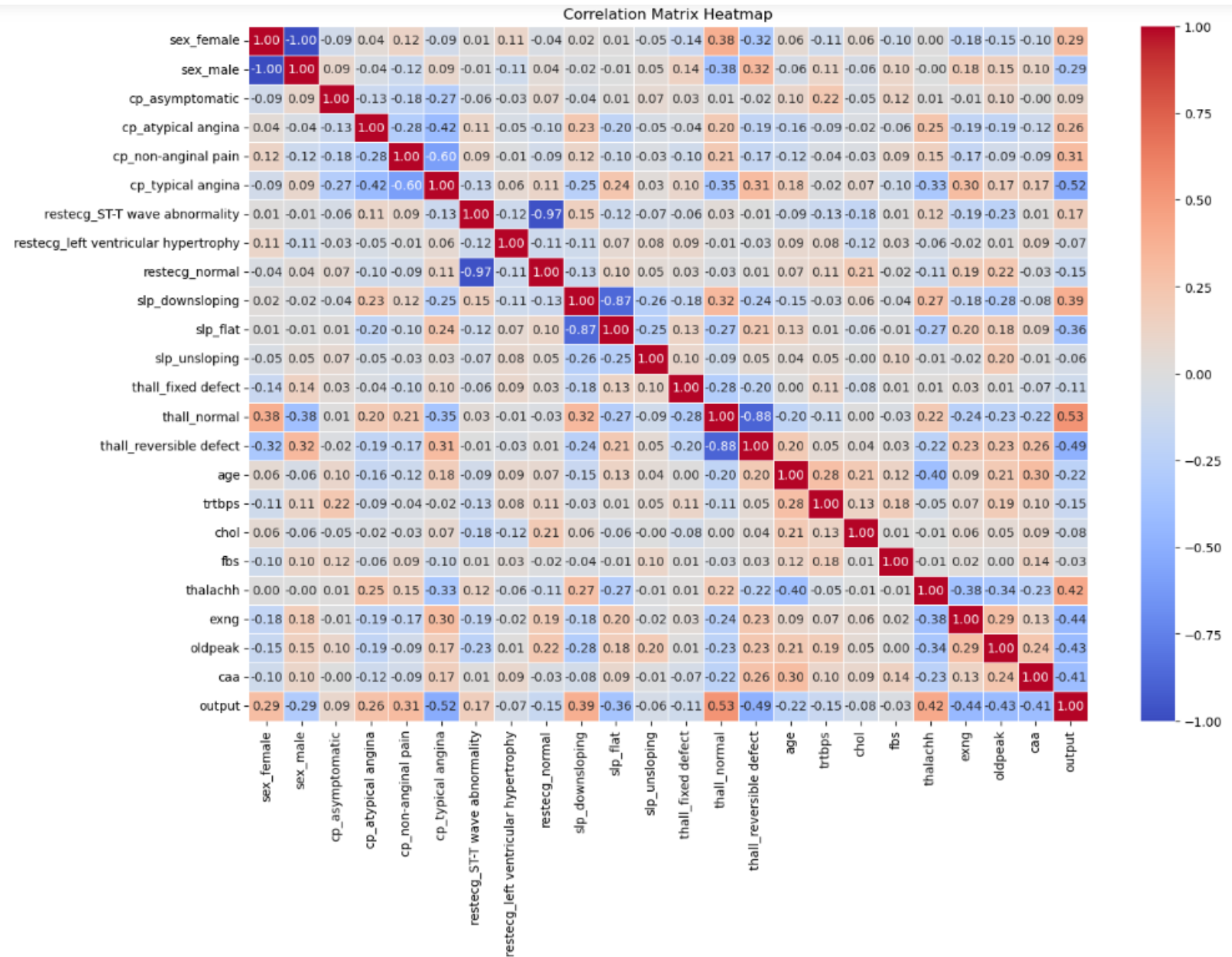
Our goal is to build a heart disease prediction model and compare the performance of various machine learning algorithms. We base on .csv dataset which includes parameters:

- Age
- Sex
- Chest Pain Type (cp)
- Resting Blood Pressure (trtbps)
- Serum Cholesterol (chol)
- Fasting Blood Sugar (fbs)
- Resting Electrocardiographic Results (restecg)
- Maximum Heart Rate Achieved (thalachh)
- Exercise Induced Angina (exang)
- ST Depression Induced by Exercise Relative to Rest (oldpeak)
- The Slope of The Peak Exercise ST Segment (slp)
- Number of Major Vessels Colored by Flourosopy (caa)
- Thallium Stress Test (thall)
- Output (Diagnosis of Heart Disease)

Data presentation

	A	B	C	D	E	F	G	H
1	age,sex,cp,trtbps,chol,fbs,restecg,thalachh,exng,oldpeak,slp,caa,thall,output							
2	63,1,3,145,233,1,0,150,0,2.3,0,0,1,1							
3	37,1,2,130,250,0,1,187,0,3.5,0,0,2,1							
4	41,0,1,130,204,0,0,172,0,1.4,2,0,2,1							
5	56,1,1,120,236,0,1,178,0,0.8,2,0,2,1							
6	57,0,0,120,354,0,1,163,1,0.6,2,0,2,1							
7	57,1,0,140,192,0,1,148,0,0.4,1,0,1,1							
8	56,0,1,140,294,0,0,153,0,1.3,1,0,2,1							
9	44,1,1,120,263,0,1,173,0,0,2,0,3,1							
10	52,1,2,172,199,1,1,162,0,0.5,2,0,3,1							
11	57,1,2,150,168,0,1,174,0,1.6,2,0,2,1							
12	54,1,0,140,239,0,1,160,0,1.2,2,0,2,1							
13	48,0,2,130,275,0,1,139,0,0.2,2,0,2,1							
14	49,1,1,130,266,0,1,171,0,0.6,2,0,2,1							
15	64,1,3,110,211,0,0,144,1,1.8,1,0,2,1							
16	58,0,3,150,283,1,0,162,0,1,2,0,2,1							
17	50,0,2,120,219,0,1,158,0,1.6,1,0,2,1							
18	58,0,2,120,340,0,1,172,0,0,2,0,2,1							

Data presentation



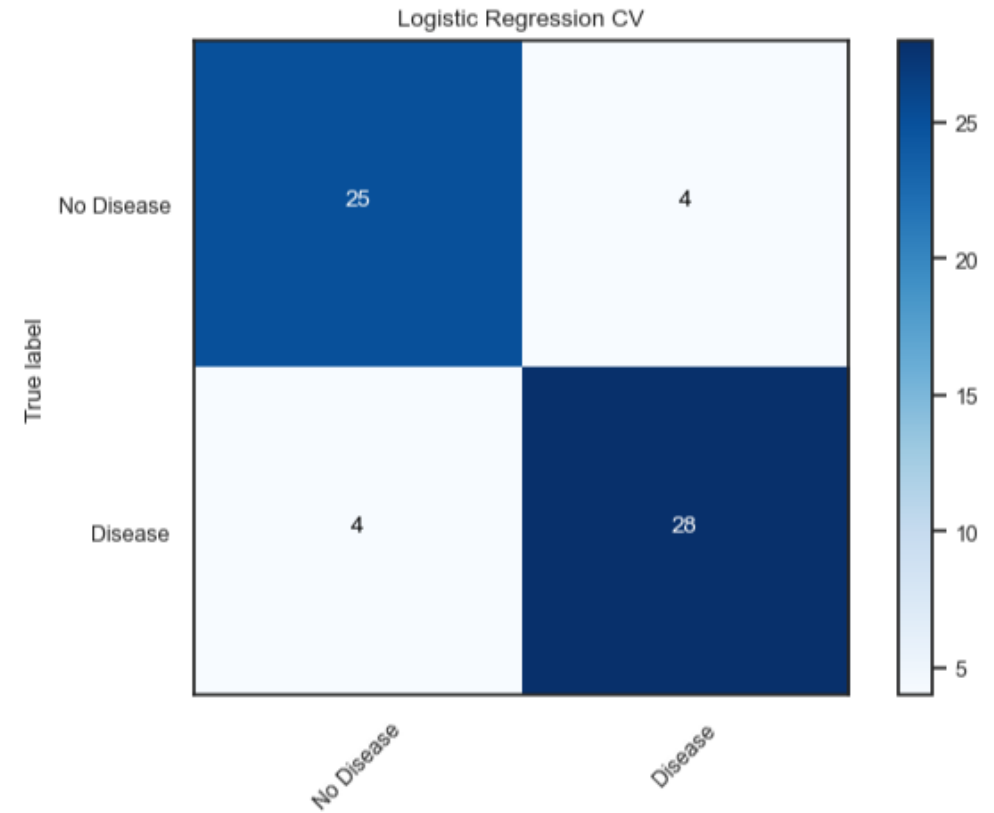
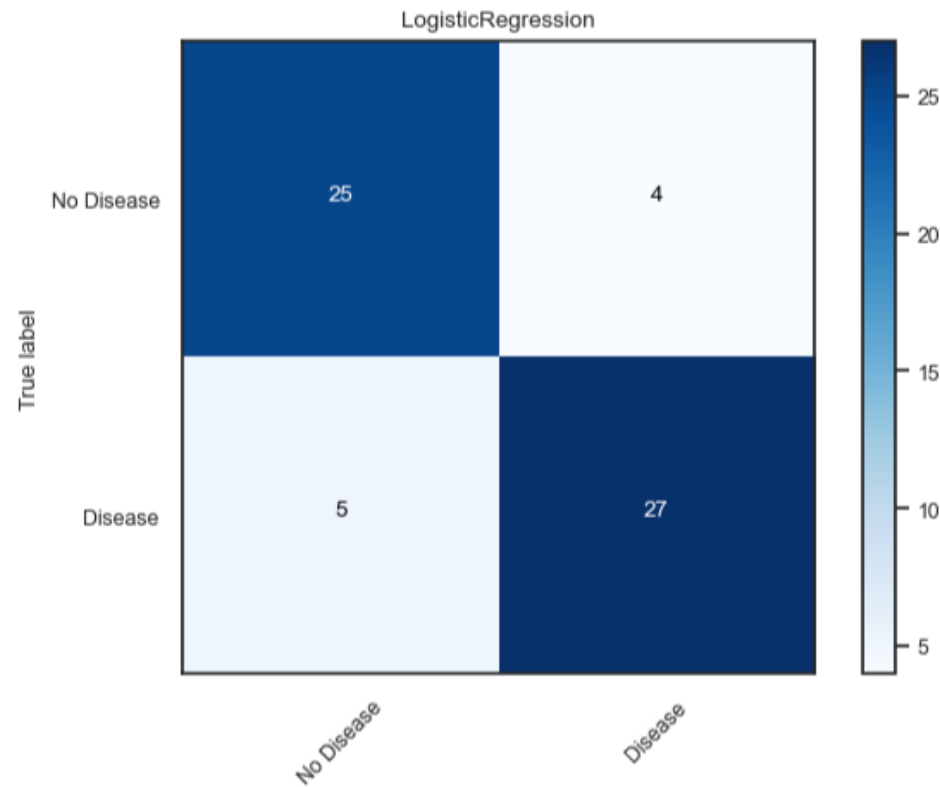
Models

Logistic Regression - statistical model used for binary classification, where it predicts the probability of an instance belonging to a certain class based on input features. It utilizes a logistic function (sigmoid) to map the output to a probability value between 0 and 1.

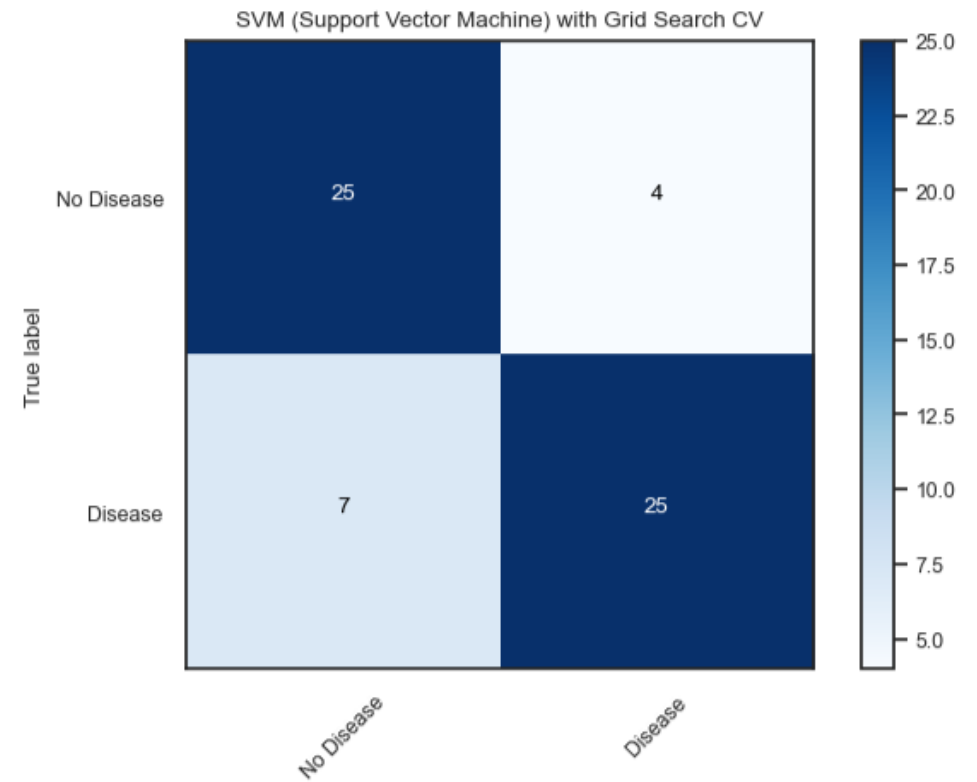
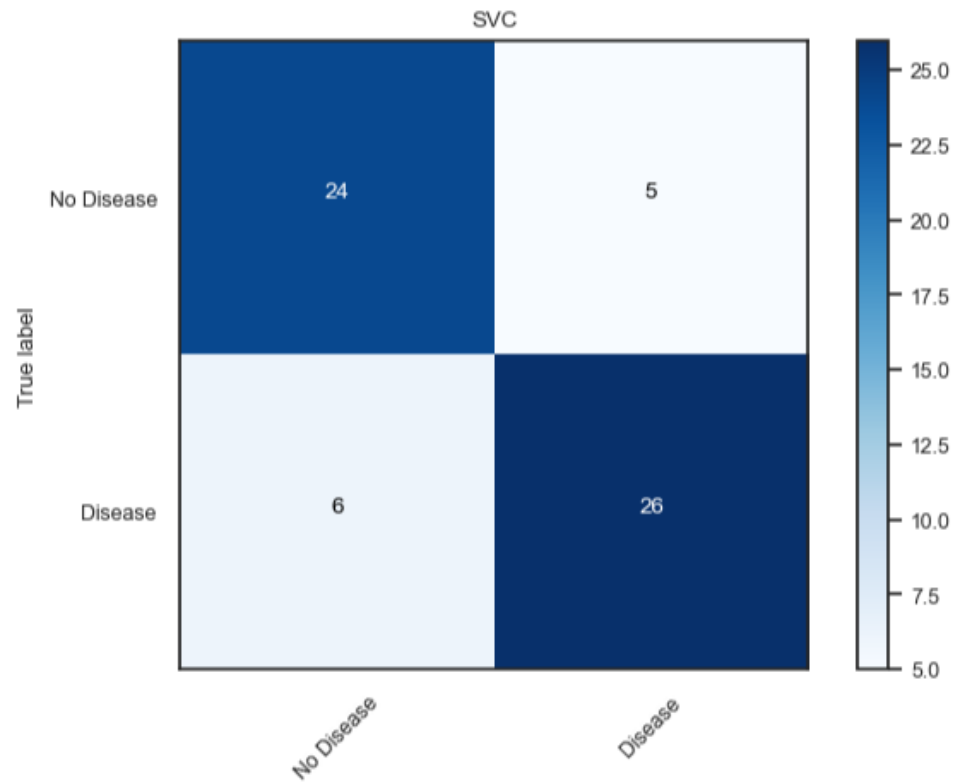
Support Vector Classifier - is a machine learning model that is commonly used for classification tasks. It aims to find an optimal hyperplane in a high-dimensional space that best separates different classes of data points, maximizing the margin between them.

Decision Trees - A decision tree is a model that recursively partitions the data into subsets based on feature values, making predictions at the leaf nodes. It uses a series of if-else conditions to traverse the tree and determine the final outcome.

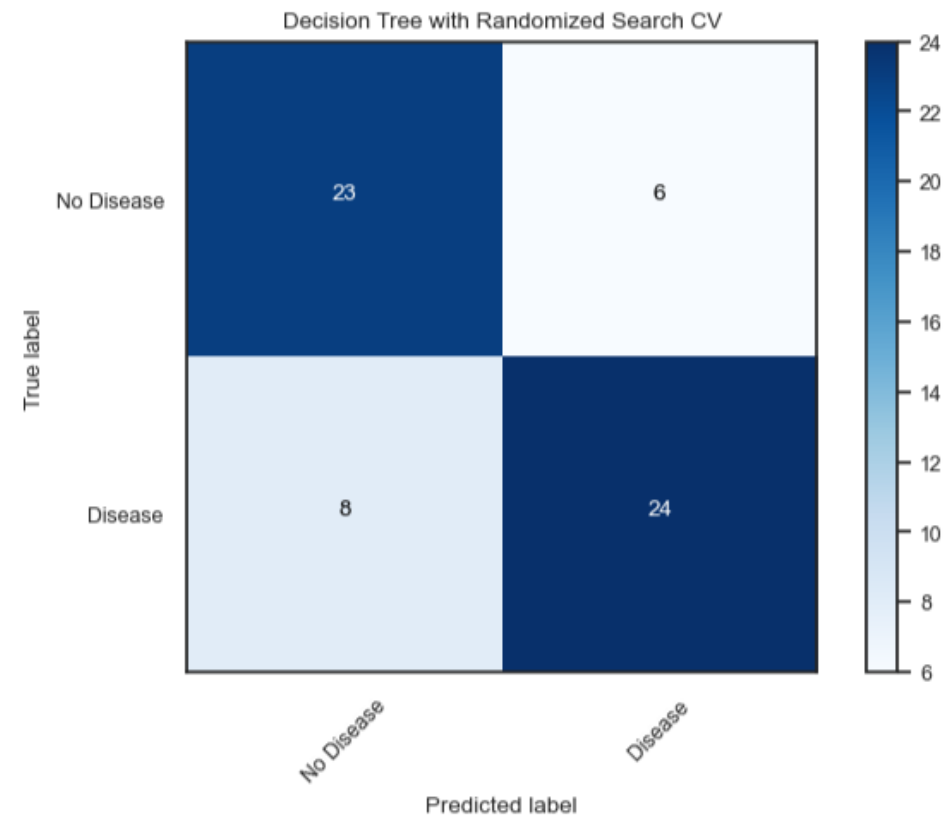
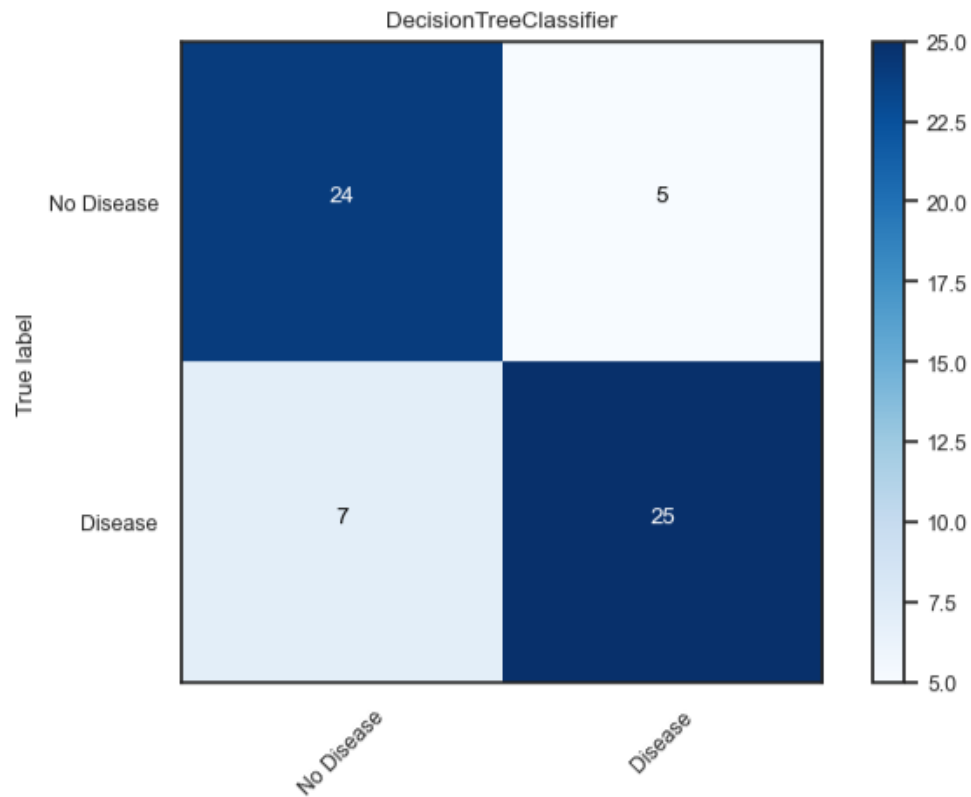
Logistic Regression



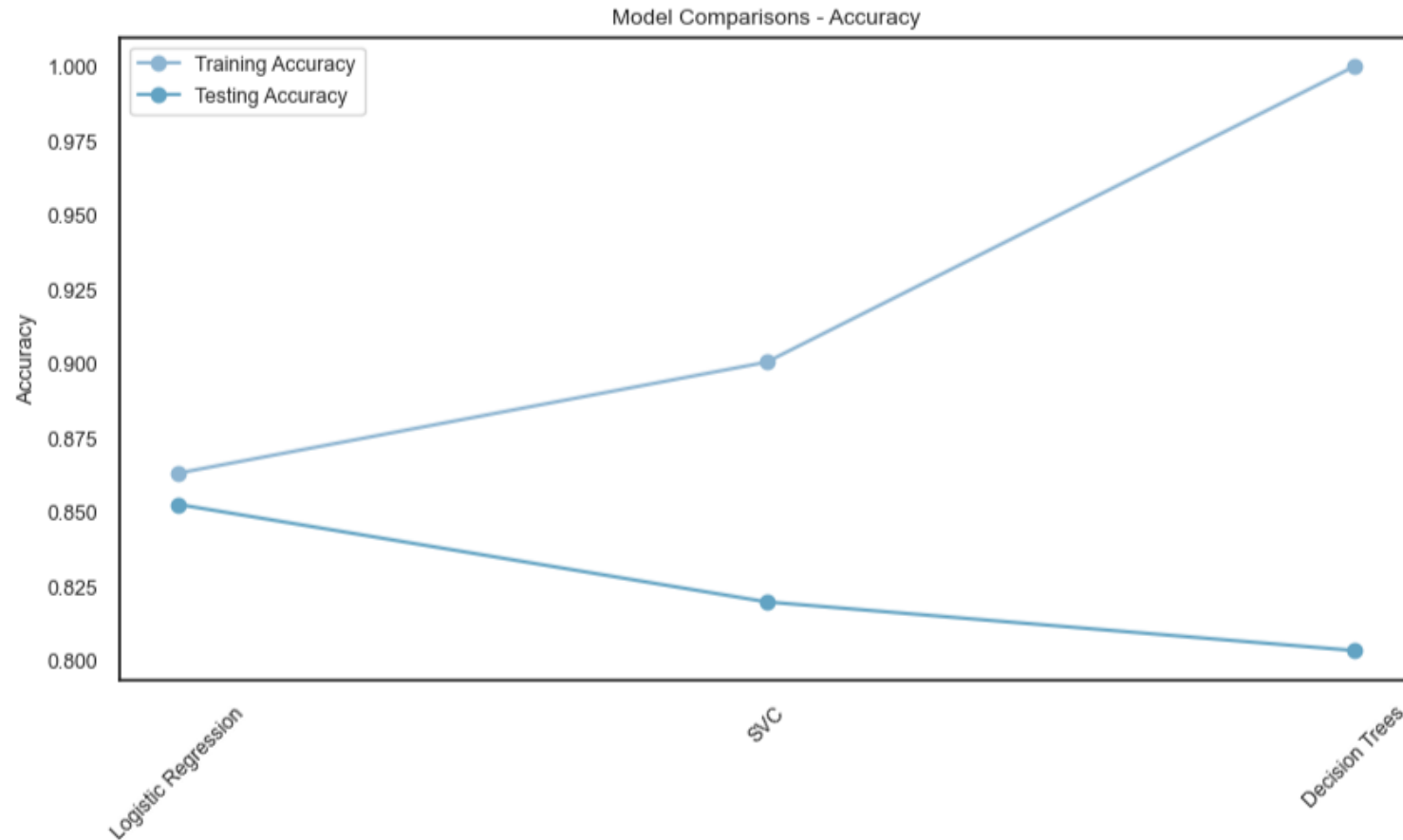
Support Vector Classifier



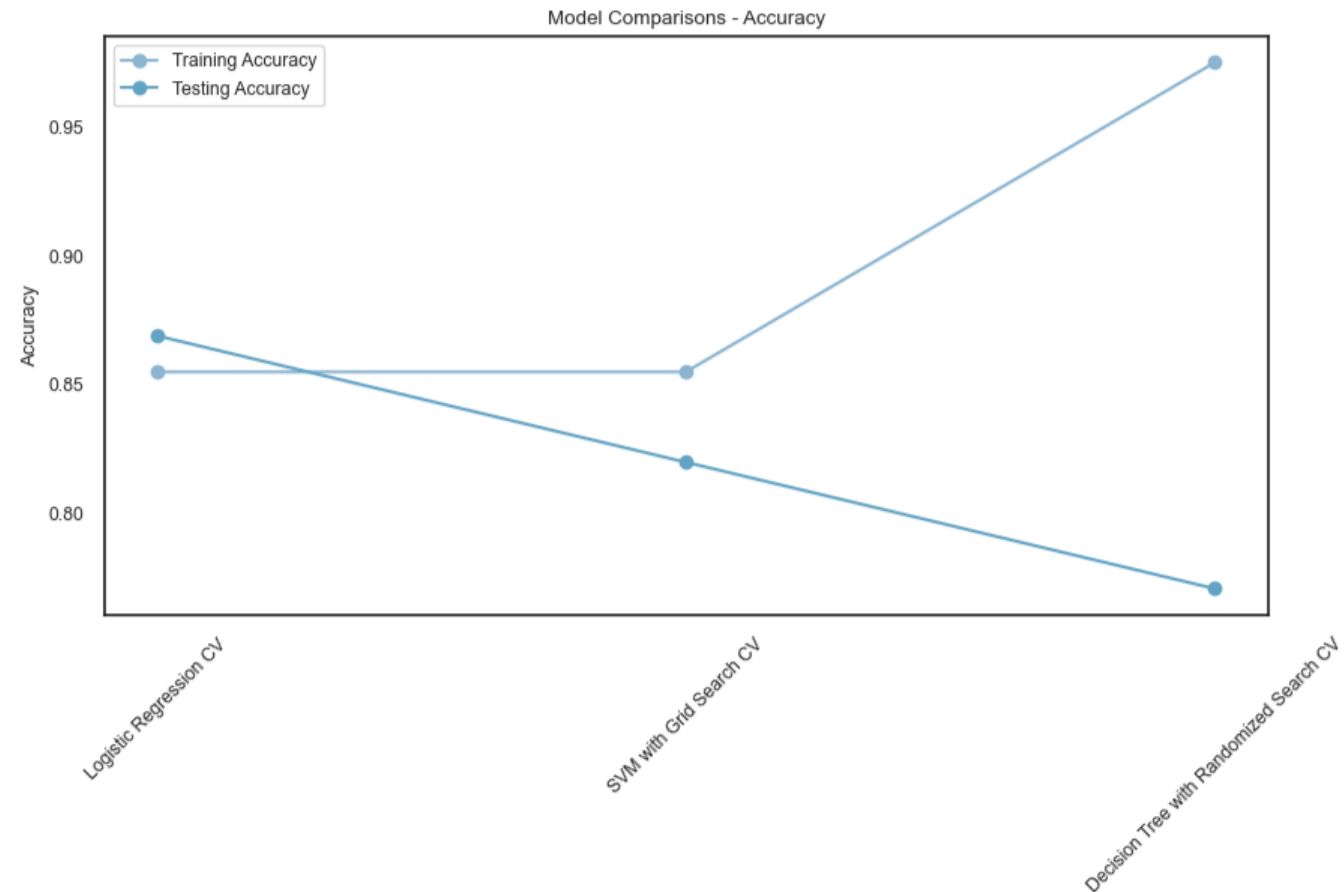
Decision Tree Classifier



Output analysis



Output analysis



Output analysis

Logistic Regression : This model had a training accuracy of 86% and a testing accuracy of 85%. This indicates that the model generalized well on unseen data and didn't overfit the training set.

Support Vector Classifier (SVC) : The SVC model had a training accuracy of 90% and a testing accuracy of 82%. The model performed well on the training set but didn't generalize as well on the test set.

Decision Tree Classifier : The model achieved a perfect training accuracy of 100%, but the testing accuracy dropped to 80%. This discrepancy suggests that the decision tree model overfitted the training data.

Conclusions

After training three distinct machine learning models, we can conclude that the data is reliable, but with slight variations depending on the chosen model. Although the training and testing accuracies differed among the models, the Linear Regression model exhibited the highest level of accuracy in both phases.

Source

www.kaggle.com

<https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a>

https://en.wikipedia.org/wiki/Machine_learning

<https://www.javatpoint.com/logistic-regression-in-machine-learning>

<https://pythonprogramming.net/linear-svc-example-scikit-learn-svm-python/>

<https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>

Thank you for the attention

JAKUB ŁAPAJ

ADRIAN PIRÓG