

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Pendidikan memiliki peran penting dalam meningkatkan kualitas dan kecerdasan bangsa. Pada jenjang Pendidikan tinggi, proses pembelajaran harus dilakukan secara terpusat pada mahasiswa (*student centered learning*) dengan tujuan agar dapat menghasilkan lulusan yang profesional di bidangnya serta berharga bagi masyarakat dan negaranya yang mampu beradaptasi terhadap kemajuan dan perkembangan ilmu pengetahuan dan teknologi (L. W. Santoso & Yulia, 2020). Dalam mempersiapkan mahasiswa menghadapi hal tersebut, pemerintah Indonesia selalu berusaha untuk memberikan inovasi dan terobosan baru dalam meningkatkan kualitas pendidikan, salah satunya adalah dengan merancang sebuah program Merdeka Belajar-Kampus Merdeka (MBKM) yang merupakan salah satu perwujudan pembelajaran yang terpusat pada mahasiswa dengan memberikan “kemerdekaan” kepada mahasiswa untuk mencari dan menemukan pengetahuan melalui kenyataan dan dinamika lapangan yang sesuai dengan minat dan bakat mahasiswa (Dirjen Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan, 2020)

Program MBKM dapat dikatakan sebagai program yang relatif baru, diterapkan di Indonesia sejak Februari 2020 lalu. Layaknya program baru, implementasi program MBKM secara tidak langsung juga diikuti oleh berbagai opini pro dan kontra dari berbagai kalangan masyarakat. Secara umum, masyarakat yang pro terkait dengan program MBKM merasa bahwa program ini merupakan inovasi dalam mempersiapkan sumber daya manusia yang berkualitas dan berkarakter, sehingga dapat membantu dalam menghasilkan lulusan yang lebih bermutu. Sedangkan, masyarakat yang kontra terhadap program MBKM merasa bahwa program MBKM masih belum memiliki peraturan dan dasar yang jelas, salah satunya terkait dengan sistem konversi SKS mahasiswa, sehingga perlu diberikan indikator yang jelas (Dasrimin, 2021).

Opini yang dilontarkan oleh masyarakat terhadap implementasi program MBKM berasal dari berbagai sumber, salah satunya adalah media sosial Twitter.

Twitter merupakan salah satu media sosial yang sering digunakan dalam mengungkapkan perasaan berupa pujian dan kritik dalam berbagai ranah, seperti sarana protes, sarana pembelajaran serta sebagai sarana pemberian informasi atau pendapat masyarakat terhadap berbagai isu dalam bentuk tulisan status atau *tweet* (Sarasvananda dkk., 2022). Twitter sendiri menyediakan sebuah *Application Programming Interface* (API) yang bersifat *open source* sehingga proses pengambilan atau *crawling* data menjadi sangat mudah yang menjadikan salah satu kelebihan Twitter dibandingkan dengan media sosial lainnya (Pintoko & L., 2018). Beberapa penelitian yang berhasil menggunakan Twitter sebagai sumber *dataset* yakni penelitian yang dilakukan oleh Singh dan Supta (2020), Masturoh dan Pohan (2021), dan Syahputra (2021). Hal tersebut menunjukkan bahwa media sosial Twitter memiliki peluang yang besar untuk dapat dimanfaatkan sebagai media dalam mengumpulkan data opini masyarakat.

Adanya opini atau umpan balik dari masyarakat tentu dapat dijadikan sebagai bahan evaluasi dan pertimbangan terhadap implementasi program di masa mendatang. Oleh karena itu, diperlukannya sebuah analisis sentimen terhadap program MBKM. Analisis sentimen merupakan salah satu bagian kegiatan dari *text mining* yang bertujuan untuk menganalisis data tekstual yang berupa sebuah opini yang diungkapkan oleh individu maupun kelompok terhadap suatu topik, sehingga dapat ditemukan sebuah informasi yang bermanfaat (Que dkk., 2020). Melalui analisis sentimen, opini atau pendapat khalayak umum yang bersifat tidak terstruktur dan dalam jumlah besar dapat dianalisis dan dikelompokkan menjadi data yang terstruktur dengan mudah (Oktaviani dkk., 2021). Beberapa metode yang sering digunakan dalam analisis sentimen adalah *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *K-Nearest Neighbor* dan lainnya.

Penelitian terkait analisis sentimen sudah banyak dilakukan sebelumnya, beberapa diantaranya adalah penelitian oleh Nurjanah dkk. (2017) terkait analisis sentimen terhadap tayangan televisi menggunakan metode *K-Nearest Neighbor* yang menghasilkan tingkat akurasi mencapai 80.83%. Penelitian lainnya dilakukan oleh Salam dkk. (2018) yang meneliti sentimen masyarakat terkait jasa ekspedisi barang J&T Ekspres Indonesia menggunakan metode *K-Nearest Neighbor* yang menghasilkan tingkat akurasi sebesar 79.21%. Penelitian terkait juga pernah

dilakukan oleh Pertiwi (2019) yang membandingkan antara metode *Naïve Bayes*, *Neural Network*, KNN dan SVM dalam melakukan sentimen opini masyarakat terhadap sarana dan transportasi mudik tahun 2019, dengan hasil metode KNN dapat memberikan nilai akurasi yang paling tinggi, yakni sebesar 90.76% dibandingkan metode lainnya.

Berdasarkan beberapa penelitian sebelumnya, metode KNN terbukti memiliki performa yang baik dalam melakukan klasifikasi data sentimen atau opini masyarakat. Oleh karena itu, pada penelitian ini akan digunakan metode KNN dalam melakukan klasifikasi sentimen. KNN adalah sebuah metode pembelajaran mesin terawasi berbasis memori (*memory-based supervised machine learning method*) yang melakukan proses klasifikasi dengan cara mencari data latih yang memiliki kemiripan dengan data uji berdasarkan jarak ketetanggaan terdekat (*nearest neighbor*) (Mwinzi dkk., 2022; Pajri dkk., 2020). Kelebihan dari metode KNN adalah metode ini dapat bekerja secara sederhana dan mudah untuk diimplementasikan terhadap data yang memiliki jumlah besar dan tangguh terhadap data latih yang *noise* (Mahardika dkk., 2018).

Dalam *machine learning*, permasalahan yang sering terjadi adalah adanya data yang tidak seimbang (*imbalance data*). Data tidak seimbang terjadi ketika terdapat sebuah kelas minoritas dan mayoritas pada *dataset*. Kelas mayoritas merupakan sebuah kelas yang memiliki jumlah data lebih banyak dibanding dengan kelas lainnya. Begitupun sebaliknya, kelas minoritas merupakan kelas yang memiliki jumlah data paling sedikit dibanding kelas lainnya, sehingga penyebaran jumlah masing-masing tidak seimbang (Brownlee, 2020). Data tidak seimbang berpotensi menyebabkan kesalahan klasifikasi pada kelas minoritas. Kelas minoritas akan cenderung lebih sulit untuk diklasifikasikan karena jumlah data pada kelas minoritas yang sedikit jika dibandingkan dengan kelas lainnya pada *dataset*. Sehingga, proses klasifikasi akan cenderung bias terhadap kelas mayoritas dan akan berpengaruh pada performa dari model klasifikasi dalam mengambil keputusan (Wijayanti dkk., 2021).

Untuk mengatasi permasalahan data tidak seimbang terdapat dua buah pendekatan yang dapat dilakukan, yakni pendekatan pada level data dan pendekatan

pada level model algoritma. Pendekatan pada level data merupakan pendekatan dengan melakukan penyeimbangan distribusi data pada kelas mayoritas dan minoritas dengan menggunakan teknik pengambilan ulang sampel (*resampling*). Sedangkan, pendekatan pada level algoritma merupakan pendekatan dengan melakukan modifikasi terhadap algoritma seperti menggunakan beberapa parameter yang berpengaruh terhadap kinerja model algoritma atau menggabungkan beberapa metode (B. Santoso dkk., 2017). Pendekatan pada level data dengan menggunakan teknik *resampling* dalam menangani kasus data tidak seimbang memiliki keuntungan, yakni cenderung lebih fleksibel dan dapat membantu model dalam melakukan klasifikasi dengan lebih mudah (Maldonado dkk., 2019). Beberapa metode *resampling* dalam *machine learning* adalah *Random Over Sampling* (ROS), ADASYN, dan SMOTE.

Penerapan metode *resampling* dalam menangani kasus data tidak seimbang telah dilakukan dalam beberapa penelitian sebelumnya, yakni penelitian oleh Umer dkk. (2021) yang melakukan perbandingan terhadap teknik *resampling* SMOTE dan ADASYN. Hasil penelitian menunjukkan bahwa penerapan SMOTE dapat memberikan performa paling optimal pada model klasifikasi *Extra Tree Classifier* (ETC) dengan *accuracy* sebesar 98.26% dan ADASYN memberikan performa paling optimal pada model ETC dengan *accuracy* sebesar 96.26%. Hal ini menunjukkan bahwa metode SMOTE lebih unggul dibandingkan ADASYN. Chamzah dkk. (2022) pernah melakukan penerapan teknik *resampling* menggunakan metode SMOTE pada metode klasifikasi KNN dengan hasil bahwa metode SMOTE dapat membantu meningkatkan akurasi KNN sebesar 8% dari nilai akurasi awal sebesar 82% meningkat menjadi 90%. Terakhir, Dauda dkk. (2023) pernah melakukan perbandingan penerapan beberapa metode *resampling*, yakni ROS, SMOTE dan ADASYN pada berbagai model klasifikasi. Hasil penelitian menunjukkan metode *resampling* terbaik adalah metode SMOTE yang dikombinasikan dengan model klasifikasi XG-BOOST dengan hasil *accuracy* sebesar 98%.

Berdasarkan hasil penelitian tersebut, maka pada penelitian ini digunakan metode *resampling Synthetic Minority Oversampling Technique* (SMOTE) dalam menangani kasus data tidak seimbang. SMOTE merupakan salah satu metode

pengambilan ulang sampel dengan tujuan untuk menyamakan distribusi kelas yang bekerja dengan cara mengambil sampel kelas minoritas dan melakukan penyisipan sampel sintesis untuk menambahkan jumlah data pada kelas minoritas (Nurhopipah & Magnolia, 2022). Beberapa kelebihan yang dimiliki oleh metode SMOTE adalah tidak adanya informasi yang hilang pada sebuah kelas, dikarenakan tidak ada proses pengurangan data. Selain itu, SMOTE juga membuat kelas minoritas memiliki peluang lebih besar untuk dipelajari oleh metode klasifikasi dengan cakupan yang lebih luas yang menyebabkan SMOTE mampu meningkatkan nilai akurasi klasifikasi pada kelas minoritas (Wijayanti dkk., 2021).

Berdasarkan uraian latar belakang yang telah dipaparkan sebelumnya, maka dirancang sebuah penelitian yang berjudul Optimasi Analisis Sentimen Program Merdeka Belajar-Kampus Merdeka (MBKM) Pada Media Sosial Twitter Menggunakan Kombinasi *K-Nearest Neighbor* dan SMOTE yang diharapkan dapat memberikan performa yang lebih optimal dalam melakukan analisis sentimen terhadap program MBKM pada media sosial Twitter.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana performa penerapan metode KNN dan SMOTE dalam mengklasifikasikan sentimen masyarakat terhadap implementasi program MBKM?
2. Bagaimana hasil analisis sentimen data opini masyarakat terhadap implementasi program MBKM pada media sosial Twitter?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka yang menjadi tujuan dalam penelitian ini adalah:

1. Untuk mengetahui performa penerapan metode KNN dan SMOTE dalam mengklasifikasikan opini masyarakat terhadap implementasi program MBKM.
2. Untuk mengetahui hasil analisis sentimen masyarakat terhadap implementasi program MBKM pada media sosial Twitter.

1.4. Batasan Masalah Penelitian

Agar penelitian menjadi lebih terarah, maka berikut merupakan batasan masalah dalam penelitian ini:

1. Kumpulan data (*dataset*) yang digunakan pada penelitian ini bersumber dari kumpulan *tweet* pada media sosial Twitter yang berhubungan dengan topik implementasi program MBKM dengan menggunakan beberapa kata kunci, yakni “MBKM”, “Merdeka Belajar”, “Kampus Merdeka”, “Program MBKM”, ‘Program Kampus Merdeka’, ‘Implementasi MBKM’, ‘Implementasi Kampus Merdeka’ serta *hashtag* #mbkm dan #kampusmerdeka.
2. Data *tweet* yang digunakan adalah *tweet* berbahasa Indonesia yang diunggah dalam rentang waktu 9 Februari 2021 sampai dengan 17 Maret 2022.

1.5. Manfaat Penelitian

Berdasarkan permasalahan dan tujuan penelitian, maka manfaat yang dapat diperoleh melalui penelitian ini adalah:

1. Bagi Pembaca
 - a. Dapat memberikan informasi terkait opini masyarakat terhadap implementasi MBKM pada media sosial Twitter.
 - b. Dapat menjadi bahan evaluasi bagi pihak penyelenggara MBKM untuk pengembangan implementasi program MBKM kedepannya.
 - c. Dapat memberikan gambaran terkait implementasi metode KNN dan SMOTE dalam melakukan analisis sentimen.
 - d. Dapat menjadi bahan referensi apabila hendak dilakukannya penelitian terkait pada masa mendatang.
2. Bagi Peneliti
 - a. Dapat mengimplementasikan pengetahuan yang telah didapatkan selama menjalankan studi di perguruan tinggi.
 - b. Dapat menambah wawasan serta ilmu pengetahuan dalam memecahkan permasalahan dalam bidang *text mining* yakni analisis sentimen dengan menggunakan metode KNN dan SMOTE.