

## ABSTRACT

The problem that occurs in the application of K-Nearest Neighbors as a classification algorithm is the frequent occurrence of overfitting in data processing. This can be overcome by using cross-validation techniques in evaluating the algorithm model and minimizing overfitting. Then the performance of diabetes prediction accuracy is unknown using the K-Nearest Neighbors algorithm with cross-validation technique. The data used comes from the National Institute of Digestive and Kidney Diabetes in 2021. The case study in this study is to find out the initial screening for diabetes is supported by the results of algorithm accuracy and real time application of streamlit-based users. The purpose of this study was to optimize the accuracy results with a cross validation technique supported by the k-nearest neighbors algorithm in the study of diabetes data. The method used is the k-nearest neighbors algorithm which is supported by cross validation technique for optimal accuracy results. Then the application of a streamlit-based interactive web application for testing the accuracy results used by the user to see the probability that the user has diabetes. The results showed that the optimization of the Cross Validation technique supported by the K-Nearest Neighbors algorithm model worked well. The results of the confusion matrix using the cross validation technique are more accurate in terms of the advantages of using the cross-validation technique itself. So that the classification report which has a value of 95% is more accurate than the accuracy which is worth 92% because of the use of cross-validation techniques that can minimize overfitting in addition to considerations of the accuracy value and the implementation of streamlit-based interactive web applications for user testing is going well.

*Keyword: Optimization, Machine Learning , Algorithm, K-Nearest Neighbors, Diabetes, Cross Validation, Streamlit*