MINI PROJECT : Heart Attack

*By Yazan Hijazi & Saleh Khalaf*

Project Description: Heart Attack Prediction using Decision Tree and Logistic Regression Models.

This project focuses on predicting the occurrence of heart attacks by utilizing a heart attack dataset. The analysis compares two machine learning models, namely the Decision Tree model and the Logistic Regression model. Using KNIME platform.

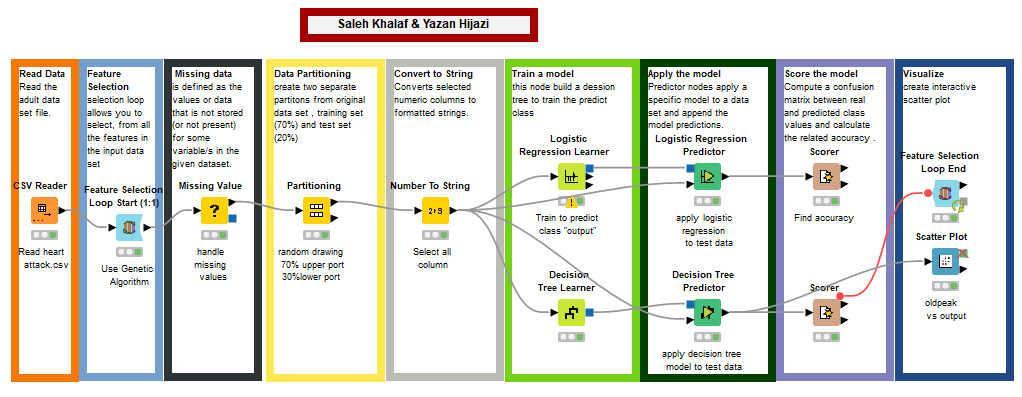
The heart attack dataset contains various features and attributes related to individuals who have experienced heart attacks. These features include age, gender, cholesterol levels, blood pressure, and other relevant medical indicators. The dataset serves as the foundation for training and evaluating the predictive models.

The first model employed in this project is the Decision Tree model. Decision trees are popular machine learning algorithms capable of handling both categorical and numerical data. By constructing a tree-like structure, the model partitions the dataset based on the available features and predicts the occurrence of heart attacks.

The second model used in this project is the Logistic Regression model. Logistic regression is a statistical technique commonly applied to predict binary outcomes. It examines the relationship between the x (output) and a y (prediction (output)).

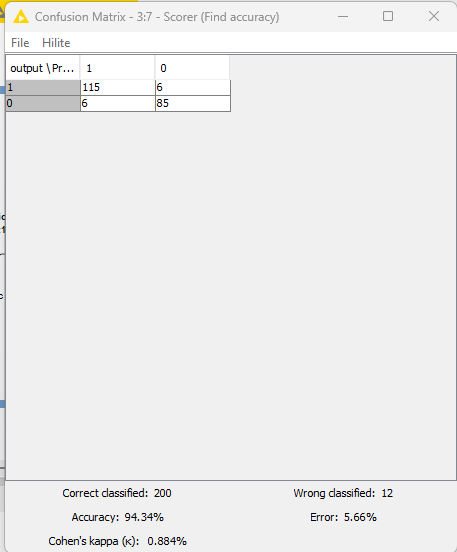
* The Workflow

The workflow involves loading the dataset with a CSV Reader, partitioning the data, handling missing values, selecting informative features, converting relevant variables, training the models on the training set, using the Predictor node to generate predictions on the testing set, and evaluating the models' performance using the Scorer node. These steps collectively enable accurate heart attack prediction and help identify the most effective model for this task.



* Logistic Regression model

After evaluating the Logistic Regression model on the heart attack dataset, an impressive accuracy of 94.43% was achieved. This indicates that the model accurately predicted the likelihood of heart attacks for a significant majority of the cases in the testing set. A high accuracy rate suggests that the model successfully captured the underlying patterns and relationships between the input variables and the occurrence of heart attacks. Such a high level of accuracy is promising and indicates the potential effectiveness of the Logistic Regression model in predicting heart attacks. With this level of accuracy, the Logistic Regression model can be considered a valuable tool for early detection and prevention of heart attacks, providing important insights into potential risk factors and enabling timely intervention for at-risk individuals.



* Decision Tree model

The Decision Tree model achieved an accuracy of 91.038% after evaluation on the heart attack dataset. This indicates its strong predictive ability in determining the likelihood of heart attacks. The model's accuracy underscores its value as a tool for early detection and prevention, helping identify crucial risk factors and informing healthcare decisions to improve patient outcomes.

A screenshot of a computer

Description automatically generated with medium confidence

When comparing the accuracy of the Logistic Regression and Decision Tree models on the heart attack dataset, the Logistic Regression model achieved a higher accuracy of 94.43% compared to the Decision Tree model's accuracy of 91.038%. This suggests that the Logistic Regression model performed slightly better in predicting the occurrence of heart attacks based on the given dataset.

* Scatter plot

A scatter plot was used to visualize the relationship between the "oldpeak" variable and the predicted output column in the heart attack dataset. It showed a positive correlation, indicating that higher values of "oldpeak" were associated with an increased likelihood of a heart attack. This suggests that greater ST depression induced by exercise relative to rest may serve as an important risk factor in predicting heart attacks.

