

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kategorisasi dokumen teks merupakan suatu kegiatan yang dilakukan untuk mengelompokkan dokumen berbentuk teks untuk ditempatkan ke dalam kategori tertentu sesuai dengan kesamaan ciri dari topik yang dibicarakan pada dokumen teks tersebut. Untuk mengkategorikan dokumen teks tersebut, terdapat disiplin ilmu yang menjelaskan secara khusus mengenai pengelompokan dokumen yang berbentuk teks, yaitu *text mining*. *Text mining* itu sendiri adalah suatu proses untuk menganalisis teks dengan mencari pola-pola informasi yang digali dari dokumen teks yang tidak terstruktur. Hasil dari mendapatkan pola-pola teks tersebut selanjutnya dapat dilakukan proses pengelompokan. Proses pengelompokan teks dapat dilakukan menggunakan metode pengkategorian yang terdapat pada *text mining*. Metode yang dapat digunakan untuk pengelompokan teks sangat banyak dan beragam, seperti ada metode *k-means*, *naïve bayes classifier*, *k-nearest neighbor*, *support vector machine (SVM)*, *artificial neural network*, *fuzzy k-nearest neighbor*, dan lain sebagainya. Namun, dari sekian banyak metode dipilihlah metode yang memiliki kinerja baik. Untuk pemilihan metode dapat dilakukan dengan membandingkan antara metode pengkategorian dokumen teks satu dengan metode yang lainnya.

Berdasarkan penelitian yang dilakukan oleh Hamzah mengenai proses klasifikasi dokumen berita dan abstrak akademis menggunakan metode *naïve bayes classifier* didapat nilai akurasi sebesar 91% [1]. Berdasarkan penelitian yang dilakukan oleh Samuel dkk didapatkan nilai akurasi untuk proses

pengkategorian subtopic bertia menggunakan metode *k-nearest neighbor* sebesar 88,29% [2]. Berdasarkan penelitian yang dilakukan oleh Hamzah dan Samuel, maka penelitian ini akan membandingkan kinerja antara metode *k-nearest neighbor* dengan *naïve bayes* untuk mendapatkan metode yang baik dalam proses pengkategorian dokumen teks.

K-Nearest Neighbor atau biasa disingkat dengan istilah KNN biasa digunakan dalam proses klasifikasi, salah satunya dalam klasifikasi teks. *K-Nearest Neighbor* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain [3]. Untuk mengetahui jarak antar data baru dengan data lama dapat dilakukan dengan menghitung bobot data-data tersebut berdasarkan fitur yang ada, kemudian bobot tersebut akan dicocokkan dan dihitung jaraknya.

Naïve Bayes merupakan suatu metode yang digunakan untuk mengklasifikasikan suatu objek berdasarkan perhitungan probabilitas. *Naïve Bayes* akan memprediksi suatu objek dengan memperhitungkan probabilitas suatu bukti untuk dijadikan sebagai hipotesis.

Pada penelitian kali ini, pengkategorian dokumen teks akan dibatasi dengan mengkategorikan artikel berita berbahasa Indonesia yang didapat dari beberapa situs berita *online* dengan judul **“Perbandingan Kinerja Metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* pada Kategorisasi Artikel Berita di Internet”**. Penelitian ini akan menggunakan akurasi ketepatan dalam proses pengkategorian dan waktu yang dibutuhkan untuk melakukan proses kategori sebagai parameter perbandingan kinerja dari kedua metode tersebut.

1.2. Perumusan Masalah

Berdasarkan rumusan masalah yang dipaparkan di atas, maka diambil suatu rumusan masalah sebagai berikut:

1. Mengkategorikan artikel berita *online* menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier*.
2. Membandingkan kinerja *K-Nearest Neighbor* dan *Naïve Bayes Classifier* dengan mengukur tingkat akurasi yang didapat dan waktu yang dibutuhkan untuk mengkategorikan artikel berita *online*.

1.3. Batasan Masalah

Penelitian ini memiliki batasan terhadap objek yang akan dibahas, mengingat adanya keterbatasan waktu dalam proses penelitian ini, selain itu juga agar penelitian yang akan dilakukan tidak melebar ke topik lainnya. Adapun batasan masalah dalam penelitian ini adalah:

1. Perangkat lunak yang akan dibuat berbasis *website* menggunakan Bahasa pemrograman PHP dan MySQL.
2. Data yang digunakan untuk proses pengujian adalah artikel berita berita berbahasa Indonesia yang diambil dari situs berita *online*.
3. Artikel berita *online* yang digunakan merupakan artikel yang berasal dari situs berita www.detik.com, www.kompas.com, www.liputan6.com, www.viva.co.id, dan www.tribunnews.com.
4. Penelitian yang akan dilakukan adalah dengan membandingkan metode klasifikasi *K-Nearest Neighbor* dengan *Naïve Bayes* menggunakan nilai akurasi ketepatan klasifikasi dan waktu yang digunakan untuk proses klasifikasi sebagai parameter perbandingan.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan di atas, maka tujuan dari penelitian ini adalah:

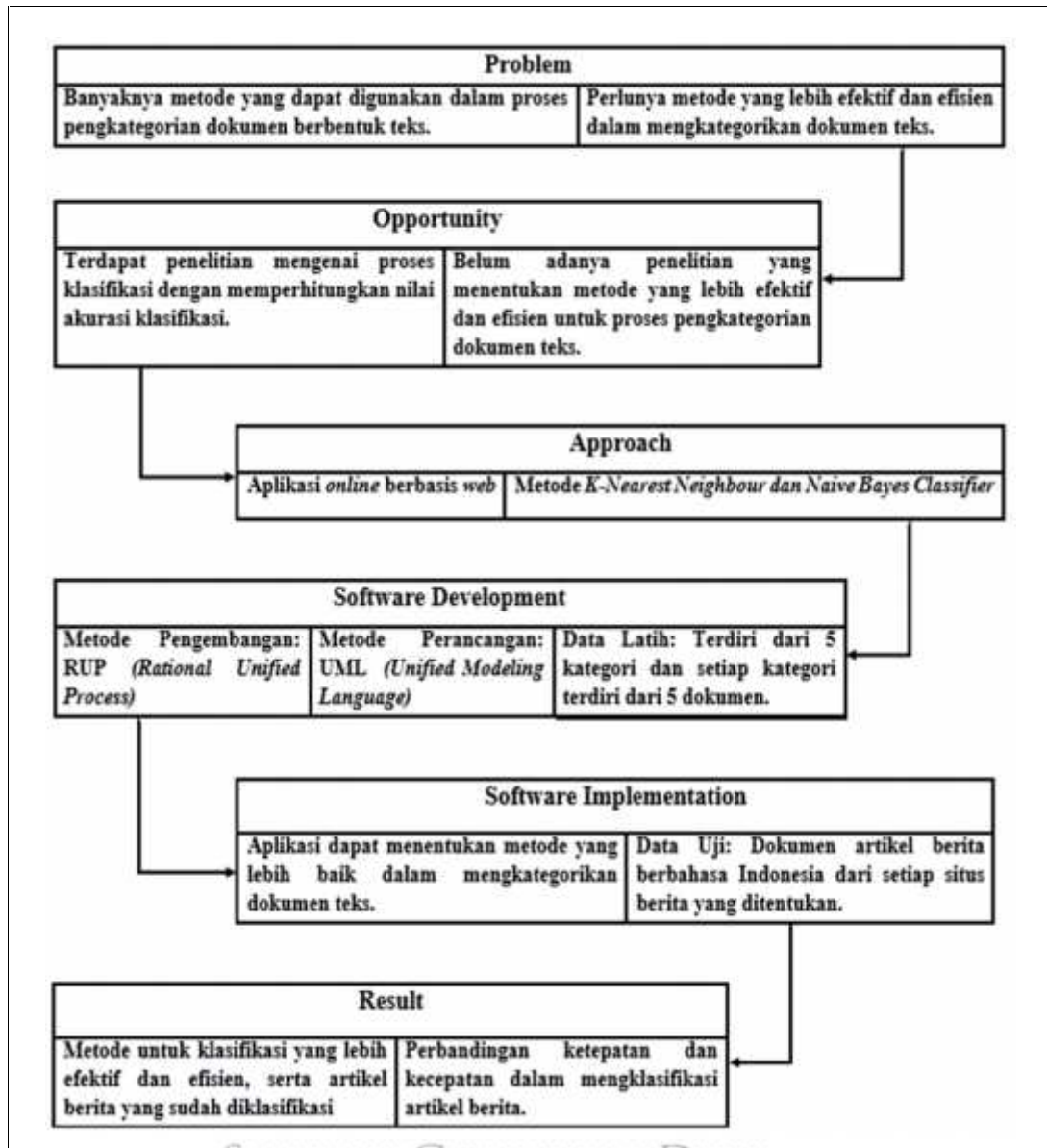
1. Untuk menerapkan metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* dalam mengkategorikan artikel berita *online*.
2. Untuk mengetahui metode yang lebih baik dalam mengkategorikan artikel berita *online*, yaitu dengan membandingkan kinerja metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier*.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini diharapkan dapat mengetahui metode yang lebih baik, yaitu antara metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* untuk digunakan dalam mengkategorikan artikel berita *online*, sehingga dapat menerapkan metode tersebut pada sistem yang dapat mengkategorikan artikel berita *online*. Selain itu, aplikasi yang dibuat dapat membantu dalam mengkategorikan artikel berita secara otomatis.

1.6. Kerangka Pemikiran

Adapun kerangka pemikiran dari Perbandingan Kinerja Metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* pada Kategorisasi Artikel Berita di Internet dapat dilihat pada Gambar 1.1 di bawah ini.



Gambar 1.1 Kerangka Pemikiran

1.7. Metode Penelitian

Metode penelitian ini terdiri dari dua tahap, yaitu tahap pengumpulan data dan tahap pengembangan perangkat lunak.

1.7.1. Pengumpulan Data

Pada tahap pengumpulan data untuk membantu pembuatan aplikasi ini yaitu dengan melakukan beberapa cara, diantaranya:

a. Observasi

Observasi di sini dilakukan dengan meninjau langsung *website* berita *online*. Peninjauan ini untuk mengamati data yang terdapat di *website* yang akan dijadikan sebagai bahan analisis kategorisasi.

b. Studi pustaka

Studi pustaka di sini merupakan panduan dalam memecahkan masalah yang akan dibahas dengan mencari beberapa referensi sebagai landasan teoritis untuk menunjang penyusunan laporan tugas akhir ini, seperti buku cetak, jurnal, maupun referensi lainnya yang berasal dari internet.

c. Studi Dokumentasi

Studi dokumentasi dilakukan dengan mengamati dokumen-dokumen tertulis yang memiliki hubungan yang sama dalam menyelesaikan masalah.

1.7.2. Pengembangan Perangkat Lunak

Model pengembangan perangkat lunak yang akan digunakan dalam membuat aplikasi ini yaitu model RUP (*Rational Unified Process*). Model RUP menggunakan konsep *object oriented* dengan aktivitas yang fokus pada pengembangan model menggunakan *Unified Model Language* (UML) [4]. Model RUP memiliki empat tahapan pengembangan perangkat lunak, diantaranya:

1. *Inception*

Pada tahap *inception*, *developer* mendefinisikan batasan kegiatan, melakukan analisis kebutuhan *user*, dan melakukan perancangan awal perangkat lunak.

2. *Elaboration*

Pada tahap *elaboration* dilakukan perancangan perangkat lunak yang terdiri dari menspesifikasikan perangkat lunak hingga merilis prototype perangkat

lunak.

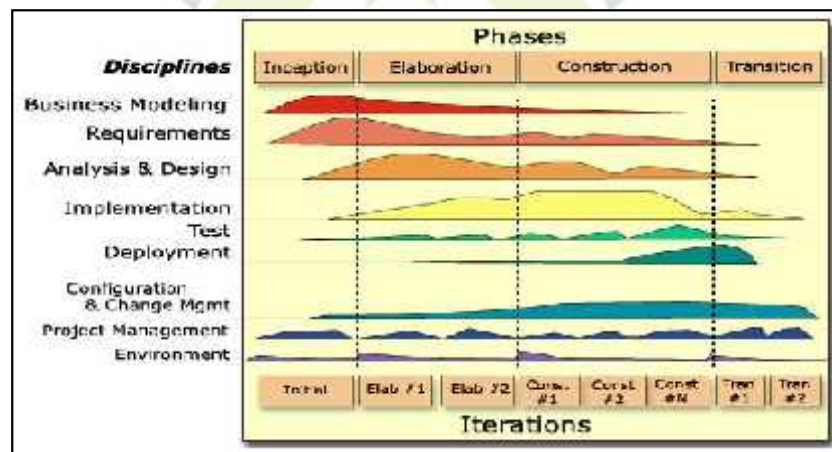
3. *Construction*

Tahap *construction* merupakan tahap implementasi rancangan perangkat lunak yang telah dibuat.

4. *Transition*

Pada tahap ini akan dilakukan penyerahan sistem kepada *user* dan dilakukan pelatihan kepada *user*. Tahap *transition* merupakan proses pembuatan sistem yang sudah dimodelkan menjadi sebuah produk jadi.

Gambar 1.2 di bawah ini merupakan arsitektur dari model pengembangan perangkat lunak RUP.



Gambar 1.2 Model Pengembangan Perangkat Lunak RUP

1.8. Sistematika Penelitian

Sistematika penulisan laporan tugas akhir ini terdiri dari beberapa bab dengan masing-masing bab berisi uraian singkat dan penjelasan mengenai pokok pembahasan. Adapun sistematika penulisan laporan tugas akhir sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, *state of the art*,

kerangka pemikiran, metode penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini menjelaskan mengenai tinjauan pustaka dan penjelasan singkat dari beberapa teori-teori yang berhubungan dengan topik yang dibuat berdasarkan hasil penelitian dan hal-hal lainnya yang dapat membantu dalam proses penulisan laporan tugas akhir ini.

BAB III ANALISIS DAN PERANCANGAN

Bab ini berisi mengenai analisis dan perancangan aplikasi yang akan dibuat, yaitu berisi tentang identifikasi masalah, perancangan aplikasi yang terdiri dari desain sistem, basis data, dan antarmuka, cara kerja aplikasi, dan evaluasi aplikasi.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi mengenai implementasi dari sistem yang dibangun baik itu yang berhubungan dengan *software* atau *hardware* yang dibutuhkan untuk mendukung aplikasi yang dibuat, selain itu juga dilakukan pengujian terhadap aplikasi yang dibuat.

BAB V PENUTUP

Bab ini berisi pembahasan singkat mengenai aplikasi yang dibuat secara keseluruhan serta saran untuk pengembangan perangkat lunak yang lebih baik untuk ke depannya.