

Dalam pelaksanaan penelitian dan penyusunan tugas akhir berupa skripsi, tesis, ataupun disertasi, dosen pembimbing menjadi peran yang krusial agar hasil penelitian yang disusun dapat maksimal. Pada DTETI FT UGM, khususnya program magisternya memberikan kesempatan bagi mahasiswa dalam menentukan dosen pembimbingnya. Namun, hal ini menjadi kesulitan tersendiri bagi mahasiswa dalam menentukan dosen pembimbing yang sesuai dengan topik yang diajukan. Penggunaan teknologi khususnya pada pengembangan model *machine learning* dapat menjadi salah satu solusi yang dapat dimanfaatkan untuk melakukan rekomendasi dosen pembimbing tugas akhir yang relevan dengan topik mahasiswa.

Tujuan dari penelitian ini adalah membangun alat rekomendasi berupa model *machine learning* yang dapat merekomendasikan dosen pembimbing tugas akhir di DTETI FT UGM berdasarkan topik mahasiswa. Model rekomendasi yang dibangun tersebut dilatih dan diuji berdasarkan data judul publikasi dosen yang bersumber dari *platform* Google Scholar. Metode yang digunakan dalam membangun model *machine learning* tersebut adalah Naive Bayes, ANN, dan LSTM yang kemudian dievaluasi satu sama lain menggunakan *5-fold cross validation*.

Tahapan penelitian ini terdiri dari pengumpulan judul publikasi dosen dengan teknik *scraping*, prapemrosesan teks, *oversampling*, ekstraksi fitur dengan TF-IDF, implementasi model *machine learning*, dan evaluasi model. Judul publikasi yang digunakan sebanyak 1474 data dari 3208 data yang berhasil dikumpulkan dari 20 dosen DTETI FT UGM. Berdasarkan evaluasi menggunakan *5-fold cross validation*, model ANN memiliki performa terbaik dengan akurasi sebesar 82%, yang lebih tinggi sekitar 9-10% dibandingkan dengan model lainnya. Model ANN juga memiliki nilai *recall*, *precision*, dan *F1-score* yang lebih baik. Pada pengujian dengan data uji, model ANN menunjukkan hasil evaluasi yang tidak berbeda jauh, yakni diperoleh akurasi sebesar 85,26%, *recall* 85,35%, *precision* 85,25%, *F1-Score* 84,74%, serta pada waktu pelatihan dan waktu prediksi dibutuhkan masing-masing selama 21 detik dan 0,12 detik.

Kata kunci: *Web Scraping*, *Machine Learning*, Naive Bayes, *Artificial Neural Network* (ANN), *Long Short-Term Memory* (LSTM).

ABSTRACT

In the process of conducting research and writing a thesis, the role of a supervisor is crucial in ensuring the quality of the undertaken research. Particularly in the master's program at DTETI FT UGM, students have the privilege of selecting their own thesis supervisor. However, this presents a challenge for students in identifying a suitable supervisor aligned with their chosen research topic. Consequently, the need arises for a recommendation tool that can provide relevant supervisor suggestions based on the student's thesis topic. One potential solution is the utilization of a machine learning model.

This research aims to develop a machine learning-based recommendation tool capable of suggesting thesis supervisors at DTETI FT UGM based on the student's specific topic. The recommendation model is trained and tested using data acquired from the titles of faculty publications obtained through web scraping from the Google Scholar platform. Naive Bayes, ANN, and LSTM methods are employed to construct the machine learning model, which is subsequently evaluated using 5-fold cross-validation.

The research process consists of collecting lecturers' publication titles through web scraping, text preprocessing, oversampling, feature extraction using TF-IDF, implementing the machine learning models, and evaluating the models. A total of 1474 publication titles were utilized from the collected dataset comprising 3208 data points from 20 lecturers in the DTETI FT UGM. Based on the evaluation using 5-fold cross-validation, the ANN model demonstrated the best performance with an accuracy of 82%, which was approximately 9-10% higher compared to the other models. The ANN model also exhibited superior recall, precision, and F1-score values. In the testing phase using unseen data, the ANN model displayed comparable evaluation results, achieving an accuracy of 85,26%, recall of 85,35%, precision of 85,25%, and an F1-score of 84,74%. The training and prediction times for the ANN model were 21 seconds and 0.12 seconds, respectively.

Keywords : Web Scraping, Machine Learning, Naive Bayes, Artificial Neural Network (ANN), Long Short-Term Memory (LSTM)