



PROJECT REPORT
COMPARATIVE PERFORMANCE ANALYSIS OF
SUPPORT VECTOR MACHINE AND RANDOM FOREST
ON DIABETES PATIENT DATA FROM HOSPITALS IN THE
UNITED STATES

DARMSTATER ALBERTUS ALBERT D.
20.K1.0035

Faculty of Computer Science
Soegijapranata Catholic University
2024

ABSTRACT

The era of technological advancement at this time has begun to help a lot in many job sections, especially in the medical section. Especially in the development of Machine Learning which has a significant impact, the model built can help to predict the patient's disease from the symptoms and tests performed. Therefore, high accuracy and a short time are required for the machine-learning model to be built. The author build a model using the Random Forest algorithm and SVM algorithm, then compare these two models. What is compared between these two models is the computation time required by each algorithm and the level of accuracy, precision, recall, and F1-Score with stepwise data usage. The result to be achieved is that one of the algorithms produces stable and maximum results with the existing data. Among the eight experiments, SVM showed better performance in experiments 1, 3, 4, and 6, while random forest showed better performance in experiments 1, 2, 3, 4, and 6. The sixth experiment yielded the highest accuracy on both the minimum and maximum datasets. Here, SVM achieved 99.992 curacy in 142.0238 seconds and Random Forest achieved 99.982 curacy in 8.045849 seconds. Random Forest was 133.977951 seconds faster, but had a slightly lower accuracy of 0.01%.

Keywords : Random Forest, Support Vector Machine, Machine Learning, Diabetes Disease