

1. Pendahuluan

1.1 Latar Belakang

Seiring dengan perkembangan jaman, maka kebutuhan manusia akan teknologi dan informasi semakin meningkat. Bahkan dengan adanya Internet, segala kebutuhan akan informasi menjadi lebih mudah didapatkan. Namun perkembangan Internet yang sangat pesat, menyebabkan penumpukan data dan informasi yang sangat banyak di Internet sehingga seringkali orang bingung untuk memilih informasi mana yang sesuai dengan yang diinginkan karena hanya dengan judul belum cukup untuk menggambarkan keseluruhan isi dari sebuah dokumen. Maka dari itu orang membutuhkan sesuatu yang mudah dan praktis untuk mengetahui apakah suatu dokumen itu sesuai dengan informasi yang dibutuhkan.

Kata kunci adalah kata atau kumpulan kata pada dokumen yang mampu memberikan deskripsi dari keseluruhan isi dokumen. Maka dengan adanya kata kunci dalam dokumen, pembaca akan lebih mudah mendapatkan dokumen yang isinya sesuai dengan yang dibutuhkan. Namun, kebanyakan dokumen justru tidak disertai dengan kata kunci. Pengekstraksian kata kunci secara manual untuk dokumen dalam jumlah besar membutuhkan waktu dan tenaga yang besar. Oleh karena itu, dibutuhkan suatu automasi pengekstraksian kata kunci. Selain itu dengan adanya kata kunci pada dokumen akan memudahkan proses information retrieval yang lain, seperti pengindeksian, klasterisasi dokumen, klasifikasi dokumen, text summarization, dan lain-lain.

Ekstraksi kata kunci merupakan suatu permasalahan klasifikasi, yaitu bagaimana mengklasifikasikan kata pada dokumen masukan menjadi kata kunci atau kata biasa (bukan kata kunci). Pada tugas akhir ini, akan diimplementasikan metode Naïve Bayes untuk mengekstraksi kata kunci pada dokumen teks. Naive Bayes merupakan sebuah pendekatan statistik yang mampu memutuskan sebuah pilihan berdasarkan probabilitas dan minimum error yang ditimbulkan atas pilihan tersebut berdasarkan training set yang diberikan. Metode Naïve Bayes ini sederhana dan tidak membutuhkan aturan semantik untuk proses ekstraksi. Naïve Bayes yang merupakan suatu metode learning, diharapkan mampu mengekstraksi kata kunci pada dokumen dan memberikan hasil yang baik dengan beberapa fitur yang digunakan pada training set, yaitu TFxIDF dan posisi kata-kata kunci pada dokumen.

1.2 Perumusan Masalah

Tugas Akhir ini mempunyai perumusan masalah sebagai berikut :

1. Bagaimana metode Naive Bayes dapat menemukan kata-kata yang merupakan kata kunci pada dokumen.
2. Bagaimana menentukan fitur-fitur yang digunakan untuk pengekstraksian kata kunci.
3. Bagaimana pengaruh penggunaan variasi fitur yang digunakan terhadap hasil ekstraksi.
4. Bagaimana pengaruh eliminasi *stopwords* sebagai tambahan pengetahuan kebahasaan.

5. Apakah metode yang diimplementasikan dapat menghasilkan kata-kata kunci yang sesuai dengan dokumen input.

Adapun batasan masalah dari tugas akhir ini adalah :

1. Dokumen teks yang digunakan dalam tugas akhir ini adalah dokumen abstraksi tugas akhir mahasiswa IT Telkom yang diambil dari *website* perpustakaan IT Telkom. Dokumen abstraksi tugas akhir yang digunakan sudah disertai dengan kata kunci, yang akan digunakan dalam proses *training* dan pengujian untuk mengetahui nilai keakurasian dari kata kunci yang dihasilkan sistem.
2. Term-term yang digunakan dalam dokumen latih dan dokumen uji merupakan kata. Frase penting dianggap sebagai term terkecil (kata).
3. Tidak melakukan penanganan terhadap frase.

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Mengimplementasikan metode Naïve Bayes untuk mengekstraksi kata kunci pada dokumen teks.
2. Menganalisis keakurasian metode Naïve Bayes dalam mengekstrak kata kunci pada dokumen teks berdasarkan kesesuaian antara kata kunci dan dokumen teks dengan parameter *precision*, *recall*, dan *f-measure*.
3. Menganalisis jumlah fitur yang digunakan dalam mengekstraksi kata kunci.
4. Menganalisis pengaruh eliminasi *stopwords* terhadap hasil ekstraksi kata kunci.

2.1 Metodologi Penyelesaian Masalah

Adapun metodologi yang digunakan dalam penulisan tugas akhir ini adalah:

1. Melakukan studi literatur mengenai ekstraksi kata kunci dan metode ekstraksi khususnya metode Naive Bayes.
2. Melakukan analisis kebutuhan perangkat lunak dan perancangan perangkat lunak ekstraksi kata kunci otomatis yang menerapkan metode Naive Bayes.
3. Melakukan implementasi perangkat lunak sesuai dengan perancangan yang telah dilakukan.
4. Melakukan pencarian data yang diperlukan untuk penelitian berupa dokumen abstraksi tugas akhir mahasiswa IT Telkom dari *website* perpustakaan IT Telkom.
5. Melakukan pengujian pada implementasi metode ekstraksi kata kunci terhadap dokumen abstraksi ta.
6. Melakukan analisis terhadap hasil pengujian.
7. Pembuatan laporan tugas akhir.