

## ABSTRAK

Sistem prediksi memainkan peran penting pada zaman sekarang. Sistem prediksi banyak digunakan saat ini untuk memberikan rekomendasi kepada pengguna berdasarkan preferensi pengguna. Salah satunya ialah untuk melakukan prediksi terkait dengan minat studi lanjut. Masih banyak yang melanjutkan studi tidak sesuai minat belajarnya, karena paksaan atau bahkan masih tidak mengetahui secara pasti minat studinya. Oleh karena itu, tujuan dari penelitian ini adalah untuk memberi rekomendasi minat studi berdasarkan preferensi pengguna. Penelitian ini menggunakan metodologi pengembangan *CRISP-DM* dan juga Algoritma K-Nearest Neighbor dengan k-value 24 maka dihasilkan akurasi model sebesar 73% dengan *distance metric Manhattan* menggunakan data training sebanyak 250 sampel data training dengan masing-masing data testing sebanyak 300 sampel data label (Academics), 100 sampel data label (Arts), 155 sampel data label (Sports). Model dibungkus dengan *pickle* untuk diintegrasikan dengan bagian *interface* sistem melalui *Flask API*, Sehingga *user* dapat menggunakan aplikasi dengan menjawab pertanyaan yang disediakan sistem secara mudah.

**Kata Kunci:** Sistem Prediksi, *Confusion matrix*, *Data Training*, *Data Testing*  
*Algoritma K-Nearest Neighbor*, *k- value*.

## **ABSTRACT**

*Prediction systems play an important role in this day and age. Prediction systems are widely used today to provide recommendations to users based on user preferences. One of them is to make predictions related to interest in further studies. There are still many who continue their studies not according to their learning interests, because of coercion or even they still do not know for sure their study interests. Therefore, the purpose of this study is to provide recommendations for study interest based on user preferences. This study uses the CRISP-DM development methodology as well as the K-Nearest Neighbor Algorithm with a k-value of 24, resulting in a model accuracy of 73% with the Manhattan distance metric using training data of 250 training data samples with each testing data of 300 samples of label data. (Academics), 100 label data samples (Arts), 155 label data samples (Sports). The model is wrapped in a pickle to be integrated with the system interface via the Flask API, so that users can use the application easily.*

**Keywords: Prediction System, Confusion matrix, Data Testing, Data Testing, Algoritma K-Nearest Neighbor, k- value.**

