

ABSTRAK

Perkembangan media digital menyebabkan jumlah berita olahraga meningkat secara signifikan sehingga diperlukan sistem klasifikasi otomatis untuk membantu pengelompokan informasi secara cepat dan akurat. Penelitian ini bertujuan untuk membangun dan mengevaluasi sistem klasifikasi teks berita olahraga menggunakan algoritma *Decision Tree* dengan pendekatan *Classification and Regression Tree (CART)*. Dataset yang digunakan berasal dari Kaggle dan berisi 148.143 judul berita olahraga berbahasa Inggris dengan berbagai kategori, antara lain *Football*, *Basketball*, *Badminton*, *MMA*, dan *Formula 1*.

Tahapan penelitian meliputi preprocessing teks yang terdiri dari *case folding*, penghapusan karakter *non-alfabet*, *tokenisasi*, *stopword removal*, dan *stemming*. Selanjutnya, data teks direpresentasikan dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*. Dataset kemudian dibagi menjadi data *training*, *validation*, dan *testing* dengan beberapa skenario rasio pembagian data. Model *Decision Tree* dilatih menggunakan berbagai kombinasi parameter *criterion* (*gini* dan *entropy*), *min_samples_leaf*, dan *max_depth* untuk memperoleh konfigurasi optimal. Evaluasi performa model dilakukan menggunakan metrik *akurasi*, *precision*, *recall*, *F1-score*, serta *confusion matrix*.

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* mampu mengklasifikasikan teks berita olahraga dengan performa yang baik. Konfigurasi terbaik diperoleh pada parameter *criterion = gini*, *min_samples_leaf = 1*, dan *max_depth = 1000*, yang menghasilkan akurasi tertinggi sebesar 92,75% pada data *testing*. Hasil ini menunjukkan bahwa metode *Decision Tree* efektif untuk klasifikasi teks berita olahraga berbasis *TF-IDF* dan dapat diterapkan sebagai solusi otomatis dalam pengelolaan informasi berita olahraga.

Kata Kunci : *Klasifikasi Teks, Berita Olahraga, Decision Tree, CART, TF-IDF, Machine Learnin*