

ABSTRAK

Penyakit *liver* merupakan salah satu penyakit yang banyak terjadi di negara maju ataupun negara berkembang. *Liver* merupakan organ yang berperan penting di dalam tubuh manusia, antara lain membersihkan darah dari senyawa berbahaya hingga membantu melakukan proses pencernaan makanan. Diagnosa penyakit *liver* sering sulit dilakukan pada penderita stadium awal diakibatkan oleh beberapa faktor antara lain penindakan yang terlambat akibat minimnya analisis terhadap informasi data penyakit *liver*. Sehingga dibutuhkan *data mining* untuk mengidentifikasi apakah pasien tersebut memiliki penyakit atau tidak, serta dapat mendiagnosa gejala penyakit *liver* secara dini pada pasien dalam waktu singkat. Pada penelitian ini menggunakan metode *Modified K-Nearest Neighbor* dengan seleksi fitur *F-Score* untuk meminimalisir atribut yang kurang relevan dengan mengurangi jumlah fitur dan juga menghitung tingkat akurasi dari penerapan metode tersebut. Dataset diambil dari *web Kaggle Repository* yang memiliki 10 fitur dan 1 kelas yang memiliki 2 target yaitu 1 dan 2 (1 untuk pasien dengan penyakit *liver* dan 2 untuk tidak ada penyakit) dengan jumlah total 583 data yang akan di *SMOTE* menjadi 832 data supaya data *minority* seimbang dengan data *majority*, persentase pembagian data 80% *training* dan 20% *testing*. Berdasarkan hasil penelitian menggunakan metode *Modified K- Nearest Neighbor* dengan 6 fitur terbaik dan nilai $k = 9$ menghasilkan akurasi 91,16%.

Kata Kunci: *Data mining, Modified K-Nearest Neighbor, Penyakit Liver, F-Score, SMOTE*

ABSTRACT

Liver disease is one of the most common diseases in developed and developing countries. The liver is an organ that plays an important role in the human body, including cleaning the blood from harmful compounds and helping to process food digestion. Diagnosis of liver disease is often difficult in patients with early stages due to several factors, including delays in treatment due to the lack of analysis of liver disease data. So that data mining is needed to identify whether the patient has the disease or not, and to be able to diagnose symptoms of liver disease early in the patient in a short time. In this study using the Modified K-Nearest Neighbor method with F-Score feature selection to minimize irrelevant attributes by reducing the number of features and also calculating the level of accuracy of applying the method. The dataset is taken from the Kaggle Repository web which has 10 features and 1 class which has 2 targets namely 1 and 2 (1 for patients with liver disease and 2 for no disease) with a total of 583 data that will be SMOTE to 832 data so that the data is the minority balanced with majority data, the percentage of data sharing is 80% training and 20% testing. Based on the results of research using the Modified K-Nearest Neighbor method with the 6 best features and a value of $k = 9$, it produces an accuracy of 91.16%.

Keywords: Data mining, Modified K-Nearest Neighbor, Liver Disease, F-Score, SMOTE