

1. Pendahuluan

Pada bagian ini berisi empat sub-bagian yaitu: Latar Belakang, Topik dan Batasannya, Tujuan dan Organisasi Tulisan. Di bawah ini akan dijelaskan dari masing-masing sub-bagian tersebut.

Latar Belakang

Berita adalah informasi yang dibuat untuk melaporkan kejadian atau peristiwa yang terjadi agar semua orang tahu kondisi selain di sekitarnya [1]. Perkembangan jaman sudah sangat maju, media penyampaian berita tidak hanya media cetak saja tetapi juga ada televisi, radio dan juga internet [2]. Walaupun banyak media penyampaian berita tetapi media cetak tetap ada sampai saat ini karena jika dibandingkan dengan Internet, media cetak jauh lebih terbukti kebenarannya karena sumber beritanya jelas tidak seperti berita di Internet yang semua orang bisa menerbitkan berita kapan pun. Kelemahan media cetak ada di segi lingkungan. Media cetak tidak ramah lingkungan karena memakai kertas yang merupakan hasil pengolahan kayu untuk mencetak berita. Jika dibandingkan dengan media elektronik seperti radio dan televisi mempunyai keunggulan dapat memvisualisasikan berita sehingga penikmat berita lebih memahami isi berita daripada media cetak yang hanya berupa tulisan tapi kekurangannya adalah penayangan beritanya hanya sekali tidak bisa diulang.

Perkembangan internet di Indonesia sangat pesat, berdasarkan survei yang dilakukan oleh Asosiasi Penyedia Jasa Internet Indonesia (APJII) pengguna internet di Indonesia terus meningkat dari tahun 1998 sampai 2017. Di tahun 2017 pengguna internet Indonesia mencapai 143,26 juta pengguna, meningkat 7% dari tahun 2016 yang penggunanya terdapat 132,7 juta pengguna. Indonesia merupakan negara dengan pengguna internet terbesar kelima di dunia setelah China, India, United States, Brazil [3], oleh karena itu media penyampaian berita melalui internet berpotensi besar menguasai pasar. Banyak berita yang diterbitkan dari internet karena target pembacanya lebih tinggi daripada melalui media cetak. Penerbit berlomba-lomba untuk menghasilkan berita yang terbaru, hal tersebut dapat menimbulkan masalah dalam mengkategorikan berita. Contohnya saya ingin membaca berita tentang olahraga, tetapi yang muncul pada kategori olahraga ternyata berita politik. Untuk mengurangi kesalahan pengkategorian perlu adanya teknik klasifikasi yang dapat mengkategorikan berita secara otomatis.

Teknik klasifikasi adalah suatu cara mengenali data-data baru berdasarkan ciri-cirinya dengan data latih yang sudah terdapat pelabelannya [4]. Teknik pengklasifikasian berita berdasarkan kategorinya dengan cara program mengecek isi berita berdasarkan per kata, dalam kata-kata tersebut mengandung ciri-ciri yang merujuk ke kategori tertentu. Dengan pemilihan metode klasifikasi yang tepat sesuai karakteristik data tersebut maka akan diperoleh hasil akurasi yang tinggi. Dalam penelitian ini, penulis memilih metode klasifikasi *decision tree* karena metode ini dapat menghasilkan akurasi sebesar 93% pada studi kasus teks Arab dengan *dataset scientific corpus* [5].

Hasil pengukuran tingkat keberhasilan klasifikasi untuk memprediksi data adalah akurasi, semakin tinggi akurasi maka semakin banyak data yang diprediksi benar. Untuk meningkatkan akurasi metode klasifikasi *decision tree*, penulis mengkombinasikan teknik klasifikasi dengan teknik pembobotan kata, karena pembobotan kata sangat penting untuk mengetahui seberapa berpengaruhnya suatu kata dalam kategori tertentu, contohnya pada sebuah dokumen mengandung kata sepak bola. Tentunya kata sepak bola harus memiliki bobot yang tinggi pada kategori olahraga.

Berdasarkan penelitian sebelumnya dengan membandingkan pembobotan *Term Frequency Absolute* (TF.ABS) dengan *Term Frequency Chi Square* (TF.CHI²) menghasilkan akurasi yang tidak jauh berbeda yaitu 95,74% untuk TF.ABS dan akurasi sebesar 95,87% untuk TF.CHI² [6]. Penelitian lain ada yang membandingkan antara TF.CHI², *Term Frequency Inverse Document Frequency* (TF.IDF) dan *Term Frequency Relevance Frequency* (TF.RF) diperoleh hasil TF.CHI² menghasilkan akurasi 89,71%, TF.RF dengan akurasi 88,07% dan TF.IDF hanya mendapat akurasi sebesar 74,93% [7]. Oleh karena itu penulis mengambil parameter TF.ABS, TF.CHI², TF.RF, TF.IDF untuk mengetahui akurasi mana yang akan menghasilkan akurasi yang optimal jika dikombinasikan dengan teknik klasifikasi *decision tree*.

Topik dan Batasannya

Topik yang diangkat dalam penelitian ini adalah membandingkan beberapa teknik pembobotan dengan mengkombinasikan teknik klasifikasi *Decision Tree* untuk mengklasifikasikan teks berita berdasarkan kategorinya. Berdasarkan topik di atas, terdapat beberapa batasan masalah yaitu:

1. Kategori berita terdiri dari 12 kelas yaitu Ekonomi, Hiburan, Hukum, Kesehatan, Gaya Hidup, Otomotif, Pendidikan, Politik, Sosial Budaya, Olahraga, Teknologi Wisata.
2. Jumlah dokumen *dataset* sebanyak 360 dokumen.
3. Proses klasifikasi menggunakan *unigram*.
- 4.

Tujuan

Tujuan dari penelitian ini untuk mengetahui teknik pembobotan apa yang menghasilkan akurasi optimal menggunakan klasifikasi *decision tree*.

Organisasi Tulisan

Penelitian ini ditulis menjadi beberapa bagian. Bagian kedua merupakan studi literatur yang terkait dengan penelitian yang dikerjakan dan menjelaskan penelitian yang sudah pernah dilakukan sebelumnya. Sistem yang akan dibangun untuk menentukan teknik pembobotan mana yang menghasilkan akurasi tinggi dijelaskan pada bagian ketiga. Bagian keempat merupakan evaluasi yang berisi hasil pengujian serta analisis pengujian pada sistem. Bagian selanjutnya adalah penarikan kesimpulan dan pemberian saran.