

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi pada saat ini berkembang begitu pesat, bahkan teknologi telah masuk dalam dunia pendidikan. Perkembangan teknologi saat ini memunculkan berbagai jenis aplikasi yang berguna untuk membantu siswa dalam belajar, salah satu aplikasi tersebut adalah Brainly. Fitri Al Qolbi (2021) menyatakan Aplikasi Brainly adalah platform media sosial berbasis pendidikan yang diciptakan oleh perusahaan pendidikan asal Polandia, yang memudahkan pengguna dalam menyelesaikan pekerjaan sekolah dengan memungkinkan pertukaran pertanyaan dan jawaban antara pengguna lainnya melalui internet. Dikutip dari laman Brainly.co.id yang diakses pada 8 Mei 2022, Brainly adalah tempat berbagi ilmu 350 juta siswa dan pakar edukasi, belajar bersama untuk menyelesaikan soal-soal yang paling rumit sekalipun. Aplikasi ini bisa diakses melalui website atau didownload gratis di Playstore maupun Appstore. Brainly menyediakan 26 kategori pelajaran mulai dari tingkatan SD, SMP, dan SMA/SMK. Siswa, orang tua, dan guru bisa menggunakan aplikasi ini untuk bertanya dan menjawab pertanyaan. Aplikasi Brainly ini sangat berguna bagi siswa karena siswa langsung mendapatkan jawaban dari suatu pertanyaan tanpa harus mencari dari berbagai sumber. Bahkan, aplikasi Brainly pada 8 Mei 2022 sudah didownload lebih dari 100 juta pengguna di Playstore Indonesia.

Media sosial adalah salah satu tempat untuk mengekspresikan kepuasan dan kekecewaan pengguna atau pendapat pengguna mengenai suatu aplikasi. Hal tersebut dapat digunakan sebagai bahan analisis sentimen terhadap suatu aplikasi khususnya Brainly. Analisis sentimen adalah suatu teknik yang digunakan untuk mengambil informasi opini, memahami, dan memproses data teks secara otomatis dengan tujuan mengidentifikasi sentimen yang terkandung dalam suatu opini. (Sari dan Wibowo, 2019).

Menurut Giovani dkk. (2020) Twitter adalah salah satu platform media sosial yang sangat diminati oleh pengguna internet. Popularitasnya disebabkan oleh antarmuka pengguna yang sederhana dan kemudahan penggunaannya. Selain

itu, pengguna Twitter memiliki kebebasan untuk menyampaikan pendapat dan opini mereka secara bebas. Pengguna Twitter akan memberikan berita atau komentar mengenai topik yang saat ini menjadi topik teratas di dunia. Hal-hal yang saat ini menjadi topik utama dan banyak dikomentari oleh pengguna Twitter akan menimbulkan suatu masalah atau trending topik di jejaring sosial.

Beberapa peneliti telah menggunakan bahan penelitian dari Twitter untuk menganalisis sentimen menggunakan berbagai algoritma. Sebagai contoh penelitian yang menganalisis sentimen tanggapan masyarakat Indonesia terhadap pandemic *Covid-19* pada media sosial Twitter menggunakan algoritma SVM, *naïve bayes*, dan KNN. Penelitian tersebut menghitung jumlah kata positif dan negatif pada setiap tanggapan, jika banyak kata positif maka tanggapan tersebut dan sebaliknya. Penelitian tersebut memberikan hasil *support vector machine* merupakan algoritma yang paling cocok untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap *Covid-19* pada media sosial Twitter dibandingkan dengan *naïve bayes* dan K-nearest neighbor (Pamungkas dan Kharisudin, 2021). Penelitian lain yang membandingkan akurasi dan waktu proses algoritma KNN dan SVM dalam analisis sentimen Twitter. Penelitian tersebut menggunakan metode ekstraksi fitur TF-IDF dan memberikan hasil Metode klasifikasi Support Vector Machine memiliki akurasi yang lebih tinggi dibandingkan metode klasifikasi K-Nearest Neighbor, yaitu sebesar 89,70% tanpa validasi K-Fold Cross Validation dan sebesar 88,76% dengan validasi K-Fold Cross Validation (Nasution dan Hayaty, 2019). Penelitian serupa membandingkan metode *lexicon-based* dan SVM untuk analisis sentimen berbasis ontologi pada kampanye pilpres Indonesia tahun 2019 di Twitter. Penelitian tersebut menghasilkan penggunaan metode SVM memiliki performa yang lebih baik dibandingkan metode *Lexicon-based* (Najib dkk., 2019). Penelitian lainnya adalah membandingkan metode klasifikasi *random forest* dan SVM pada analisis sentimen PSBB yang menghasilkan *Support Vector Machine* dianggap lebih baik karena mampu mengenali *tweet* dengan label “Positif” (Adrian dkk., 2021). Berdasarkan berbagai referensi penelitian yang telah dilakukan, algoritma *Support Vector Machine* (SVM) menjadi metode yang digunakan dalam penelitian ini.

Pada penelitian yang dilakukan oleh Rofiqoh dkk. (2017) terdapat beberapa tahapan yang dilakukan dalam analisis sentimen yaitu memasukkan *dataset*, *Preprocessing*, pembobotan kata (TF-IDF), pembobotan kata *Lexicon Based Features*, klasifikasi dengan SVM. Dalam penelitian tersebut ditemukan sebuah masalah yaitu tingkat akurasi dari penggunaan pembobotan kata dengan *Lexicon Based Features* lebih rendah dibandingkan dengan analisis sentimen yang tidak menggunakan *Lexicon Based Features*. Berdasarkan penelitian tersebut penulis berinisiatif untuk membandingkan metode *Lexicon Based Features* dengan metode ekstraksi fitur lain agar peneliti atau analis dapat menggunakan ekstraksi fitur yang lebih baik untuk menganalisis sentimen masyarakat di internet.

Penelitian yang dilakukan oleh Rangkuti dkk. (2018) menggunakan ekstraksi fitur *ensemble features* dan *bag of words*. Hasilnya yang tidak menggunakan kata baku *Accuracy* 82%, *Precision* 86%, *Recall* 79,62%, *F-Measure* 82,69%. Sedangkan yang menggunakan kata baku secara manual seleksi Fitur optimal dengan *threshold* 85% dan hasil evaluasi yang didapatkan *Accuracy* 90%, *Precision* 92%, *Recall* 88,46% dan *F-Measure* 90,19%. Penelitian yang dilakukan oleh Permatasari (2018) menggunakan metode ekstraksi fitur *ensemble features* dan *ensemble features* digabung dengan *bag of words*. Rausanfitita dkk. (2018) melakukan penelitian dengan ekstraksi fitur TF-IDF digabung dengan *ensemble features* dan metode klasifikasinya adalah menggunakan *extreme learning machine*. Hasilnya penggunaan *ensemble features* terhadap hasil analisis sentimen opini film pada Twitter menghasilkan akurasi sebesar 61,33%, *precision* sebesar 0,6308, *recall* sebesar 0,5467, dan *f-measure* sebesar 0,5857. Dalam penelitian yang dilakukan oleh Rausanfitita dkk. (2018) membandingkan metode ekstraksi fitur *ensemble features*, *bag of words*, *lexicon based features* digabung dengan *bag of words* untuk klasifikasi menggunakan metode *extreme learning machine*. Hasilnya, rata-rata dari *bag of words*, dan *lexicon based features* digabung dengan *bag of words* lebih besar dari *ensemble features* namun *ensemble features* memiliki akurasi yang stabil. Berdasarkan penelitian yang berhasil menggunakan *ensemble features*, penulis menggunakan *ensemble features* sebagai pembanding yang dibandingkan dengan *lexicon based features*.

Berdasarkan uraian diatas, penulis berinisiatif membandingkan metode *Lexicon Based Features*, dan *Ensemble Feature* untuk ekstraksi fitur analisis sentimen Aplikasi Brainly berdasarkan akurasi dan waktu. Dengan adanya penelitian ini diharapkan dapat digunakan sebagai acuan dalam memilih metode ekstraksi fitur untuk analisis sentimen menggunakan algoritma *Support Vector Machine*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, terdapat rumusan masalah yang menjadi fokus dalam penelitian ini, yaitu:

- a. Bagaimana membuat perangkat lunak perbandingan metode ekstraksi fitur *Lexicon Based Feature*, dan *Ensemble Feature*?
- b. Berapakah akurasi dan waktu metode ekstraksi fitur *Lexicon Based Feature*, dan *Ensemble Feature*?
- c. Bagaimana mengimplementasikan metode *Support Vector Machine* (SVM) untuk analisis sentimen Brainly di media sosial Twitter?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

- a. Membuat perangkat lunak perbandingan metode ekstraksi fitur *Lexicon Based Feature*, dan *Ensemble Feature*.
- b. Mengukur akurasi dan waktu metode ekstraksi fitur *Lexicon Based Feature*, dan *Ensemble Feature*.
- c. Menerapkan algoritma *Support Vector Machine* untuk menganalisis sentimen aplikasi Branly di Twitter.

1.4 Manfaat

Manfaat yang diharapkan dari hasil penelitian ini adalah:

- a. Peneliti menjadi lebih memahami implementasi dari algoritma *Support Vector Machine* (SVM).

- b. Mengetahui akurasi hasil klasifikasi *Support Vector Machine* (SVM) dan waktu dari metode ekstraksi fitur *Lexicon Based Feature*, dan *Ensemble Feature*.
- c. Mengetahui metode ekstraksi fitur yang baik digunakan dalam analisis sentimen dengan *Support Vector Machine* (SVM).
- d. Menjadi acuan dalam memilih metode ekstraksi fitur untuk analisis sentimen menggunakan algoritma *Support Vector Machine*.

1.5 Batasan Masalah

- a. Data yang digunakan adalah sentimen masyarakat mengenai Brainly di Twitter yang Berbahasa Indonesia
- b. Metode yang digunakan untuk mengklasifikasi adalah metode *Support Vector Machine*.
- c. Tahapan *preprocessing* yang digunakan dalam penelitian ini adalah *case folding*, *data cleaning*, *tokenisasi*, *stopword removal*, dan *stemming*.