

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Seiring dengan perkembangan teknologi web, ada banyak dokumen yang mengandung informasi biografi yang dibuat dan diterbitkan secara online. Informasi yang disajikan dalam bentuk web digital bermacam-macam, diantaranya adalah majalah online, artikel. Menurut Hermione Lee, biografi adalah kisah atau keterangan tentang hidup seseorang yang ditulis oleh orang lain [1]. Teks dalam biografi memiliki salah satu ciri naratif dan dapat dikategorikan sebagai teks yang semi terstruktur atau tidak terstruktur.

Dokumen web teks biografi sering mengandung informasi relevan pada individu yang diceritakan seperti fakta individual dan tidak relevan seperti informasi lain yang tidak berhubungan pada individu yang diceritakan. Informasi yang relevan dapat digunakan untuk membuat database structural biografi atau membantu kepentingan penelitian tentang kemanusiaan dan sejenisnya. Meskipun manusia kebanyakan dapat menyaring informasi yang diinginkan secara manual, akan tetapi seiring dengan berkembangnya teknologi web, jumlah dokumen yang mengandung informasi biografi semakin banyak dipublish secara online. Hal ini akan menjadi masalah ketika jumlah dokumen yang disaring sangat banyak. Untuk itu dibutuhkan sebuah sistem untuk mengekstrak informasi secara otomatis guna membantu manusia dalam memperoleh informasi tersebut dengan cepat.

Ekstraksi Informasi merupakan salah satu cara untuk mendapatkan informasi tertentu dari suatu teks pada dokumen. Ekstraksi Informasi adalah pencarian otomatis pada informasi yang terstruktur seperti entitas, hubungan antar entitas, dan atribut yang menggambarkan entitas dari sumber yang tidak terstruktur [2]. Contoh dari ekstraksi informasi adalah mengekstraksi info web, jurnal, makalah ilmiah, dan lainnya. Teks pada halaman web memiliki bentuk yang tidak terstruktur. Kemudian ditambah dengan penulisan pada teks biografi tidak selalu memiliki format yang sama. Untuk itulah perlu dikembangkan sistem *machine learning* atau

pembelajaran mesin yang mampu mengenali pola pada teks biografi untuk menghasilkan informasi sesuai yang diinginkan.

Penelitian kali ini akan mencoba menggunakan salah satu algoritma *Machine Learning* yaitu *Naïve Bayes*, namun dengan pendekatan objek yang berbeda yaitu khususnya pada biografi di halaman web. Algoritma ini memiliki kelebihan sederhana, cepat dan tingkat akurasi yang tinggi [3]. Hal ini dibuktikan pada penelitian yang dilakukan oleh Xhemali dan rekannya pada perbandingan Algoritma *Naïve Bayes*, *Decision Tree*, dan *Neural Network* membuktikan bahwa Algoritma *Naïve Bayes Classifier* memiliki tingkat akurasi lebih baik dari kedua metode lain [3]. Maka pada penelitian ini akan dicoba implementasi algoritma *Naïve Bayes* untuk ekstraksi informasi pada teks biografi berbahasa indonesia dengan harapan algoritma tersebut dapat memberikan hasil yang lebih baik dari penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan. maka rumusan masalah dari penelitian ini adalah bagaimana performansi algoritma *Naïve Bayes* dalam menerapkan Ekstraksi Informasi pada teks biografi berbahasa indonesia.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk mengimplementasikan algoritma *Naïve Bayes* ke dalam ekstraksi informasi pada biografi. Sedangkan tujuan dari penelitian ini adalah untuk mengetahui tingkat performansi *Naïve Bayes* dalam ekstraksi informasi khususnya biografi teks Indonesia.

1.4 Batasan Masalah

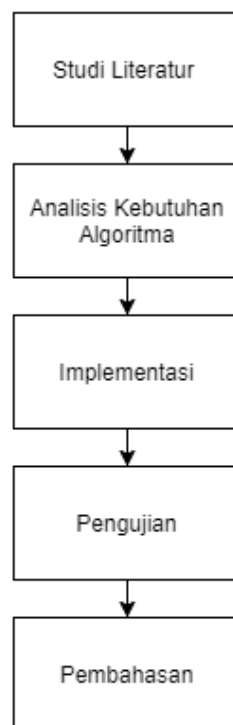
Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data set diperoleh dari halaman web internet menggunakan tools bernama httrack.
2. Teks biografi yang digunakan adalah teks berbahasa indonesia yang informasi biografi tokoh terkenal.

3. Informasi teks yang diekstrak adalah biografi tokoh terkenal pada situs web www.biografipedia.com. [4]
4. Output yang dideteksi adalah kalimat dengan kelas nama, tempat dan tanggal lahir, pendidikan.

1.5 Metode Penelitian

Metodologi penelitian merupakan proses atau cara ilmiah untuk mendapatkan data yang digunakan untuk keperluan penelitian [5]. Metode penelitian adalah salah satu metode yang digunakan khusus dalam penelitian ini. Adapun tahapan pada penelitian ini terdiri dari beberapa tahap seperti Studi Literatur, Analisis Kebutuhan Algoritma, Implementasi, dan Pengujian. Alur dari tahapan dapat dilihat pada Gambar 1.1 berikut:



Gambar 1.1 Metode Penelitian

1.5.1 Studi Literatur

Studi literatur dilakukan untuk mendapatkan pengetahuan tentang ekstraksi informasi dan algoritma *Naïve Bayes* yaitu dengan mengumpulkan informasi dari buku, jurnal, paper, dan pencarian materi di *internet*.

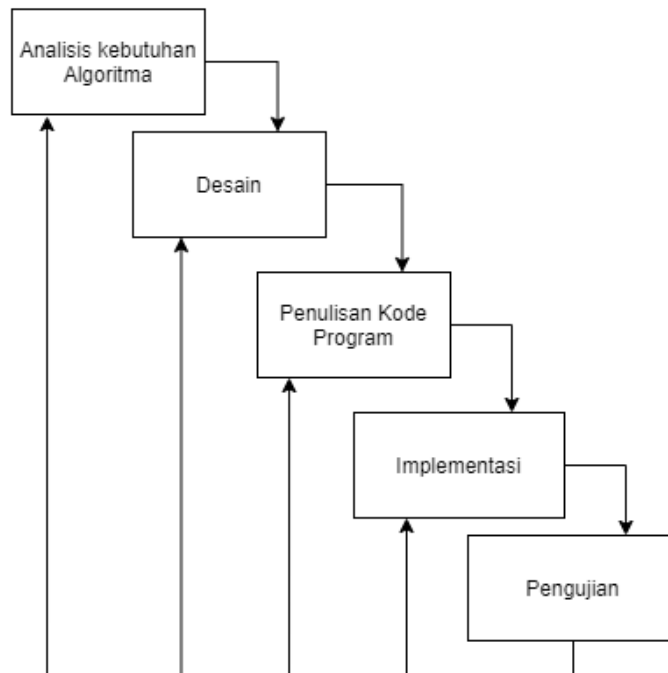
1.5.2 Analisis Kebutuhan Algoritma

Tahapan analisis yang digunakan setelah mengkaji literatur dari berbagai sumber jurnal, paper dan materi di internet adalah:

1. Menganalisis kebutuhan dataset teks biografi untuk memahami pola dan bentuk Informasi yang terkandung dalam teks. Kemudian analisis preprocessing dengan ekstraksi fitur.
2. Menganalisis tahapan klasifikasi *Naïve Bayes* yang terdiri dari melakukan perhitungan data latih yang kemudian akan digunakan ketika tahap pengujian. Hasil dari analisis pelatihan ini adalah parameter naïve bayes terhadap data baru yang akan diuji.
3. Menganalisis tahap pengujian yang setelah penerapan aplikasi dengan metode *naïve bayes* dilakukan untuk mengekstrak informasi dari masukan baru.
4. Menganalisis data keluaran hasil ekstraksi dari tahap pengujian untuk mengetahui apakah *Naïve Bayes* memiliki nilai akurasi yang baik atau tidak.

1.5.3 Implementasi

Implementasi pada penelitian ini adalah membangun perangkat lunak menggunakan model *Waterfall*. Model ini menggunakan pendekatan secara sistematis dan berurutan dari tahap awal hingga akhir [6]. Proses dari model *Waterfall* dapat dilihat pada Gambar 1.2.



Gambar 1.2 Model *Waterfall*

Berikut adalah penjelasan tiap proses pada model *waterfall*.

1. Analisis Kebutuhan

Pada tahap ini dilakukan analisis kebutuhan yang diperlukan yang berkaitan pada pembangunan perangkat lunak.

2. Desain

Pada tahap ini dilakukan perancangan arsitektur sistem yang akan dibangun. Penggunaan bahasa pemrograman serta tampilan antarmuka yang sederhana dan mudah dipahami.

3. Penulisan Kode Program

Pada tahap ini penulisan kode program dilakukan sesuai dengan desain sistem yang dirancang pada tahap sebelumnya dengan bahasa yang dimengerti oleh komputer.

4. Implementasi

Pada tahap ini implementasi sistem yang telah dibangun apakah sesuai dengan tujuan yang telah ditentukan.

5. Pengujian

Pada tahap akhir ini dilakukan pengujian terhadap sistem yang dibangun apakah sesuai dengan hipotesa.

1.5.4 Pengujian

Setelah tahap implementasi selanjutnya dilakukan pengujian untuk mengetahui hasil dari tahap implementasi. Pengujian menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi dan performa yang diberikan *Naïve Bayes* pada penelitian ini. Parameter pengujian yang dilakukan adalah dengan mengukur nilai akurasi, presisi, *recall*, dan *f-measure*. Serta perbandingan hasil penelitian dengan penelitian telah dilakukan sebelumnya yang berkaitan.

1.6 Sistematika Penulisan

Sistematika penulisan ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan secara umum adalah sebagai berikut:

BAB 1 PENDAHULUAN

Berisi penjelasan tentang latar belakang mengapa penelitian ini dilakukan, rumusan masalah berdasarkan latar belakang, maksud dan tujuan, batasan-batasan dari penelitian yang dilakukan, metodologi penelitian, serta sistematika penulisan yang digunakan.

BAB 2 LANDASAN TEORI

Berisi teori dan konsep yang digunakan sebagai landasan atau acuan yang diperlukan yaitu antara lain, penjelasan tentang konsep dasar *Artificial Intelligence* dan *Machine Learning*, *Naïve Bayes*, *Information Extraction* serta penjelasan singkat mengenai model *waterfall* dan lain-lain.

BAB 3 ANALISIS DAN PERANCANGAN

Berisi pembahasan analisis dari rumusan masalah yang sudah dijelaskan di bagian pendahuluan, analisis kebutuhan algoritma mulai dari analisis masalah, analisis *preprocessing*, analisis metode *Naïve Bayes*, sampai analisis kebutuhan fungsional dan *non-fungsional*.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Berisi pembahasan implementasi yang dilakukan, mulai dari pengujian dari sistem yang dibangun, *hardware* dan *software* yang digunakan, serta hasil pengujian dari metode *Naïve Bayes*. Pengujian yang dilakukan antara lain pengujian parameter *Confusion Matrix* yang memiliki akurasi, *precision*, *recall* pada sistem aplikasi untuk menguji tingkat akurasi dari ekstraksi informasi pada halaman web.

BAB 5 KESIMPULAN DAN SARAN

Berisi kesimpulan dari hasil pengujian yang dilakukan, apakah tujuan awal dalam penelitian telah tercapai atau tidak, serta berisi saran untuk penelitian selanjutnya.

