

MEDICAL RISK FACTOR

REPORT

The project aims to predict disease diagnosis with over **85%** accuracy. It incorporates insights from research on handling imbalanced data to enhance predictive accuracy. The prediction is made using 'Random Forest' model.

WORK PROCESS

We started by gathering and thoroughly analyzing the data, identifying features essential for accurate forecasting while filtering out less relevant ones. After testing five models, we selected the one that delivered the best results - Random Forest and meticulously combined all the collected information to achieve the desired, precise outcome.

SELECTED MODEL

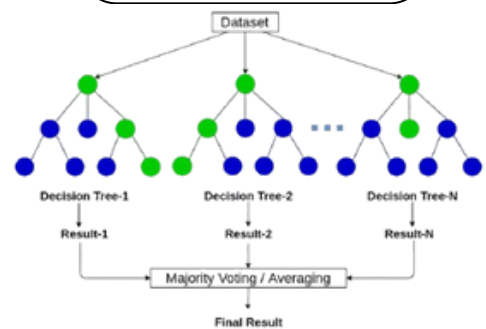
A Random Forest is an ensemble learning algorithm that builds multiple decision trees during training and combines their outputs (via averaging for regression or voting for classification) to improve accuracy and reduce overfitting.

It works by selecting random subsets of features and data samples for each tree, making it robust to noise and less prone to overfitting compared to individual decision trees.

BEST PARAMETERS

max_features ('sqrt'):
Square root selected features.
min_samples_split (2):
Minimum samples to split.
n_estimators (100):
Number of decision trees.

RANDOM FOREST



FINAL RESULT

F1_score



Supervision:
Mayya Lihovodov

Students:
Eti Perah
Matan Cohen
Bar Azarya
Aviv Meir Ovadia