

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dewasa ini, informasi merupakan hal yang sangat dibutuhkan oleh banyak orang. Internet merupakan salah satu sumber informasi terbesar dan tercepat. Pesatnya penggunaan Internet mengakibatkan pertumbuhan dan pertukaran informasi yang sangat cepat dibandingkan sebelumnya. Begitu juga dengan jumlah informasi berupa artikel berita berbahasa Indonesia berbasis web yang terdapat pada Internet juga sangat berlimpah dan beragam, sehingga mengakibatkan jumlahnya meningkat secara eksponensial. Perkembangan informasi ini memungkinkan informasi dapat diakses dengan mudah oleh pengguna. Namun, dengan jumlah informasi yang terus bertambah setiap harinya menimbulkan masalah dan tantangan yang cukup besar, karena dapat menyulitkan pembaca dalam mencari topik berita yang diinginkan.

Oleh karena itu, diperlukan suatu pengkategorisasian terhadap informasi yang berupa artikel berita yang memudahkan pembaca untuk mencari topik berita yang mereka inginkan. Salah satu cara yang dapat mengkategorikannya adalah dengan menggunakan teknik kategorisasi. Sekumpulan dokumen artikel berita tidak hanya memiliki *noise* tapi juga memiliki *feature space* yang berdimensi tinggi. Hal ini dapat menjadi masalah karena dapat menyebabkan rendahnya tingkat keakuratan dalam pengkategorisasian teks. Untuk mengatasinya digunakan teknik *feature selection* yang dapat mengurangi dimensionalitas data yang besar, membuang data yang tidak relevan, dan meningkatkan akurasi hasil.

*Feature selection* terdapat beberapa pendekatan, salah satunya adalah pendekatan *wrapper feature selection*. Melalui pendekatan ini dapat menghasilkan akurasi yang tinggi. *Wrapper* bergantung kepada classifier untuk mengevaluasi setiap feature subset. Akan dilakukan dengan menggunakan algoritma induksi *Naive Bayes* sebagai *classifier*. Proses pemilihan feature dilakukan dengan melakukan pencarian terhadap feature subset yang kemudian akan dievaluasi dengan menggunakan pengukuran performansi dari *classifier* itu sendiri.

Pada tugas akhir ini akan dibahas metoda pencarian feature subset untuk proses *feature selection* pada artikel berbahasa Indonesia dengan pendekatan *wrapper*, yaitu dengan metoda *Hill Climbing search* dan *Best First Search*. Pencarian ini memiliki komputasi yang tinggi dengan cost yang banyak. Tugas akhir ini merupakan bagian dari riset text mining.

## 1.2 Perumusan Masalah

Permasalahan yang akan diselesaikan dalam tugas akhir ini yaitu:

1. Bagaimana implementasi metoda pencarian feature subset dengan menggunakan algoritma pencarian *Hill Climbing search* dan *Best First Search*?
2. Bagaimana menganalisis pencarian feature subset dengan menggunakan algoritma pencarian heuristik yaitu *Hill Climbing search* dan *Best First search*?

Adapun batasan masalah pada tugas akhir ini yaitu:

1. Dataset yang akan digunakan merupakan artikel berita berbahasa indonesia yang diperoleh dari web dan data bersifat *offline* yang disimpan dalam file berekstensi .txt.
2. Klasifikasi yang digunakan dengan menggunakan algoritma induksi *Naive Bayes* yang ada di WEKA.
3. Feature hanya berupa kata bukan berupa frasa.

## 1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Mengimplementasikan metode *feature selection* untuk memilih atribut secara tepat dengan pendekatan *wrapper*.
2. Menganalisis pengaruh algoritma pencarian feature subset menggunakan *Hill Climbing search* dan *Best First Search* terhadap performansi hasil klasifikasi berdasarkan nilai *macro-average F-measure*.

## 1.4 Metodologi Penelitian

Metodologi yang digunakan untuk menyelesaikan masalah dalam tugas akhir ini adalah :

1. Studi literatur.  
Mencari dan mengumpulkan informasi serta memahami dan mempelajari konsep *wrapper* pada *feature selection* serta konsep *text mining* melalui literatur berupa makalah, buku, atau jurnal yang berhubungan dengan *feature selection*, *wrapper feature selection*, *text categorization*.
2. Pencarian dan pengumpulan data.  
Data yang akan digunakan berupa artikel berita berbahasa Indonesia yang diperoleh dari web.
3. Analisis kebutuhan dan implementasi aplikasi yang akan dibangun.  
Analisis kebutuhan dilakukan dengan merancang sistem kebutuhan perangkat lunak. Sedangkan implementasi akan dilakukan terhadap hasil analisis sistem kebutuhan perangkat lunak.

4. Pengujian

Melakukan pengujian dan analisis terhadap dataset yang diklasifikasikan dengan perangkat lunak yang dibangun serta mengukur performansi dari *classifier*.

5. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.