



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Teknologi merupakan salah satu sumber yang banyak diakses dan digunakan untuk memperoleh informasi terkait segala hal. Contohnya seperti melihat berita di televisi, menonton berita melalui streaming Youtube, dan membaca artikel berita yang tersedia di berbagai website yang tersedia di Internet. Hasil observasi yang dilakukan membuktikan bahwa teknologi informasi masih memiliki kesulitan dalam menyaring informasi yang diterbitkan, sehingga pengguna harus mengklasifikasikan dan merespon setiap informasi yang dibaca dari internet.

Semakin cepat berita tersebut dipublikasikan maka informasi dari berita tersebut akan tersebar dengan cepat. Berbagai kebutuhan informasi mulai dari politik, ekonomi, olahraga, bisnis hingga hiburan menjadi konsumsi yang penting di masyarakat. Sebuah dokumen berita seringkali terkait lebih dari satu kategori, untuk itu diperlukan pemanfaatan metode kategorisasi yang tidak hanya cepat tetapi juga dapat mengelompokkan sebuah berita ke dalam banyak kategori [1].

Klasifikasi teks adalah topik klasik untuk pemrosesan bahasa alami, di mana seseorang perlu menetapkan kategori yang telah ditentukan untuk dokumen teks bebas. Ada berbagai teknik untuk algoritma pembelajaran yang dibuat untuk menunjukkan kinerja yang wajar untuk klasifikasi teks. Hasil dari observasi saat ini, hampir semua teknik klasifikasi teks didasarkan pada kata-kata, di mana statistik sederhana dari beberapa kombinasi kata terurut biasanya berkinerja terbaik.

Kisaran penelitian klasifikasi teks berubah dari merancang fitur terbaik ke memilih klasifikasi pembelajaran mesin sebaik mungkin. Seperti yang kita ketahui banyak berita yang dipisahkan menurut kategorinya masing-masing sehingga mempermudah pembaca untuk menemukan topik apa yang ingin dibaca. Akan tetapi terkadang berita tersebut tidak hanya memiliki satu kategori saja. Dalam hal ini *text classification* berperan untuk menentukan kategori judul berita secara langsung berita yang telah ditulis dan diterbitkan tersebut masuk kategori mana saja.

Dalam kasus ini, penulis menggunakan algoritma Support Vector Machine (SVM) dibanding algoritma Naïve Bayes, K-Nearest Neighbors(KNN), dan Decion Tree yang dimana algoritma tersebut merupakan metode klasifikasi yang banyak digunakan dalam hal klasifikasi karena mudah untuk diinterpretasi oleh manusia. Naive Bayes memiliki kelebihan yaitu sederhana, cepat dan memiliki akurasi yang tinggi. SVM mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Nilai akurasi yang dihasilkan akan menjadi tolak ukur untuk menemukan pola uji terbaik untuk kasus klasifikasi sentimen [1]. K-Nearest Neighbor dapat melakukan pengklasifikasian data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan [3]. Sedangkan Decision Tree mendefinisikan berbagai alternatif yang memecahkan masalah dan faktor-faktor yang mungkin mempengaruhi alternatif bersama dengan perkiraan akhir jika memiliki alternatif lain [4].

Hasil penelitian menunjukkan nilai akurasi untuk algoritma Naive Bayes sebesar 84.50%. Sedangkan nilai akurasi algoritma Support Vector Machine (SVM)

lebih besar dari Naive Bayes yaitu sebesar 90.00%. [1]. Support Vector Machine (SVM) merupakan metode perhitungan yang baik di dalam mendapatkan hasil klasifikasi dengan tingkat akurasi tinggi [2]. Hasil penelitian nilai akurasi pada 266 data yang didapat dari aplikasi ini membantu memprediksi nasabah dalam menggunakan produk banknya, algoritma SVM memperoleh akurasi sebesar 69,55% yang lebih unggul dibandingkan dengan algoritma KNN yang memperoleh akurasi sebesar 49,62% [3]. Hasil penelitian lain tentang analisis sentimen review komentar pada aplikasi transportasi online, Decision tree memperoleh akurasi 89,80% sedangkan SVM memperoleh akurasi 90,20% [4]. Dari beberapa penelitian diatas, dapat menunjukkan bahwa Algoritma SVM memiliki akurasi yang cukup baik dalam hal klasifikasi dibandingkan beberapa algoritma lainnya.

Metode pembobotan kata berdasarkan kombinasi *term frequency-inverse document frequency* (TF-IDF) memberikan bobot lebih kepada istilah yang lebih penting. Efektivitas kinerja mesin pencari selalu dikaitkan dengan tingkat relevansi hasil pencarian. Untuk menemukan dokumen yang relevan, metode pembobotan TF-IDF memberikan bobot lebih pada istilah yang lebih penting. Istilah yang lebih penting yang dimaksud adalah istilah yang jika muncul dalam suatu dokumen maka dokumen tersebut dapat dianggap relevan dengan *query* [3].

## 1.2. Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalahnya adalah :

1. Bagaimana cara implementasi Algoritma Support Vector Machine dalam mengklasifikasikan kategori dari judul berita?

2. Berapa nilai *precision*, *recall*, *accuracy*, dan *F1-Score* yang dihasilkan oleh metode Support Vector Machine untuk klasifikasi judul berita?

### 1.3. Batasan Masalah

Batasan masalah dalam merancang dan membangun aplikasi implementasi ini adalah:

1. Judul berita menggunakan bahasa Indonesia.
2. Pengaplikasian Text Classification ini hanya diklasifikasikan kedalam 6 topik, antara lain *finance*, *food*, *health*, *otomotif*, *sport*, dan *travel* yang bersumber dari portal berita *detik.com*.
3. Total data yang diambil untuk melakukan penelitian ini adalah 6000 judul berita, terdiri dari 1000 judul berita untuk masing-masing kategori.
4. Rentan waktu yang diperoleh dalam *dataset* adalah tanggal 01 Januari 2020 sampai tanggal 13 Juni 2020 yang dibuat oleh Ibrahim pada tahun 2020.

### 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan, maka tujuan dari penelitian ini adalah :

- a) Mengimplementasikan Algoritma Support Vector Machine untuk klasifikasi judul berita.
- b) Mendapatkan atau menghitung nilai *precision*, *recall*, *accuracy*, dan *F1-Score* pada metode Support Vector Machine untuk klasifikasi judul berita.

- c) Hasil dari penelitian yang dilakukan dapat diterapkan penggunaanya pada mesin, seperti mensortir berita yang diambil dari portal berita.

### 1.5. Manfaat Penelitian

Manfaat yang diharapkan dari pelaksanaan penelitian ini adalah :

- a) Mempersingkat dan mempermudah proses pengkategorian judul berita secara cepat dan akurat secara otomatis serta referensi dalam melakukan penelitian terkait.
- b) Menambah wawasan dan referensi penelitian yang berkaitan dengan topik *text classification* dengan menggunakan Algoritma SVM dan ekstraksi fitur TF-IDF.
- c) Menyediakan aplikasi berbasis web dengan tujuan untuk mempermudah masyarakat umum dalam menentukan kategori berita hanya dengan menggunakan judul berita.

### 1.6. Sistematika Penulisan

Sistematika penulisan pada skripsi ini dibagi menjadi lima bagian, diantaranya sebagai berikut :

#### 1. BAB 1 PENDAHULUAN

Pada bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

## 2. BAB 2 LANDASAN TEORI

Pada bab ini berisi mengenai landasan-landasan teori yang digunakan dalam penelitian ini.

## 3. BAB 3 METODOLOGI PENELITIAN

Pada bab ini berisi tentang metode yang digunakan dalam melakukan penelitian ini, serta tahap-tahap dalam proses perancangan aplikasi.

## 4. BAB 4 HASIL DAN DISKUSI

Pada bab ini berisi tentang spesifikasi sistem, implementasi sistem, hasil uji coba dari penelitian yang dilakukan.

## 5. BAB 5 SIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan yang menjawab tujuan penelitian dan saran untuk membantu penelitian yang akan dilakukan lebih lanjut.

