

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kebutuhan akan sistem kerja yang berbasis teknologi informasi (TI) telah berkembang seiring dengan perkembangan zaman. Pengembangan, pemeliharaan, dan pengoperasian jaringan komputer, perangkat lunak, serta sistem pemrosesan dan distribusi data adalah semua bidang teknologi informasi (TI). Situasinya mirip dengan klasifikasi jurnal yang sebelumnya dilakukan secara manual. Namun, seiring berjalannya waktu, mengklasifikasi jurnal berubah secara otomatis seiring dengan perkembangan teknologi informasi. Versi digital dari berbagai karya akademik tersedia, termasuk tesis, laporan penelitian, laporan kerja praktek. Namun, fenomena ini tidak menambah jumlah informasi atau pengetahuan yang digali dari dokumen elektronik tersebut (Akromunnisa & Hidayat, 2019).

Konsep text mining digunakan sebagai klasifikasi dokumen teks, di mana dokumen yang ada di klasifikasikan menurut dokumen yang akan diproses. Dengan bantuan konsep ini, membuat artikel yang diteliti menjelaskan kategori jenis dengan memunculkan kata-kata dari artikel yang ada (Hendriyanto & Sari, 2022).

Pada penelitian terdahulu sudah ada beberapa yang mengembangkan metode ini salah satunya yaitu menghasilkan bahwa algoritma hasil *K-Nearest Neighbor* kinerja klasifikasi terbaik dicapai dengan persentase accuracy sebesar 72.91% dan precision mencapai 73,36% (Ma'rifah et al., 2020). Kemudian ada yang menunjukkan bahwa algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dapat digunakan dalam mengklasifikasi teks, Pada penelitian ini akurasi *k-Nearest Neighbor* lebih baik dengan akurasi 80% dibandingkan *Naïve Bayes* yang mendapatkan akurasi 73% (Indriani, 2020). Adapun penelitian selanjutnya membandingkan ketiga metode yaitu dari hasil perbandingan algoritma klasifikasi yang terbaik adalah algoritma *Support Vector Machine* sebesar 81%. Sedangkan *Naïve Bayes* sebesar 67%, dan *K-Nearest Neighbor* sebesar 80% (Aziz Syahputro et al., 2019).

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang yang sudah dijelaskan, maka dapat diidentifikasi permasalahannya sebagai berikut :

1. Belum adanya analisis menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* pada dataset artikel yang berjumlah lebih dari 100.

Penelitian sebelumnya hanya menggunakan 40 jurnal.

2. Akurasi yang dihasilkan pada penelitian sebelumnya baru sebesar 70%.

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menganalisis kinerja *Naïve Bayes* dan *K-Nearest Neighbor* untuk pengklasifikasian jurnal dengan dataset lebih dari 100.
2. Menghasilkan akurasi yang lebih besar dari 70%.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan dua algoritma yaitu *Naïve Bayes* dan *K-Nearest Neighbor* untuk klasifikasi jurnal dan melakukan perbandingan terhadap kedua algoritma..
2. Aplikasi menggunakan Bahasa pemrograman JavaScript dan menggunakan HTML CSS.

1.5 Kontribusi

Hasil penelitian ini diharapkan dapat berguna untuk pengklasifikasian, khususnya dalam mengklasifikasi teks. Dengan adanya hasil penelitian ini semoga memberikan referensi bagi mahasiswa yang sedang mengerjakan karya ilmiah. Ini juga diharapkan kedepannya penelitian ini dapat dilanjutkan seperti menggunakan algoritma yang berbeda.