



PROJECT REPORT

COMPARISON OF PREDICTION ALGORITHMS FOR PREDICTING DIABETES USING ORANGE DATA MINING

JUAN ANDREAS SIHALOHO
18.K1.0074

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT

Diabetes is one of the diseases with the highest mortality rate and has experienced a very significant increase in the number of sufferers. According to the IDF Diabetes Atlas (2021) from International Diabetes Federation, 10.5% of the adult population (20-79 years) has diabetes, with almost half unaware that they are living with the condition.

With machine learning technology, we can create models that can later predict whether someone has diabetes or not so they can treat it immediately. By using simplified data from The Behavioral Risk Factor Surveillance System (BRFSS) through surveys and uploaded to Kaggle, we can train models with various algorithms which can then predict whether someone has diabetes or not. There are 4 algorithms that are used in this study, namely Naive Bayes, Logistic Regression, Gradient Boosting, and Neural Network. The results of this study show that these algorithms can predict whether someone has diabetes or not with an accuracy of up to 74%.

Keyword: Diabetes, Neural Network, Gradient Boosting, Naive Bayes, Logistic Regression

