

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri berita mengalami perubahan dan inovasi pada model bisnis yang lebih efektif dan efisien merupakan hasil dari Revolusi industri 4.0 [1]. Revolusi industri mengakibatkan perusahaan berita melakukan berbagai inovasi untuk produktivitas dan meningkatkan daya saing perusahaan, diantaranya pada sisi teknologi. Inovasi itu adalah penerapan *Artificial Intelligence*, *Machine Learning* untuk menangani data teks berita yang beragam dan berjumlah besar (Big Data).

Pada umumnya berita yang disajikan dalam situs berita terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi dan lain sebagainya. Namun, dalam proses membagi berita ke dalam kategori-kategori tersebut sejauh ini dilakukan secara manual oleh editor[2]. Seiring dengan meningkatnya jumlah pengguna internet di Indonesia, jumlah berita semakin meningkat. Data yang begitu banyak dan tidak terstruktur sulit untuk diorganisasikan atau dikelompokkan kedalam kategori yang tepat[3]. Permasalahan lain muncul ketika data yang akan dikelompokkan dalam masing-masing kategori memiliki kemiripan isi. Hal ini membutuhkan ketelitian dan waktu yang tidak sebentar[2]. Oleh karena itu, perlu adanya peran *text mining* yang dapat mengklasifikasikan berita sesuai dengan kategori yang sebenarnya.

Text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dimana, *text mining* merupakan variasi dari data mining

yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[4].

Untuk mengatasi permasalahan diatas, dilakukan penelitian untuk mengklasifikasikan teks konten berita menggunakan algoritma klasifikasi *Machine learning*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka didapatkan rumusan masalah yaitu :

1. Bagaimana cara mengklasifikasikan teks konten berita menggunakan algoritma klasifikasi *Machine learning* ?
2. Bagaimana kinerja *Support Vector Machine (SVM)* dalam mengklasifikasikan teks berita?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Algoritma klasifikasi *machine learning* yang digunakan pada penelitian ini adalah *Support Vector Machine (SVM)*.
2. Data teks berita yang digunakan berbahasa Indonesia.
3. Sumber dataset terdiri dari 3 kelas atau 3 kategori (politik, kuliner, sepakbola).

1.4 Maksud dan Tujuan Penelitian

1. Untuk mengetahui bagaimana cara mengklasifikasikan teks berita menggunakan algoritma klasifikasi *Machine Learning*.
2. Untuk mengetahui tingkat akurasi metode *Support Vector Machine (SVM)* dalam mengklasifikasikan teks berita.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat untuk :

1. Hasil penelitian dapat dijadikan referensi bagi para peneliti berikutnya yang akan membahas mengenai klasifikasi data teks.
2. Mengetahui cara implementasi dan akurasi algoritma *Support Vector Machine (SVM)* pada kasus klasifikasi data teks berita.
3. Memperoleh informasi dan menambah pengetahuan dalam mengelola data teks berita untuk melakukan klasifikasi.
4. Untuk kebutuhan jurnalistik.

1.6 Metode Penelitian

1.6.1 Pengumpulan Data

a. Studi Pustaka

Adapun metode yang digunakan adalah studi pustaka. Studi pustaka merupakan teknik pengumpulan data dengan berbagai buku, literatur, catatan, serta berbagai laporan yang berkaitan dengan masalah yang ingin dipecahkan.

b. Sumber Dataset

Dataset yang akan dipakai pada penelitian ini berupa data teks konten berita dari situs media CNN Indonesia.

1.6.2 Melakukan Preprocessing Data

Yaitu casefolding, penghapusan stopword, cleaning, filtering, stemming, dan tokenizing.

1.6.3 Melakukan Klasifikasi Menggunakan SVM

Pada tahap ini Membagi data menjadi data training dan data testing, pembobotan kata menggunakan TF-IDF, membangun model menggunakan SVM dengan fungsi kernel.

1.6.4 Pengujian Hasil

Evaluasi performa model klasifikasi menggunakan *Confusion Matrix* dan Validasi model menggunakan *K-Fold Cross Validation*.

1.7 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini merupakan pengantar pokok permasalahan dan gambaran penelitian secara keseluruhan, adapun hal-hal yang dibahas adalah latar belakang, rumusan masalah, maksud dan tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini memuat dasar-dasar teori yang digunakan dalam penelitian ini, juga menjelaskan tentang teori-teori yang membahas secara mendetail.

BAB III METODE PENELITIAN

Bab ini Menguraikan sumber data yang dipakai dalam penelitian, alur penelitian.

BAB IV HASIL DAN PEMBAHASAN

Membahas tentang implementasi dan hasil dari *pre-processing* data, TF-IDF, model klasifikasi serta pembahasannya.

BAB V PENUTUP

Bab ini berisikan tentang kesimpulan penelitian, fakta-fakta, yang didapatkan selama melaksanakan penelitian, dan saran yang diberikan berdasarkan hasil-hasil dari pelaksanaan penelitian, agar pengembangan sistem yang sejenis dengan penelitian ini dapat memahami gambaran penelitian yang telah dilakukan secara keseluruhan, sehingga dapat memperbaiki bagian-bagian yang masih terdapat kekurangan.

