

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Saat ini diperkirakan 80% teks digital berbentuk tak teratur, tingginya volume teks di picu sebagai aktivitas seperti berita online, aktivitas akademi, jurnal digital, dsb. Jurnal digital berkembang melalui internet sehingga perkembangannya tidak dapat di bendung. Tingginya kuantitas dokumen akan memicu terjadinya penumpukan data baik itu secara offline maupun online. Document yang menumpuk menyebabkan pembengkakan memory. Akibat lain dari penumpukan dokumen adalah susahya mengkategorikan document, dan dokument akan sulit di temukan. Sehingga di pelukan sebuah *tool* untuk mengkategorikan document dengan cepat agar dokumen bisa tertata dan mudah di temukan.

Naïve Bayes Classifier merupakan metode yang dapat di gunakan untuk menyelesaikan permasalahan di atas, Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes . Metode pengklasifikasian menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Naive Bayes Classifier bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan oleh Xhemali , Hinde Stone dalam jurnalnya mengatakan bahwa “Naïve Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya”[1].

Tingkat akurasi Naïve Bayes juga sangat di tentukan dengan kualitas data yang di gunkan. Selain itu Naïve bayes juga dapat di kombinasikan dengan algoritma lain seperti TF-IDF, TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus.

Dengan di buatnya aplikasi berbasis Naïve Bayes di harapkan biasa membantu untuk menata ulang document yang awalnya tak teratur menjadi lebih teratur, agar bisa mempermudah peneliti sesudahnya untuk mencari jurnal sesuai dengan kategorinya.

1.2. Rumusan Masalah

Dari uraian permasalahan di atas maka kita dapat mengentehuai :

1. Bagaimana cara mengimplmentasikan Naïve Bayes Classifier dan TF-IDF?
2. Bagaimana cara menggolongkan teks berdasarkan tema menggunakan Naïve Bayes Classifier dan TF-IDF ?
3. Berapa tingkat akurasi Naïve bayes dan TF-IDF?

1.3. Batasan Masalah

Adapun batasan masalah yang sudah di tentukan yaitu:

1. Abstract yang digunakan mencapai 800.
2. Sistem yang digunakan khusus untuk mengidentifikasi teks.
3. Bahasa pemrograman yang digunakan adalah Python dan php.
4. Platform yang digunakan adalah website.
5. Waktu pemrosesan perhitungan sistem mencapai 25 detik

1.4. Maksud dan Tujuan Penelitian

Adapun batasan masalah adalah sebagai berikut:

1. Mengetahui cara mengimplemntasikan Naïve Bayes Classifier dan TF-IDF.
2. Mengetahui cara menggolongkan teks berdasarkan tema menggunakan Naïve Bayes Classifier dan TF-IDF.

3. Mengetahui tingkat akurasi Naïve Bayes Classifier dan TF-IDF.

1.5. Metode penelitian

Pada penelitian ini, penulis menggunakan beberapa metode penelitian. Adapun metode-metode penelitian yang digunakan adalah sebagai berikut:

1. Metode Pengumpulan Data

Dalam penelitian ini data-data yang didapat dari beberapa metode, antara lain:

a. Arsip

Menghimpun dan mempelajari teori-teori mengenai naïve bayes dan TF-IDF. Sumber bacaan berupa e-book, jurnal-jurnal, artikel yang di peroleh dari internet guna menunjang penelitian.

b. Metode Deskriptif

Merumuskan dan menafsirkan data yang ada sehingga memberikan gambaran yang jelas mengenai objek penelitian secara umum.

2. Metode analisis

Dalam penelitian ini penulis menganalisis sistem menggunakan metode analisis KUANTITATIF yang digunakan untuk menguji suatu hipotesa yang berkaitan dengan *text mining*.

3. Metode Pengembangan

Dalam penelitian ini tahap perancangan dalam membuat sistem meliputi :

a. Pemodelan Proses

Pemodelan proses ini menggambarkan bagaimana bisnis beroperasi. Menggambarkan perintah yang dilakukan dan bagaimana cara data berpindah diantara perintah – perintah tersebut.

b. Pemodelan Data

Pemodelan data ini menggambarkan data yang digunakan dan diciptakan dalam suatu sistem bisnis. Model ini menunjukkan orang, tempat atau benda dimana data diambil dan hubungan antar data tersebut.

c. Desain Antarmuka Pengguna

Desain ini merupakan tampilan dimana pengguna atau user berinteraksi dengan sistem. Karena ada berbagai tingkat pengguna untuk mendesain suatu antarmuka pengguna diasumsikan pengguna yang menggunakannya merupakan pengguna akhir.

d. Metode Testing

Pengujian sistem pakar ini dilakukan untuk memastikan bahwa sistem telah sesuai dengan kebutuhan. Metode pengujian ini menggunakan metode whitebox dan blackbox. Pengujian – pengujian tersebut meliputi pengujian kesalahan peulisan (syntax error), kesalahan saat proses (runtime error) dan kesalahan logika (logical error).

1.6. Sistematika Penulisan

Untuk mempermudah dalam menyusun dan memahami isi dari Tugas Akhir ini, penulis menggunakan sistematika penulisan sebagai berikut :

Bab I. Pendahuluan

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, Batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab II. Landasan Teori

Bab ini membahas mengenai pengetahuan yang menjadi dasar teori yang berhubungan erat dan dipergunakan dalam mendukung pokok pembahasan tugas akhir yaitu Sistem Pakar Metode Naive Bayes.

Bab III. Analisis dan Perancangan

Dalam bab ini diuraikan tentang analisis dan perancangan sistem yang dibuat. Analisis meliputi analisa masalah dan analisa kebutuhan, sedangkan untuk perancangan terdiri dari perancangan proses, perancangan basis data serta perancangan antarmuka.

Bab IV. Implementasi dan Pembahasan

Bab ini berisi informasi tentang implementasi dari perancangan yang telah dibuat, yang meliputi coding program, serta hasil input dan output program. Serta pembahasan mengenai kelebihan dan kekurangan sistem yang dibuat.

Bab V. Penutup

Bab ini berisi kesimpulan dan saran untuk pengembangan sistem pada waktu yang akan datang.

