

BAB II

LANDASAN TEORI

2.1 Penelitian Terkait

Penelitian terkait dilakukan untuk meninjau berbagai studi sebelumnya yang membahas analisis sentimen menggunakan algoritma Support Vector Machine (SVM). Kajian ini bertujuan untuk memahami pendekatan yang telah digunakan, tahapan pengolahan data, serta performa model yang diperoleh dalam penelitian terdahulu. Styawati dan Handrastuty (2021) menerapkan algoritma SVM guna mengklasifikasikan pandangan publik terhadap program Kartu Prakerja berdasarkan data Twitter.

Penelitian tersebut melalui proses prapemrosesan, pembobotan fitur berbasis TF-IDF, serta pengujian menggunakan beberapa kernel. Analisis yang dilakukan mengidentifikasi bahwa kernel linear menghasilkan akurasi sebesar 98,67%, sedangkan kernel RBF memperoleh akurasi sebesar 98,34%. Hasil tersebut menunjukkan bahwa SVM memiliki performa klasifikasi yang sangat tinggi pada data sentimen media sosial.

Aulia et al. (2021) melakukan perbandingan beberapa jenis kernel pada implementasi SVM dalam pengelompokan opini publik terkait vaksinasi COVID-19. Penelitian tersebut menggunakan pendekatan text mining serta tahapan preprocessing sebelum proses klasifikasi. penelitian yang dilakukan menunjukkan bahwa kernel linear memperoleh akurasi tertinggi sebesar 83,06%, diikuti oleh kernel polynomial dan sigmoid dengan nilai yang sedikit lebih rendah.

Temuan ini menunjukkan bahwa pemilihan kernel berpengaruh terhadap kinerja model.

Darwis dan Pratiwi (2022) menerapkan algoritma SVM untuk menganalisis sentimen masyarakat terhadap isu Komisi Pemberantasan Korupsi (KPK) di Twitter. Penelitian tersebut melakukan tahapan preprocessing dan evaluasi model untuk mengukur tingkat keberhasilan klasifikasi. Temuan studi tersebut memperlihatkan bahwa penerapan svm menghasilkan capaian akurasi yang memadai dalam proses klasifikasi sentimen.

Idris et al. (2023) dalam penelitian terhadap ulasan aplikasi Shopee menggunakan SVM dengan tahapan preprocessing dan pembobotan fitur. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi sebesar 98% dengan nilai F1-score yang tinggi, sehingga menunjukkan performa klasifikasi yang sangat baik.

Aisah et al. (2023) juga menerapkan SVM dalam analisis sentimen ulasan aplikasi Al-Qur'an digital dan memperoleh akurasi sebesar 85,11%. Nilai precision dan recall yang di hasilkan mengidentifikasi bahwa model menunjukan tingkat kestabilan yang memadai dalam proses klasifikasi data. Sejumlah studi terdahulu menunjukkan bahwa algoritma SVM memperlihatkan tingkat kinerja yang relatif stabil dalam tugas klasifikasi teks berbasis sentimen.

Namun, sebagian besar penelitian sebelumnya hanya menggunakan satu konfigurasi klasifikasi tanpa membandingkan beberapa skenario jumlah kelas dalam satu penelitian. Oleh karena itu, penelitian ini memiliki perbedaan pada objek

kajian serta penerapan tiga skenario klasifikasi untuk mengetahui konfigurasi yang memberikan performa terbaik.

2.2 Text Mining

Text mining merupakan proses pengolahan data teks yang bersifat tidak terstruktur menjadi informasi yang dapat dianalisis secara sistematis. Data teks seperti komentar media sosial, ulasan pengguna, atau opini publik umumnya tidak memiliki format yang terstruktur sehingga memerlukan tahapan pengolahan khusus sebelum dapat digunakan dalam proses analisis.

Dalam penelitian analisis sentimen berbasis media sosial, text mining umumnya melibatkan beberapa tahapan utama, yaitu preprocessing, ekstraksi fitur, dan klasifikasi (Wati & Ernawati, 2021). Tahapan tersebut bertujuan untuk meningkatkan kualitas data serta mentransformasikan teks menjadi vektor numerik yang selanjutnya di gunakan dalam tahap klasifikasi dengan SVM.

Sebagaimana diterapkan dalam penelitian Ananda dan Pristyanto (2021) serta Husada dan Suryaputra (2023), text mining menjadi langkah penting dalam mempersiapkan data agar siap digunakan dalam proses klasifikasi menggunakan algoritma SVM.

2.3 Analisis Sentimen

Analisis sentimen merupakan teknik dalam text mining yang digunakan untuk mengkaji dan mengklasifikasikan opini ke dalam kategori tertentu, seperti positif, negatif, dan netral. Melalui pendekatan ini, kecenderungan pandangan

masyarakat terhadap suatu isu dapat diidentifikasi secara lebih sistematis dan terukur (Styawati & Handrastuty, 2021; Idris et al., 2023).

Dalam konteks media sosial, analisis sentimen di pandang sebagai metode relevan karena pengguna secara aktif menyampaikan pendapat, kritik, dan dukungan terhadap berbagai kebijakan secara terbuka. Data tersebut dapat dimanfaatkan untuk menggambarkan opini publik secara sistematis berdasarkan kategori sentimen yang telah ditentukan.

2.4 Text Preprocessing

Text preprocessing merupakan tahap awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan menyiapkan data sebelum dilakukan proses klasifikasi. Data dari media sosial umumnya mengandung noise seperti URL, simbol, tanda baca, singkatan, serta variasi penulisan kata yang dapat memengaruhi hasil analisis.

Tahapan preprocessing penelitian ini mencakup proses cleaning, case folding, tokenisasi, normalisasi, stopwords removal, dan stemming. Tahapan ini sejalan dengan penelitian Tinilah dan Elizabeth (2022) serta Aisah et al. (2023) yang juga menerapkan proses pembersihan teks sebelum dilakukan ekstraksi fitur. Proses tersebut berperan dalam membersihkan data dari elemen yang mengganggu guna mendukung performa model klasifikasi.

2.5 Term Frequency–Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode pembobotan kata yang digunakan guna mengonversi teks menjadi representasi numerik. Metode ini memberikan bobot

pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat keunikannya dalam keseluruhan kumpulan dokumen.

Penggunaan TF-IDF sebelum proses klasifikasi menggunakan SVM diterapkan dalam penelitian Styawati dan Handrastuty (2021), Idris et al. (2023), serta Aulia et al. (2021). Pembobotan fitur memungkinkan model menilai tingkat kepentingan setiap kata sehingga proses penentuan kategori sentimen menjadi lebih akurat..

2.6 Support Vector Machine (SVM)

Support Vector Machine merupakan algoritma pembelajaran mesin yang digunakan untuk melakukan klasifikasi dengan menentukan hyperplane optimal yang memisahkan data antar kelas. Algoritma ini bekerja dengan mencari margin terbesar antara data dari kelas yang berbeda sehingga menghasilkan kemampuan model dalam menangani data baru secara efektif.

Berbagai penelitian terdahulu menggambarkan kemampuan SVM dalam mencapai tingkat akurasi yang tinggi pada analisis sentimen media sosial (Styawati & Handrastuty, 2021; Idris et al, 2023). Dengan mempertimbangkan keunggulan yang dimiliki, algoritma ini digunakan sebagai metode mengklasifikasikan sentimen masyarakat terhadap Program Makanan Bergizi Gratis.

2.7 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk mengukur tingkat keberhasilan algoritma dalam mengklasifikasikan data sentimen. Metrik evaluasi yang umum digunakan meliputi accuracy, precision, recall, dan F1-score (Idris et al., 2023;

Aisah et al., 2023). Penggunaan metrik tersebut memungkinkan penilaian performa model secara menyeluruh serta membandingkan hasil pada setiap skenario klasifikasi yang diterapkan dalam penelitian ini. Melalui evaluasi tersebut, dapat diketahui sejauh mana algoritma SVM mampu mengelompokkan data sentimen secara efektif dan konsisten.

2.8 Platform X sebagai Sumber Data

Platform X merupakan media sosial berbasis teks yang banyak digunakan untuk menyampaikan opini terhadap berbagai isu publik. Penelitian ini memanfaatkan Twitter sebagai platform media sosial dalam analisis sentimen karena menyediakan data teks dalam jumlah besar yang dapat diolah menggunakan pendekatan text mining (Darwis & Pratiwi, 2022; Husada & Suryaputra, 2023).