

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Media sosial merupakan suatu platform digital untuk melakukan aktivitas sosial dalam jaringan internet. Saat ini banyak sekali media sosial yang digunakan oleh masyarakat Indonesia salah satunya adalah Twiter. Berdasarkan data dari statista [1], per Januari 2022 pengguna Twitter di Indonesia mencapai angka 18,45 juta. Dengan menggunakan media sosial Twitter seseorang dapat mengekspresikan opini mereka melalui sebuah cuitan. Cuitan opini yang dibagikan pada media sosial twitter dapat mengandung emosi atau perasaan dari sang penulis. Emosi yang di deteksi dari setiap cuitan tersebut dapat digunakan untuk pengambilan keputusan pada bidang bisnis, hukum, kesehatan atau dalam kehidupan bersosial. Seperti pada penelitian Hilda Rachmi, Suparni, dan Ahmad Al Kaafi [2] dilakukan analisis emosi mengenai opini masyarakat terhadap diterapkannya sistem ganjil genap di kota Bogor melalui postingan masyarakat di sosial media.

Penelitian mengenai analisis emosi saat ini telah banyak dilakukan. Salah satunya adalah penelitian yang dilakukan oleh Fildza Sakinah Alnaz dan Warih Maharani [3]. Pada penelitian tersebut dilakukan klasifikasi emosi pada cuitan pengguna di media sosial Twitter dengan membandingkan performansi antara kombinasi metode *Naïve Bayes* dengan fitur N-Gram dan *Naïve Bayes* dengan fitur TF-IDF. Dimana akurasi terbaik yang didapatkan sebesar 0.6485 dengan rata-rata nilai *precision* 0.69, *recall* 0.64, dan *f1-score* 0.66. Terdapat masalah pada penelitian tersebut yaitu adanya ketidakseimbangan data antar kelas. Masalah ketidakseimbangan data cukup sering terjadi dalam masalah klasifikasi karena pada dasarnya data yang ditambang langsung dari sumber data merupakan data yang tidak seimbang sehingga membuat proses klasifikasi menjadi lebih menantang [4]. Untuk mengatasi ketidakseimbangan data dapat dilakukan proses *oversampling*. Penelitian yang dilakukan oleh Afifah Nurlaila, Wiranto, dan Ristu Saptono [5] mengenai klasifikasi emosi dari pelanggan produk perawatan kulit dari cuitan

Twitter menggunakan metode *Naïve Bayes* dilakukan proses *oversampling* menggunakan *Random Over Sampling* (ROS) karena saat pengujian terdapat satu kelas yang memiliki jumlah data yang lebih banyak dari kelas lainnya sehingga menghasilkan nilai *precision* dan *recall* yang sangat rendah mencapai 0%. Proses *oversampling* yang dilakukan pada penelitian tersebut dapat meningkatkan nilai *recall*, *precision*, dan *f1-measure*. Namun, dikarenakan cara kerja metode ROS yang menduplikasi data yang sama persis dari kelas minoritas sehingga mungkin menyebabkan *overfitting* [6].

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu perkembangan dari metode *oversampling*. Teknik ini mirip dengan ROS, perbedaannya ada pada sampel yang dihasilkan, tidak diduplikat secara random dari sampel yang sudah ada, tetapi sampel tersebut dibuat dengan mempergunakan konsep *Nearest Neighbour* [6]. Saat ini telah banyak penelitian yang merapkan metode SMOTE. Pada penelitian Cristian Padurariu dan Mihaela Elena Breaban [7] dalam melakukan perbandingan performansi pada kombinasi antara berbagai metode pembelajaran mesin, metode sampling, dan metode pembobotan kata untuk menghadapi ketidakseimbangan data didapatkan bahwa metode *oversampling* SMOTE mampu memberikan performansi yang sama bahkan melebihi metode ROS pada beberapa kasus pengujian. Penelitian yang dilakukan oleh Saifurrachman Chohan, Arifin Nugroho, Achmad Maezar Bayu Aji, dan Windu Gata [8] dalam melakukan analisis sentimen pada aplikasi Duolingo menggunakan metode *Naïve Bayes* dan SMOTE mampu memberikan performansi yang sangat baik dimana akurasi yang dihasilkan sebesar 91.95% dan 0.740 untuk AUC.

Berdasarkan uraian latar belakang diatas maka pada penelitian ini akan dilakukan penerapan metode SMOTE untuk mengatasi ketidakseimbangan data pada analisis emosi cuitan dari media sosial Twitter dengan menggunakan *Naïve Bayes* sebagai metode klasifikasi yang akan digunakan. Sehingga judul dari penelitian ini adalah “Analisis Emosi Pada Media Sosial Twitter Menggunakan Metode *Naïve Bayes* Dan *Synthetic Minority Oversampling Technique*”.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas maka didapatkan rumusan masalah pada penelitian ini yaitu bagaimana performansi dari metode *Naïve Bayes*-SMOTE dalam melakukan analisis emosi pada cuitan media sosial Twitter.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk melakukan penerapan metode *Naïve Bayes*-SMOTE dalam melakukan analisis emosi pada cuitan media sosial Twitter. Tujuan dari penelitian ini adalah untuk mengetahui nilai performansi dari metode *Naïve Bayes*-SMOTE dalam melakukan analisis emosi pada cuitan media sosial Twitter.

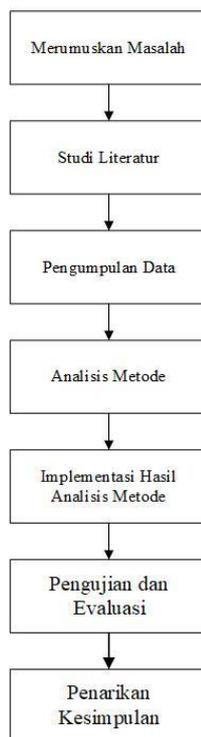
1.4 Batasan Masalah

Agar masalah yang sedang ditinjau lebih terarah dan mencapai tujuan yang telah ditentukan, maka akan dibatasi masalah sebagai berikut yang terbagi ke dalam beberapa aspek diantaranya.

1. Dataset yang digunakan merupakan cuitan dari media sosial Twitter.
2. Dataset cuitan Twitter yang digunakan diambil dari (<https://github.com/meisaputri21/Indonesian-Twitter-Emotion-Dataset>).
3. Dataset cuitan Twitter merupakan file dengan format .csv
4. Dataset merupakan cuitan berbahasa Indonesia.
5. Kelas emosi yang dianalisis terbagi menjadi lima yaitu *anger*, *fear*, *happy*, *sadness*, dan *love*.
6. Metode ekstraksi fitur yang digunakan adalah N-Gram.
7. Sistem hanya mampu mengenali satu kelas emosi di setiap cuitan.

1.5 Metodologi Penelitian

Metodologi penelitian merupakan langkah-langkah yang dilakukan oleh peneliti dalam suatu penelitian dan merupakan suatu analisis kritis dari sebuah metode penelitian [9]. Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian kuantitatif. Adapun tahapan dari metode penelitian yang digunakan pada tugas akhir ini dapat dilihat pada gambar 1.1 di bawah ini.



Gambar 1.1 Tahapan Metode Penelitian

1.5.1 Merumuskan Masalah

Pada tahapan ini dilakukan perumusan masalah yang menjadi fokus penelitian dari latar belakang yang telah dijabarkan sebelumnya yaitu bagaimana kinerja dari metode *Naïve Bayes-SMOTE* dalam melakukan analisis emosi pada cuitan media sosial Twitter.

1.5.2 Studi Literatur

Studi literatur dilakukan dengan cara mengumpulkan dan mempelajari data pustaka yang menunjang penelitian yang sedang dilakukan. Data pustaka dapat berupa jurnal, buku, artikel atau sumber-sumber lain yang berkaitan dengan tema analisis emosi pada text, metode *Naïve Bayes*, metode *SMOTE*, atau metode *preprocessing* yang digunakan.

1.5.3 Pengumpulan Data

Dataset pada penelitian ini diambil dari sumber terbuka GitHub yang berasal dari penelitian Mei Silviana Saputri, Rahmad Mahendra, dan Mirna Adriani [10]. Dataset yang digunakan merupakan cuitan berbahasa Indonesia dari media sosial Twitter yang disatukan dalam file dengan format *.csv*.

1.5.4 Analisis Metode

Analisis metode dilakukan terhadap metode-metode yang digunakan dalam penelitian ini. Adapun metode-metode yang digunakan pada sistem ini terbagi menjadi dua bagian yaitu:

1. *Preprocessing*

Tahapan *preprocessing* dilakukan untuk mengolah data masukan agar mempermudah proses ekstraksi fitur. Adapun metode *pre-processing* yang dilakukan pada penelitian ini adalah *case folding*, *convert slangword*, *data cleaning*, *convert negation*, *tokenization*, *stopword removal*, dan *stemming*.

2. Ekstraksi Fitur

Pada tahapan ini akan dilakukan ekstraksi fitur pada data cuitan dari media sosial Twitter yang telah melalui tahap *preprocessing*.

3. *Sampling*

Setelah dilakukan *preprocessing* dan pembobotan maka akan dilakukan oversampling terhadap kelas minoritas padadata latih menggunakan metode SMOTE.

4. Klasifikasi

Pada tahapan klasifikasi akan dilakukan klasifikasi emosi terhadap data latih cuitan dari media sosial Twitter. Klasifikasi pada penelitian ini dilakukan dengan menggunakan metode *Naïve Bayes*.

1.5.5 Implementasi Hasil Analisis Metode

Tahapan ini merupakan tahapan implementasi dari hasil analisis metode yang telah dilakukan pada tahapan sebelumnya dalam bentuk kodingan program. Dengan memaparkan *tools* atau *library* apa saja yang digunakan dalam melakukan proses implementasi.

1.5.6 Pengujian dan Evaluasi

Tahapan ini merupakan tahapan pengujian terhadap performansi metode *Naïve Bayes-SMOTE* dalam melakukan analisis emosi pada cuitan media sosial Twitter. Pengujian dilakukan dengan menggunakan dataset yang telah disiapkan. Evaluasi terhadap hasil pengujian juga dilakukan pada tahapan ini.

1.5.7 Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan dari hasil penelitian penerapan metode *Naïve Bayes-SMOTE* dalam melakukan analisis emosi pada cuitan media sosial Twitter dan dilakukan pemberian saran untuk penelitian selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan sebuah gambaran umum terhadap penelitian yang sedang dilakukan. Adapun sistematika yang digunakan pada tugas akhir ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Pada bab ini berisi uraian latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, tahap pengumpulan data, model pengembangan perangkat lunak dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab ini akan membahas berbagai konsep-konsep dasar dan teori-teori pendukung yang berhubungan dengan penelitian dan hal-hal yang berguna dalam proses analisis permasalahan, dan menggunakannya sebagai acuan untuk pemecahan masalah pada penelitian ini.

BAB 3 ANALISIS MASALAH DAN METODE

Pada bab ini akan membahas tentang deskripsi analisis masalah, analisis data, gambaran umum sistem, dan analisis metode algoritma.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini berisi hasil implementasi analisis dari BAB 3, serta hasil pengujian sistem baik untuk mengetahui pengujian kinerja sistem yang telah dibuat.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan yang diperoleh dari hasil pengujian sistem, serta saran untuk pengembangan pada penelitian selanjutnya