

## ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan tingkat kecanduan rokok pada mahasiswa Universitas Malikussaleh dengan menggunakan algoritma C4.5. Perilaku merokok merupakan permasalahan yang dapat berdampak negatif terhadap kesehatan, konsentrasi belajar, dan kualitas hidup mahasiswa. Data penelitian diperoleh dari hasil pengisian kuesioner serta *input* data secara langsung oleh pengguna melalui sistem yang dikembangkan. Total data yang digunakan berjumlah 310 data mahasiswa, yang selanjutnya dibagi menjadi 240 data sebagai data pelatihan (*training*) dan 60 data sebagai data pengujian (*testing*). Dataset disusun berdasarkan beberapa atribut, yaitu jumlah batang rokok per hari, lama merokok, alasan merokok, dan waktu mulai merokok. Hasil klasifikasi menunjukkan bahwa sebanyak 98 mahasiswa termasuk dalam kategori kecanduan ringan, 150 mahasiswa pada kategori sedang, 58 mahasiswa pada kategori berat, dan 4 mahasiswa pada kategori sangat berat. Mayoritas mahasiswa berada pada tingkat kecanduan sedang, yang menunjukkan bahwa kebiasaan merokok di kalangan mahasiswa masih cukup tinggi. Proses pembentukan model dilakukan melalui perhitungan nilai *Entropy*, *Information Gain*, *Split Info*, dan *Gain Ratio*, di mana atribut jumlah konsumsi rokok per hari terpilih sebagai atribut akar karena memiliki nilai *Gain Ratio* tertinggi sebesar 0,2717. Model klasifikasi yang dihasilkan memiliki tingkat akurasi sebesar 80%, dengan nilai *precision* 64,5%, *recall* 56,8%, dan *F1-score* 58,9%. Selain itu, sistem juga dilengkapi fitur *input* data mandiri dan notifikasi hasil klasifikasi melalui email menggunakan *MailHog*, sehingga pengguna dapat mengetahui tingkat kecanduan rokok secara langsung. Hasil penelitian ini menunjukkan bahwa algoritma C4.5 efektif digunakan dalam mengklasifikasikan tingkat kecanduan rokok mahasiswa.

Kata Kunci : Klasifikasi, C4.5, Rokok, Pohon Keputusan, Mahasiswa

## **ABSTRACT**

*This study aims to classify the level of cigarette addiction among Malikussaleh University students using the C4.5 algorithm. Smoking behavior is a problem that can have a negative impact on students' health, concentration, and quality of life. The research data was obtained from questionnaires and direct data input by users through the developed system. A total of 310 student data were used, which were then divided into 240 data as training data and 60 data as testing data. The dataset was compiled based on several attributes, namely the number of cigarettes per day, duration of smoking, reasons for smoking, and time of starting smoking. The classification results showed that 98 students were in the mild addiction category, 150 students in the moderate category, 58 students in the severe category, and 4 students in the very severe category. The majority of students were at a moderate level of addiction, indicating that smoking habits among students are still quite high. The model was developed by calculating Entropy, Information Gain, Split Info, and Gain Ratio values, with the attribute of daily cigarette consumption selected as the root attribute because it had the highest Gain Ratio value of 0.2717. The resulting classification model has an accuracy rate of 80%, with a precision value of 64.5%, recall of 56.8%, and an F1-score of 58.9%. In addition, the system is equipped with a self-input data feature and classification result notifications via email using MailHog, so that users can find out their level of cigarette addiction directly. The results of this study indicate that the C4.5 algorithm is effective in classifying the level of cigarette addiction among students.*

*Keywords: Classification, C4.5, Smoking, Decision Tree, Students*