

## ABSTRACT

One of the problems faced by healthcare facilities is the influx of patients every day, resulting in long queues. This ultimately leads to a decrease in public satisfaction with healthcare services. Additionally, medical record data continues to grow every day but is not optimally utilized, even though if processed properly, it can provide valuable information. This study proposes a solution to these problems by developing a disease diagnosis classification model that will receive symptoms experienced by patients and then display the classification results of the disease diagnosis. Data mining is used in this research with the CRISP-DM methodology. This study compares the classification results using the Naive Bayes and KNN algorithms combined with Bag of Words and TF-IDF for text preprocessing. The results show that the KNN algorithm and Bag of Words produce the best results with an accuracy score of 75%, while the Naive Bayes algorithm and Bag of Words yield an accuracy score of 74%. The Naïve Bayes and TF-IDF and The KNN and TF-IDF yield an accuracy score of 73%. This prediction system is expected to provide doctors with an overview of patient illnesses, making examinations more efficient.

**Keywords** : data mining, k-nearest neighbor, medical record, naïve bayes