

ABSTRAK

Skripsi merupakan salah satu persyaratan akhir dalam pendidikan akademis mahasiswa yang harus ditulis sesuai dengan bidang ilmunya. Proses awal dalam pembuatan skripsi adalah pengajuan judul, di mana mahasiswa harus memastikan bahwa judul yang diajukan tidak mirip dengan judul skripsi sebelumnya untuk menghindari *plagiarisme*. Deteksi kemiripan judul sangat penting untuk mencegah plagiarisme dan meningkatkan akurasi sistem rekomendasi dengan mengidentifikasi judul yang relevan. Dengan kemajuan teknologi informasi, tantangan seperti plagiarisme menjadi lebih kompleks, memerlukan sistem yang efisien untuk mengecek kemiripan judul. Salah satu metode yang digunakan adalah *Support Vector Machine* (SVM), sebuah teknik pembelajaran mesin yang efektif dalam klasifikasi teks, termasuk judul skripsi, berkat kemampuannya dalam menangani hubungan *non-linear* melalui kernel. Penggunaan *Support vector machine* dalam deteksi kemiripan judul bertujuan untuk memastikan keakuratan dan efisiensi dalam mengidentifikasi potensi *plagiarisme*. Kombinasi metode *Support Vector Machine* (SVM) dan *Natural Language Processing* (NLP), khususnya melalui vektorisasi menggunakan TF-IDF, telah terbukti efektif dalam mendeteksi kemiripan judul skripsi. Proses ini melibatkan pengumpulan data dari database judul_skripsi, diikuti dengan preprocessing yang meliputi tokenisasi, penghapusan kata-kata tidak relevan, dan stemming. Data yang bersih dan terstruktur ini kemudian diubah menjadi representasi numerik melalui TF-IDF, yang memberikan bobot lebih tinggi pada kata-kata yang spesifik namun signifikan. Model SVM yang dilatih dengan data ini berhasil mendeteksi kemiripan dengan tingkat akurasi yang memadai.

Kata kunci : Deteksi Kemiripan.,Judul Skripsi., SVM., NLP, Klasifikasi Teks., Pembelajaran Mesin

ABSTRACT

A thesis is one of the final requirements in a student's academic education that must be written in accordance with their field of study. The initial process in writing a thesis is submitting a title, where students must ensure that the title submitted is not similar to the title of a previous thesis to avoid plagiarism. Detecting title similarity is very important to prevent plagiarism and improve the accuracy of the recommendation system by identifying relevant titles. With the advancement of information technology, challenges such as plagiarism have become more complex, requiring an efficient system to check title similarity. One of the methods used is the Support Vector Machine (SVM), a machine learning technique that is effective in text classification, including thesis titles, thanks to its ability to handle non-linear relationships through kernels. The use of Support Vector Machines in title similarity detection aims to ensure accuracy and efficiency in identifying potential plagiarism. The combination of the Support Vector Machine (SVM) and Natural Language Processing (NLP) methods, especially through vectorization using TF-IDF, has proven effective in detecting the similarity of thesis titles. This process involves collecting data from the thesis_title database, followed by preprocessing including tokenization, removal of irrelevant words, and stemming. This clean and structured data is then transformed into a numeric representation through TF-IDF, which gives higher weight to specific but significant words. The SVM model trained with this data successfully detects similarities with a reasonable level of accuracy.

Keywords: *Similarity Detection., Thesis Title., SVM., NLP, Text Classification., Machine Learning*