



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 I

PENDAHULUAN

1.1. Latar Belakang

Informasi menjadi kebutuhan pokok bagi setiap orang, namun tidak semua informasi yang ada dapat menjadi kebutuhan. Dipengaruhi oleh kemajuan teknologi internet sehingga informasi mengalami pelonjakan yang besar, karena kebutuhan masyarakat akan berita semakin meningkat setiap harinya, berbagai pihak mencoba untuk menyajikan sajian berita berbahasa Indonesia yang sesuai dengan kebutuhan masyarakat Indonesia secara cepat, tepat, akurat, dan terpercaya bagi para pembaca berita. Berita dapat diperoleh dari media cetak maupun elektronik seperti koran, televisi, radio, dan internet. Berita yang disajikan dalam bentuk teks pada media elektronik, biasanya dikelompokkan berdasarkan isinya seperti berita olahraga, ekonomi, sains, dan lain sebagainya.

Permasalahan yang muncul adalah penggunaan media digital dalam penyampaian informasi menyebabkan jumlah artikel berita digital yang dirilis oleh portal berita tiap harinya menjadi sangat banyak. Hal ini berdampak pada ketersediaan artikel berita yang jumlahnya sangat melimpah. Berdasarkan permasalahan tersebut diperlukan adanya pengorganisasian dokumen artikel berita. Salah satu cara yang dapat dilakukan dengan cepat dan dapat dipahami oleh para penerima informasi adalah dengan melakukan klasifikasi dokumen artikel berita berdasarkan topiknya. Klasifikasi merupakan salah satu metode dalam *text mining* yang bertujuan untuk menentukan label atau kategori objek (Prasetyo, 2012).

Text mining merupakan sebuah proses pengetahuan intensif di mana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis (Feldman & Sanger, 2007). *Text mining* mencoba untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi dari suatu pola menarik. Sumber data berupa sekumpulan dokumen dan pola menarik yang tidak ditemukan dalam bentuk *database record*, tetapi dalam data teks yang tidak terstruktur.

Untuk mengetahui kategori pada sebuah artikel berita secara otomatis, diperlukan pengukuran kemiripan dokumen terkait dengan menggunakan metode *Naïve Bayes Classifier* (NBC) (Saepudin, 2013). NBC adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah kelas, dan kelebihan dari metode ini adalah tingkat akurasi yang tinggi juga waktu komputasi yang lebih cepat (Destuardi & Surya, 2009).

Web scraping merupakan teknik yang digunakan untuk mengekstrak sejumlah besar dokumen artikel berita dari situs web portal berita. Data yang sudah diekstraksi disimpan ke sebuah *file* lokal di komputer atau ke basis data dalam format tabel (*spreadsheet*). Inilah yang memungkinkan *user* untuk mengeksplorasi isi dari situs web tanpa mengunjungi situs web yang bersangkutan, sehingga *user* bisa melakukan berbagai bentuk analisis semantik tanpa mengganggu *resource* situs web yang bersangkutan (Haddaway, 2015).

Beberapa penelitian yang berkaitan dengan klasifikasi *naïve bayes* di antaranya implementasi *Naïve bayes classifier* pada program bantu penentuan buku referensi mata kuliah menghasilkan nilai akurasi 69% (Nurani et al, 2007). Klasifikasi emosi untuk teks bahasa Indonesia menggunakan metode *Naïve bayes*

menghasilkan akurasi nilai maksimum 60,45 dengan *precision* dan *recall* (Destuardi & Sumpeno, 2009). Klasifikasi dokumen teks berbahasa Indonesia dengan menggunakan *naïve bayes* menghasilkan akurasi 83,57% dan terus meningkat hingga 87,63% (Samodra et al, 2009). Klasifikasi dokumen menggunakan algoritma *naïve bayes* dengan penambahan parameter probabilitas *parent category* menghasilkan 61,77% untuk klasifikasi menggunakan *Naïve bayes* dan *parent category* sedangkan 60,49% menggunakan *Naïve bayes* saja (Trisedya & Jai, 2009). Klasifikasi teks dengan *Naïve Bayes Classifier* (NBC) untuk pengelompokan teks berita dan abstrak akademis menghasilkan akurasi yang lebih tinggi maksimal 91% dibandingkan dengan dokumen akademik maksimal 82% (Hamzah, 2012).

Berdasarkan uraian, maka disusun proyek skripsi yang berjudul “Implementasi Algoritma Naïve Bayes Classifier Untuk Klasifikasi Kategori Berita Pada Website Cnnindonesia.Com”, diharapkan dapat membantu dalam mengategorikan berita.

1.2. Rumusan Masalah

Berdasarkan kebutuhan yang dijelaskan sebelumnya, terdapat rumusan masalah yaitu , Bagaimana mengklasifikasi berita menggunakan metode klasifikasi *Naïve Bayes Classifier*? Dan berapakah akurasi yang dapat dicapai oleh aplikasi?

1.3. Batasan Masalah

Berdasarkan rumusan masalah yang telah ditetapkan, maka batasan masalah penelitian ini adalah sebagai berikut.

1. *Dataset* berita pelatihan dari artikel berita bahasa Indonesia disimpan ke dalam format csv.
2. Berita yang dibutuhkan untuk pelatihan sebanyak 396 berita yang diperoleh dari portal pada tahun 2018 -2019.
3. Penentuan kategori dokumen berita berupa olahraga, teknologi, ekonomi, dan politik.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini ialah:

1. Mengimplementasikan klasifikasi berita menggunakan metode *Naïve Bayes Classifier*.
2. Mencapai performa akurasi dari klasifikasi yang tinggi menggunakan metode *Naïve Bayes Classifier*.

1.5. Manfaat Penelitian

Dengan dilakukannya penelitian ini diharapkan mampu memberikan manfaat secara keilmuan berupa kontribusi pada klasifikasi berita menggunakan metode *Naïve Bayes Classifier*, sedangkan manfaat kepada masyarakat adalah memberikan kemudahan dalam proses klasifikasi berita terutama untuk penyedia media online di Indonesia.

1.6. Sistematika Penulisan

Sistematika penulisan laporan skripsi ini dijabarkan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang latar belakang penelitian yang dilakukan, membatasi masalah, perumusan masalah, memaparkan tujuan dan manfaat penelitian yang baik dilakukan untuk pihak eksternal maupun bagi penulis sendiri, serta sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan dibahas mengenai dasar-dasar teori, baik dari buku, jurnal maupun literatur lainnya, serta konsep secara tertulis yang berkaitan dengan permasalahan yang dikemukakan dalam penelitian.

BAB III METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM

Pada bab ini menjelaskan tentang metode penelitian, dan perancangan sistem yang akan dibuat pada penelitian ini.

BAB IV IMPLEMENTASI SISTEM

Bab ini terdiri dari hasil implementasi sistem yang telah dibuat oleh penulis serta pengujian terhadap sistem yang telah dibuat.

BAB V SIMPULAN DAN SARAN

Bab terakhir ini berisikan atas kesimpulan, keterbatasan dan saran yang berdasarkan hasil penelitian.