

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam penentuan kategori artikel khususnya pada artikel kesehatan, manusia terkadang menemukan kesulitan dalam mengklasifikasi atau mengelompokkan suatu artikel berdasarkan masing-masing kategori artikel. Karena teks dari suatu artikel terkadang memiliki suatu kemiripan antara artikel yang satu dengan artikel lainnya yang akan membuat pembaca akan mengalami kesulitan dalam mengklasifikasi. Dalam pengolahan semua kata yang ada pada artikel yang membuat sulit adalah bagaimana mengidentifikasi kata dari artikel tersebut menurut masing-masing kategori yang ada pada artikel kesehatan, diantaranya Spesialis Anak, Spesialis Mata, Spesialis Kulit, Spesialis Jantung, Spesialis Gizi, Spesialis Syaraf, maka digunakanlah metode *Naive Bayes Classifier*.

Naive Bayes Classifier merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naive Bayes Classifier* adalah Teorema Bayes, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang, Bayes Optimal Classifier menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

Pada proses pengelompokan artikel, permasalahan yang terjadi adalah mengelompokkan suatu artikel berdasarkan masing-masing kategori. Dalam pengelompokan tersebut salah satu cara untuk mengelompokkannya adalah memanfaatkan sebuah sinopsis yang terkandung didalam sebuah artikel, dari sinopsis itulah pengelompokan dapat dilakukan, sinopsis merupakan ringkasan atau inti dari sebuah teks artikel, maka dengan data sinopsis artikel bisa dapat membantu dalam mencari dan mengelompokkan artikel berdasarkan masing–masing kategori pada sebuah artikel.

Pada penelitian Tugas Akhir yang berjudul “**Klasifikasi Teks dengan Naive Bayes Classifier untuk Pengelompokan Teks Artikel**” penulis ingin melakukan

pengujian kinerja *Naive Bayes Classifier* untuk pengklasifikasian teks artikel berdasarkan masing-masing kategori. Objek penelitian adalah suatu teks artikel dalam bentuk digital atau media massa dalam bentuk elektronik. *Naive Bayes Classifier* digunakan untuk pengklasifikasian tingkat akurasi pada teks artikel.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan sebelumnya, terdapat beberapa permasalahan yang akan diangkat dalam penelitian ini, antara lain :

1. Bagaimana melakukan pre-prosesing menggunakan *text minning (tokenezing, filtering, stemming)* dan perhitungan *Naive Bayes Classifier*?
2. Bagaimana mengklasifikasikan artikel berdasarkan masing-masing katerogi menggunakan metode *Naive Bayes Classifier*?
3. Bagaimana melakukan uji kinerja metode *Naive Bayes Classifier*?

1.3 Batasan Masalah

Agar tidak menyimpang jauh dari permasalahan, maka penelitian ini mempunyai batasan masalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini merupakan dokumen teks artikel. Teks artikel yang digunakan pengujian adalah teks artikel dari sebuah media massa elektronik berbasis web.
2. Artikel yang digunakan adalah artikel pada ruang lingkup kesehatan yang berbahasa Indonesia.
3. Diklasifikasikan menjadi 6 kelas yaitu Spesialis Anak, Spesialis Mata, Spesialis Kulit, Spesialis Jantung, Spesialis Gizi, Spesialis Syaraf.
4. Aplikasi yang dibuat adalah aplikasi berbasis PHP (PHP Hypertext Preprocessor).
5. Metode yang digunakan adalah algoritma *Naive Bayes Classifier*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengklasifikasi teks artikel kesehatan dengan menggunakan algoritma *Naive Bayes Classifier*.
2. Mengolah data atau dokumen dengan *text mining*.
3. Pengimplementasian sistem untuk menguji kinerja algoritma *Naive Bayes Classifier*.

1.5 Manfaat Penelitian

Penelitian ini dilakukan dengan harapan dapat memberikan manfaat diantaranya sebagai berikut:

1. Dapat melakukan proses *text mining* dengan perhitungan akurasi teks artikel menggunakan metode *Naive Bayes Classifier*.
2. Dapat mengklasifikasikan teks artikel kesehatan berdasarkan masing-masing kategori.
3. Memberikan informasi artikel berdasarkan data yang relevan.