

# 1 PENDAHULUAN

Bab ini menjabarkan mengenai latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah dan sistematika penulisan.

## 1.1 Latar Belakang

Berita merupakan laporan atau cerita mengenai kejadian atau peristiwa aktual dan faktual yang diproduksi setiap hari untuk memberikan informasi terbaru kepada masyarakat[1]. Survei yang dilakukan UC Browser pada tahun 2016 lalu menyatakan bahwa 56,5% pengguna internet di Indonesia rata-rata membaca 4-12 artikel berita per hari[2]. Data ini menunjukkan bahwa konsumsi artikel berita di Indonesia cukup besar. Pembaruan berita yang diproduksi setiap harinya menyebabkan jumlah berita mengalir dengan cukup besar. Aliran berita yang terus bertambah membuat data dalam bentuk teks tidak mudah diolah sehingga perlu dilakukan klasifikasi artikel berita secara otomatis.

Klasifikasi data dalam bentuk teks merupakan bagian ilmu dari *text mining* yang mampu memberikan solusi dalam klasifikasi artikel berita berdasarkan *learning* terhadap data yang sudah dipersiapkan sebelumnya. Berita yang telah diklasifikasikan akan memudahkan kita untuk manajemen data teks serta dapat memberikan informasi singkat tentang artikel berita kepada pembaca. Selain itu juga dengan klasifikasi artikel berita bahasa Indonesia yang dilakukan secara otomatis dapat mengefisienkan waktu pengklasifikasian.

Namun dalam klasifikasi sendiri terdapat masalah yang berhubungan dengan unsur ketidakpastian dalam menentukan sebuah kelas, yaitu dimana dalam beberapa kelas memiliki kata atau fitur penciri yang sama sehingga sistem akan sulit dalam mengklasifikasikannya[5, 6]. Salah satu metode dalam bidang *Artificial Intelligence* yang dapat menangani masalah yang berkaitan dengan unsur ketidakpastian adalah *Naive Bayes*[5]. Dimana metode *Naive Bayes* sendiri menerapkan teorema *Bayesian* yang memiliki ciri utama asumsi sederhana (*naive*) terhadap independensi antar variabel input dan menggunakan perhitungan probabilistik dalam pengambilan keputusan[4]. *Multinomial Naive Bayes Classifier* merupakan salah satu metode perhitungan probabilistik yang menerapkan teorema

*Bayesian*. Oleh karena itu, pada penelitian ini penulis menggunakan *Multinomial Naive Bayes Classifier* dalam mengklasifikasikan artikel berita bahasa Indonesia.

Selain itu, permasalahan lain dalam klasifikasi teks adalah tingginya dimensi fitur. Dimana masalah ini dapat menyebabkan kompleksitas komputasi tinggi dan akurasi yang rendah[6]. Oleh karena itu, untuk mengatasi masalah tersebut diusulkan teknik pemilihan fitur atau *feature selection* untuk mengurangi tingginya dimensi fitur data.

Untuk mengurangi dimensi fitur secara tepat perlu dilakukan perhitungan untuk mengukur kuatnya hubungan antara kata dengan kelasnya. *Chi-square* merupakan metode *feature selection* yang sangat efisien untuk memilih kata penciri suatu dokumen[5, 6]. Dalam teori statistik, pengujian metode *Chi-Square* digunakan untuk menguji independensi antara *term* dalam suatu kategori/kelas tertentu[8]. Oleh karena itu, dalam penelitian ini penulis menggunakan metode *Chi-Square feature selection* untuk mengurangi dimensi data agar dapat meningkatkan keefektifan dan mengurangi waktu komputasi *Multinomial Naive Bayes* dalam mengklasifikasikan artikel berita bahasa Indonesia.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang dipaparkan, maka pada penelitian ini penulis merumuskan masalah sebagai berikut:

1. Bagaimana membangun suatu sistem klasifikasi dalam mengklasifikasikan artikel berita berbahasa Indonesia?
2. Bagaimana menganalisa performa *Chi-Square feature selection* dan *Multinomial Naive Bayes Classifier* dalam pengklasifikasian artikel berita berbahasa Indonesia.

Adapun batasan masalah dalam pengerjaan tugas akhir ini yaitu:

1. Artikel berita yang diambil hanya judul dan isi artikel berita yang disimpan dalam bentuk (.txt). Dataset yang dikumpulkan oleh penulis sebanyak 720 data teks dimana setiap kelas memiliki 60 data.
2. Artikel berita yang diambil adalah artikel berita dari rentang tanggal 1 November 2016 – 30 April 2017.

3. Artikel berita yang diambil bersumber dari halaman *website kompas.com, tribunnews.com, republika.com, mediaindonesia.com* dan *sindonews.com*.
4. Sistem klasifikasi yang dibangun untuk menangani masalah *multiclass*.
5. Pada proses *stemming* menggunakan library Sastrawi.
6. Terdapat 12 kelas dalam penelitian tugas akhir ini yaitu : Ekonomi, Gaya Hidup, Hiburan, Hukum, Kesehatan, Olahraga, Otomotif , Pendidikan, Politik, Properti, Teknologi dan Wisata.
7. Implementasi dilakukan menggunakan bahasa pemrograman Python.

### 1.3 Tujuan

Adapun tujuan yang ingin dicapai penulis dalam menyelesaikan penelitian ini adalah sebagai berikut:

1. Menganalisis dan mengimplementasikan *Chi-Square feature selection* dan *Multinomial Naive Bayes Classifier* dalam mengklasifikasikan artikel berita berbahasa Indonesia.
2. Menganalisis performa model yang dihasilkan oleh *Multinomial Naive Bayes* yang dipengaruhi oleh parameter  $\alpha$  pada *Chi-Square feature selection*.

### 1.4 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah pada pelaksanaan tugas akhir ini adalah sebagai berikut:

1. Studi Literatur  
Mempelajari literatur yang relevan dengan topik klasifikasi yang diambil guna memahami dan menganalisis langkah kerja sistem dalam mengklasifikasikan artikel berita berbahasa Indonesia, baik dari segi *preprocessing data, feature selection Chi Square* dan klasifikasi menggunakan metode *Multinomial Naive Bayes Classifier* serta pengimplementasiannya dalam menyelesaikan masalah yang diangkat.
2. Pengumpulan Data

Dataset yang digunakan penulis pada penelitian ini adalah artikel berita bahasa Indonesia yang diambil langsung dari halaman *website kompas.com, tribunnews.com, republika.com, mediaindonesia.com* dan *sindonews.com*. Artikel berita bahasa Indonesia yang diambil hanya judul artikel dan isi artikel berita.

### 3. Analisis Awal dan Perancangan Sistem

Analisis awal dilakukan dengan diskusi bersama dosen pembimbing untuk mengetahui kebutuhan sistem yang dibangun. Setelah itu, dilakukan perancangan sistem sesuai dengan kebutuhan dan masalah yang diangkat penulis dalam penelitian tugas akhir ini.

### 4. Implementasi Sistem

Melakukan implementasi terhadap rancangan sistem yang telah dibuat sebelumnya. Implementasi dilakukan dengan menggunakan bahasa pemrograman Python 2.7, IDE PyCharm Community Edition 2017.1.5.

### 5. Pengujian dan Analisa Hasil

Sistem yang telah dibangun dilakukan pengujian (*testing*). Sistem dievaluasi dengan data uji *fold ke-i* pada *k-fold cross validation*. Setelah dilakukan pengujian kemudian penulis melakukan analisis terhadap hasil performa sistem berupa nilai rata-rata *precision, recall, dan f1-measure* dari hasil pengujian dengan menggunakan *k-fold cross validation*.

### 6. Penarikan Kesimpulan

Pada tahap ini, penulis melakukan analisis terhadap hasil pengujian untuk mendapatkan kesimpulan dari kinerja sistem yang telah dibangun dan telah diimplementasikan oleh penulis.

### 7. Penyusunan Laporan Tugas Akhir

Penyusunan laporan tugas akhir ini dilakukan dengan mendokumentasikan hasil dari penelitian yang telah dilakukan dan melampirkan dokumen pendukung yang berkaitan dengan penelitian tugas akhir yang dilakukan oleh penulis.

## 1.5 Sistematika Penulisan

Penulisan laporan tugas akhir ini terbagi menjadi lima bab yang masing-masing akan diuraikan penulis sebagai berikut:

## **1. Pendahuluan**

Pada bab ini dijabarkan mengenai latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah dan sistematika penulisan.

## **2. Kajian Pustaka**

Pada bab ini terdapat ringkasan hasil kajian pustaka yang terkait dengan masalah yang diajukan penulis pada bab pendahuluan dan menjelaskan teori-teori pendukung yang digunakan untuk menyelesaikan masalah yang diajukan oleh penulis.

## **3. Perancangan Sistem**

Pada bab ini berisi rancangan sistem yang dibangun, pengumpulan dan ilustrasi pengolahan data yang dilakukan dalam membangun sistem pengklasifikasian artikel berita bahasa Indonesia serta skenario pengujian yang dilakukan.

## **4. Pengujian dan Analisis Sistem**

Pada bab ini membahas hasil pengujian berdasarkan metode dan skenario pengujian yang telah dituliskan pada bab Perancangan Sistem serta pada bab ini juga menjelaskan hasil analisis terhadap hasil pengujian yang telah dilakukan.

## **5. Kesimpulan dan Saran**

Pada bab ini menjelaskan kesimpulan dari keseluruhan hasil pengerjaan tugas akhir yang mengacu pada tujuan penelitian, skenario pengujian dan analisis hasil pengujian pada bab-bab sebelumnya serta memberikan saran untuk penelitian selanjutnya.