

# 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 pemanfaatan jaringan komputer, perangkat lunak, dan sistem untuk pemrosesan dan distribusi data adalah semua aspek teknologi informasi (TI). Begitu juga untuk mengklasifikasi jurnal dimana dahulu masih dilakukan secara manual, kemudian seiring perkembangan zaman, mengklasifikasi jurnal menjadi otomatis dengan berkembangnya teknologi informasi. Versi digital dari berbagai karya ilmiah akademik, termasuk tesis, laporan penelitian, laporan kerja praktek, dan sebagainya, tersedia. Namun, jumlah informasi atau pengetahuan yang dapat dipetik dari dokumen elektronik tersebut umumnya tidak bertambah seiring dengan fenomena tersebut (Akromunnisa and Hidayat 2019).

Metode *Text Mining* merupakan salah satu teknik data mining yang dapat digunakan untuk mengatasi masalah tersebut. Algoritma *Text Mining* dirancang untuk mengenali data semi-terstruktur seperti konten dokumen, abstrak, dan sinopsis (Sipayung, Fauziah, and Nurhayati 2020).

Klasifikasi teks merupakan salah satu proses yang dapat dilakukan dengan text mining. Proses pengklasifikasian suatu dokumen teks dapat disebut sebagai klasifikasi teks. *Support Vector Machine (SVM)*, *Naive Bayes*, *k-Nearest Neighbor (KNN)*, *Decision Tree*, dan *Artificial Neural Networks (ANN)* adalah beberapa dari berbagai algoritma yang dapat digunakan dalam proses klasifikasi teks (Akromunnisa and Hidayat 2019).

Pada penelitian terdahulu sudah ada beberapa yang mengembangkan metode ini salah satunya yaitu menghasilkan bahwa algoritma *Naive Bayes* dapat digunakan dalam otomatisasi klasifikasi teks dengan mendapatkan akurasi rata-rata sebesar 97,7% (Rizqiyani et al 2017). Kemudian ada yang menunjukkan bahwa algoritma *Naive Bayes* dan *K-Nearest Neighbor* dapat digunakan dalam mengklasifikasi teks, pada penelitian ini akurasi *k-Nearest Neighbor* lebih baik dengan akurasi 91,6% dibandingkan *Naive Bayes* yang mendapatkan akurasi 86,1% (Asril et al 2019). Penelitian selanjutnya mendapatkan hasil bahwa *Naive Bayes* lebih baik yang mampu mengklasifikasikan 28 dokumen dari 40 data uji dibanding *K-Nearest Neighbor* yang mampu mengklasifikasikan 16 dokumen dari 40 data uji (Devita et al 2018). Penelitian lainnya mendapatkan hasil bahwa *K-Nearest Neighbor* dapat melakukan klasifikasi data dengan mendapatkan akurasi sebesar 98,7% (Hasanah et al. 2022). Adapun yang membandingkan metode *SVM*, *NB*, dan *K-NN* untuk dokumen teks yang berbahasa Inggris dan menentukan mana dari tiga kategori yang telah ditentukan yang termasuk dalam kategori tersebut. *NB* memiliki akurasi rata-rata 72,3%, *K-NN* memiliki akurasi rata-rata 84,53%, dan *SVM* memiliki akurasi rata-rata 86,06% (Yusra, Olivita, and Dkk. 2016).

Pada penelitian ini akan dilakukan perbandingan dua metode klasifikasi yaitu *Naive Bayes* dan *K-NN*. Kedua metode tersebut digunakan untuk klasifikasi jurnal menggunakan

judul artikel. Penelitian ini juga akan menganalisis hasil akurasi dari kedua metode yang cocok digunakan untuk klasifikasi jurnal berdasarkan judul artikel.

## 1.2 Rumusan masalah

Rumusan masalah pada penelitian ini adalah, 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 Penelitian

Tujuan Penelitian ini adalah untuk menjawab identifikasi diatas, 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

Berdasarkan dari perumusan masalah di atas, maka batasan masalah dalam penelitian analisis klasifikasi jurnal berbasis web sebagai berikut:

1. Penelitian ini menggunakan dua algoritma yaitu *Naïve Bayes* dan *K-Nearest Neighbor* untuk klasifikasi jurnal.
2. Bahasa pemrograman yang digunakan adalah *Python*.

## 1.5 Kontribusi

Hasil penelitian ini diharapkan dapat berguna untuk mengklasifikasikan, khususnya dalam mengklasifikasikan 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.