

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi di era digital berkembang sangat pesat dan memiliki dampak yang positif dalam kehidupan sehari-hari termasuk dalam dunia pendidikan tinggi. Berkembangnya teknologi menuntut mahasiswa untuk memiliki pengetahuan dan skill yang baik untuk menghadapi era perkembangan zaman (Zhafira dkk., 2021). Oleh karena itu, Kementerian Pendidikan dan Kebudayaan (Kemendikbud) meluncurkan program kampus merdeka untuk menjawab kebutuhan tersebut. Perguruan Tinggi didorong untuk membuat kurikulum yang inovatif dan adaptif bagi mahasiswa untuk menunjang program merdeka belajar kampus merdeka. Program ini memberikan kesempatan kepada mahasiswa untuk mengembangkan kemampuan sesuai minat dan bakat untuk mempersiapkan diri mereka terjun ke dunia karir setelah lulus kuliah.

Kebijakan MBKM ini banyak menuai opini pro dan kontra di masyarakat sejak awal kemunculannya. Banyak yang mengkritisi pelaksanaan dari program MBKM ini melalui media sosial. Media sosial merupakan platform yang paling banyak digunakan dan menjadi tempat untuk menyampaikan pendapat. *Twitter* merupakan salah satu media sosial yang mewadahi opini masyarakat tentang banyak topik. Banyaknya pendapat dan tanggapan yang muncul dari masyarakat mengenai kebijakan MBKM di media sosial terutama *Twitter*, menjadi suatu sentimen yang dialami oleh masyarakat (Rozaq dkk., 2022). Analisis sentimen terhadap program MBKM perlu dilakukan untuk mengetahui kepuasan masyarakat terhadap program MBKM (Rozaq dkk., 2022).

Sentimen terhadap program MBKM melalui cuitan di *Twitter* terus bertambah setiap hari. Dalam waktu sekejap ratusan bahkan ribuan cuitan mengenai MBKM bisa memenuhi jagat *Twitter*. Setiap cuitan di *Twitter* perlu diklasifikasikan dan dianalisis untuk keperluan evaluasi dari suatu kebijakan (Naufal & Kusuma., 2022). Pengklasifikasian cuitan berdasarkan sentimen secara manual akan membutuhkan waktu yang lama. Dengan jumlah data yang sangat besar diperlukan teknologi untuk pengolahan data. Oleh karena itu perlu pendekatan *machine learning* untuk melakukan klasifikasi sentimen (Zhafira dkk., 2021).

Salah satu cara untuk menganalisis sentimen adalah dengan menggunakan menggunakan metode *supervised learning*, yang melibatkan proses pembelajaran mesin dengan menggunakan dataset yang sudah diberi label. *Supervised Learning* merupakan teknik dalam menciptakan kecerdasan buatan, yang bertujuan untuk mengenali pola – pola pada kumpulan data yang belum diberi label atau tidak terklasifikasi (Atmevia dkk., 2023). Kinerja metode *supervised learning* sangat bergantung pada kualitas *data training* dan cara data tersebut diolah. Salah satu algoritma yang sering digunakan untuk menganalisis sentimen dengan metode *supervised learning* adalah *Random Forest*.

Penelitian sebelumnya mengenai pengklasifikasian sentimen pada data *Twitter* menggunakan metode *Multinomial Naïve Bayes* menunjukkan hasil klasifikasi yang kurang baik dan penulis tersebut mengusulkan untuk menggunakan metode klasifikasi yang lain salah satunya *Random Forest* (Rahman, Umar S., 2020). *Random Forest* merupakan salah satu algoritma *machine learning* yang terkenal efektif dalam pengklasifikasian teks yang mana pengklasifikasiannya itu didasarkan pada suara terbanyak dari klasifikasi *decision tree* (Ramli, R. G., & Sibaroni, Y., 2022). Algoritma *Random Forest* memiliki performa yang cukup baik dalam kasus pengklasifikasian teks berdasarkan penelitian sebelumnya mengenai analisis sentiment komentar di *YouTube* dengan nilai *accuaracy* 79% dan *F1-Score* 86,26% (Afdhal dkk., 2022).

Beberapa penelitian terkait analisis sentimen menggunakan Algoritma *Random Forest* belum melibatkan perbandingan pembobotan fitur dalam penilaian kinerjanya. Pembobotan fitur dilakukan karena komputer hanya dapat memproses input yang numerik (Kwon dkk., 2018). Pembobotan fitur dapat mempengaruhi kinerja klasifikasi (Sabri dkk., 2021). Klasifikasi menggunakan *machine learning* yang dikombinasikan dengan pembobotan fitur yang tepat akan menghasilkan performa yang baik (Zhang dkk., 2018). Berdasarkan kondisi ini dinilai perlu dilakukan penelitian yang dapat meningkatkan kinerja klasifikasi pada Algoritma *Random Forest*.

Pada penelitian ini penulis melakukan komparasi pembobotan fitur *Bag of Words* dan *TF-IDF* yang dikombinasikan dengan Algoritma *Random Forest*. *Bag of Words* dan *TF-IDF* keduanya adalah teknik yang umum digunakan dalam pemrosesan bahasa alami untuk

mengubah teks menjadi representasi numerik yang dapat digunakan oleh model pembelajaran mesin. Namun, keduanya memiliki pendekatan yang berbeda dalam membangun representasi tersebut. Pada intinya model *Bag of Words* mengubah dokumen teks menjadi vektor, di mana setiap dokumen menghasilkan vektor yang menunjukkan kemunculan semua istilah yang berbeda di dalam ruang vektor dokumen untuk dokumen tersebut (Canedo & Mandes., 2020). Sedangkan *TF-IDF* adalah sebuah kalkulasi statistik yang dapat mengukur relevansi setiap kata dalam dokumen dalam suatu kumpulan dokumen (Gutam & Bansal., 2022). Penelitian ini bertujuan memilih pembobotan fitur yang sesuai dengan Algoritma *Random Forest* melalui proses komparasi. Hasil komparasi akan menentukan pembobotan fitur yang menghasilkan kinerja klasifikasi yang paling optimal dengan beberapa split data yang berbeda.

Sebelum melakukan klasifikasi, data akan melalui tahap praproses dan seleksi fitur. Praproses data dilakukan agar data menjadi terstruktur (Pratomo dkk., 2021). Pada tahap ini, semua huruf kapital diubah menjadi huruf kecil dan semua karakter yang kurang penting dihapus, Tahap selanjutnya adalah seleksi fitur. Metode seleksi fitur yang digunakan dalam penelitian ini adalah *Chi-Square* (χ^2). Penelitian yang dilakukan oleh Burhanudin (2022) menemukan bahwa *Chi-Square* (χ^2) merupakan metode seleksi fitur yang efektif. Setelah tahap praproses dan seleksi fitur, data akan diklasifikasikan lalu performa klasifikasi akan diukur menggunakan perhitungan nilai *accuracy*, *precision*, *recall* dan *F1-score*.

1.2 Rumusan Masalah Penelitian

1. Bagaimana implementasi teknik *feature selection* menggunakan metode *Chi-Square* untuk mengklasifikasikan sentimen program MBKM pada data *Twitter*?
2. Bagaimana implementasi pembobotan fitur menggunakan *Bag of Words* dan *TF-IDF* untuk mengklasifikasikan sentimen program MBKM pada data *Twitter*?
3. Bagaimana perbandingan performa dari klasifikasi sentiment program MBKM pada data *Twitter* menggunakan *Random Forest* dengan *Bag of Words* dan *TF-IDF*?

1.3 Tujuan Penelitian

1. Mengimplementasikan teknik *feature selection* menggunakan metode *Chi-Square* untuk mengklasifikasikan sentimen program MBKM pada data *Twitter*.

2. Mengimplementasikan pembobotan fitur menggunakan *Bag of Words* dan *TF-IDF* untuk mengklasifikasikan sentimen program MBKM pada data *Twitter*.
3. Mengetahui perbandingan performa dari klasifikasi sentiment program MBKM pada data *Twitter* menggunakan *Random Forest* dengan *Bag of Words* dan *TF-IDF*

1.4 Manfaat Penelitian

1. Dapat memberikan pengetahuan mengenai dampak penggunaan teknik *feature selection* dengan metode *Chi-Square* untuk mengklasifikasikan data teks.
2. Dapat memberikan pengetahuan mengenai dampak penggunaan *Bag of Words* dan *TF-IDF* untuk pembobotan fitur.

1.5 Batasan Masalah

1. Data yang digunakan dikumpulkan dalam rentang waktu antara tanggal 14 April 2023 dan 12 Juni 2023
2. Data yang digunakan adalah *tweet* dari *platform Twitter* yang mengandung kata kunci “mbkm”, “msib”, “kampus merdeka” dan “merdeka belajar”.

1.6 Sistematika Penulisan

1. BAB I PENDAHULUAN

Pada bab pendahuluan dijelaskan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian dan sistematika penulisan

2. BAB II KAJIAN PUSTAKA

Berisi teori – teori yang berkaitan dengan klasifikasi teks dan metode klasifikasi dengan algoritma *Random Forest*

3. BAB III METODOLOGI PENELITIAN

Berisi mengenai sumber data yang digunakan pada penelitian dan tahap penelitian mulai dari tahap awal sampai tahap analisis

4. BAB IV HASIL DAN PEMBAHASAN

Berisi mengenai hasil implementasi pembobotan fitur *Bag of Words* dan *TF-IDF* pada klasifikasi sentimen program MBKM

5. BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya