

ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan memprediksi popularitas artikel berita di *Mashable* menggunakan metode klasifikasi *Vector Machine (SVM)* dalam data mining. Kumpulan data "Online News Popularity" dari Repositori Pembelajaran Mesin *UCI* digunakan sebagai dasar penelitian, mencakup 39.644 sampel dengan 61 karakteristik termasuk berbagai faktor seperti keterbacaan kata, keterbacaan gambar, kategori artikel, dan metrik sosial lainnya. Proses preprocessing data meliputi pembersihan, transformasi dan pemilihan fitur menggunakan metode *Recursive Feature Elimination (RFE)*. Model *SVM* dilatih menggunakan data pelatihan dan dioptimalkan oleh *Grid Search* dengan validasi silang *5-fold*. Pengujian model menunjukkan performa luar biasa dengan akurasi 89,3% Ini menunjukkan bahwa model memiliki kinerja yang cukup baik dalam memprediksi, presisi 89,3% prediksi menunjukkan kinerja yang kuat dan positif benar, *recall* 100% recall yang sempurna ini menunjukkan bahwa model berhasil menemukan semua artikel yang benar-benar populer, dan *F1-Score* 94,4% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara presisi dan recall. Hasil penelitian menunjukkan bahwa faktor-faktor seperti jumlah kata pada judul, jumlah gambar, dan kategori artikel berpengaruh terhadap popularitas artikel. Penerapan model ini secara efektif dapat membantu perusahaan media mengembangkan strategi konten yang lebih efektif dan relevan bagi pembaca. Meskipun penelitian ini memiliki keterbatasan, namun hasilnya memberikan landasan yang kuat untuk penelitian lebih lanjut dan penerapan praktis dalam industri media.

Kata Kunci: Support Vector Machine (SVM), Klasifikasi, Data Mining, Preprocessing Data, Prediksi popularitas

ABSTRACT

This study aims to analyze and predict the popularity of news articles on Mashable using the Support Vector Machine (SVM) classification method in data mining. The "Online News Popularity" dataset from the UCI Machine Learning Repository is used as the basis for the research, comprising 39,644 samples with 61 features, including various factors such as word readability, image readability, article categories, and other social metrics. The data preprocessing steps include cleaning, transformation, and feature selection using the Recursive Feature Elimination (RFE) method. The SVM model is trained using training data and optimized through Grid Search with 5-fold cross-validation. The model testing demonstrates outstanding performance with an accuracy of 89.3%. This indicates the model has strong predictive capabilities, with a precision of 89.3% showing robust and correct predictions, a recall of 100% indicating the model successfully identifies all truly popular articles, and an F1-Score of 94.4% reflecting an excellent balance between precision and recall. The results of the study indicate that factors such as the number of words in the title, the number of images, and the article category significantly influence article popularity. The application of this model can effectively help media companies develop more effective and relevant content strategies for readers. Although this study has certain limitations, its findings provide a solid foundation for further research and practical applications in the media industry.

Keyword: Support Vector Machine (SVM), Classification, Data Mining, Data Pre-processing, Popularity prediction