

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi saat ini memberi dampak kemajuan telah memberikan perubahan pada beberapa segmen kehidupan, seolah-olah dalam bidang pengajaran, dll. zaman sekarang perkembangan web dan media sosial telah berkembang sangat pesat hingga memberikan data kesimpulan yang bermacam-macam dari individu lain. Gambaran yang dilakukan oleh Afiliasi Pelaku Usaha Manfaat Web (APJII) pada tahun 2014 ternyata jumlah klien web di Indonesia telah berkembang dalam waktu yang lama. Pada tahun 2005, ada 16 juta klien web di Indonesia dan jumlah ini terus bertambah 20 juta orang adalah klien web pada tahun 2006. Pertanyaan tentang ini juga menghasilkan pemanfaatan saat ini sebagian besar web masuk ke sosial media (87,4% dari seluruh responden), media sosial saat ini, khususnya Twitter salah satunya. Platform media sosial yang sangat populer yaitu *Twitter* di antara berbagai demografi. Hal ini terlihat dari semakin banyaknya pengguna *Twitter*, termasuk di Indonesia. *Twitter* memiliki basis klien dinamis pada tahun 2016 sebesar ± 312 juta per bulan. Topik yang paling banyak dibicarakan dan menjadi tren di *Twitter* saat ini menyebabkan isu atau subjek tertentu menjadi tren di platform [1].

Kerapan sape, dikenal juga sebagai pacuan sapi, merupakan sebuah tradisi dalam budaya Madura yang menggambarkan balapan sapi. Dalam perlombaan ini, dua ekor sapi saling berlomba dengan dihubungkan oleh alat bernama pangonong dan kaleles. Kompetisi kerapan sapi ini biasanya diadakan di berbagai tingkatan, dimulai dari tingkat kecamatan, lalu meningkat ke tingkat kabupaten, dan puncaknya diadakan di tingkat Madura. Namun, tradisi ini juga menuai berbagai opini dari masyarakat, baik yang positif maupun negatif. Analisis sentimen terhadap opini publik ini dapat memberikan wawasan berharga bagi pemangku kepentingan dalam memahami persepsi masyarakat terhadap sapi kerapan [2].

Dampak positif dari sapi kerapan antara lain : Sapi kerapan Madura adalah tradisi unik yang mencerminkan kekayaan budaya dan sejarah pulau Madura,

Perlombaan ini menjadi daya tarik wisata yang menarik banyak pengunjung dari berbagai daerah dan mancanegara, Karapan sapi memperkuat rasa kebersamaan dan solidaritas antarwarga desa yang terlibat dalam persiapan dan pelaksanaan acara, Sapi-sapi yang ikut serta dalam karapan diperlakukan dengan sangat baik dan dijaga kesehatannya secara ketat, Karapan sapi memberikan kesempatan bagi peternak lokal untuk menunjukkan hasil kerja keras mereka dalam merawat sapi. Dan dampak negatif dari karapan sapi sebagai berikut : Beberapa orang mengkritik karapan sapi karena dianggap sebagai bentuk kekerasan terhadap hewan, terutama dalam cara sapi dipacu untuk berlari cepat, Ada laporan tentang penggunaan zat-zat yang tidak aman atau tidak etis untuk meningkatkan kinerja sapi dalam perlombaan, Karapan sapi dapat menyebabkan stres dan cedera pada sapi yang ikut serta dalam perlombaan, Tradisi ini membutuhkan biaya yang tidak sedikit yang bisa menjadi beban bagi peternak dengan sumber daya terbatas, Perjudian yang sering kali terjadi di sekitar karapan sapi dapat menimbulkan dampak negatif, seperti meningkatnya potensi konflik antarpemonton, masalah sosial, dan ekonomi bagi keluarga yang terlibat dalam perjudian.

Analisis sentimen adalah bidang *Natural Language Processing (NLP)* untuk membantu mengidentifikasi sentimen (negatif dan positif) dari kumpulan data-data terdiri dari opini atau pandangan yang diungkapkan dalam format teks tentang suatu isu atau peristiwa. Ini adalah area penelitian populer yang dianggap bermanfaat di berbagai aspek. Analisis sentimen umumnya dipakai di banyak bidang seperti prediksi isu politik, harga saham, layanan atau kepuasan dan lain-lainnya. Salahsatunya cara untuk melakukan analisis sentimen ialah dengan menganalisis tweet di *Twitter* yang berisi opini, yang digunakan sebagai sumber untuk menganalisa opini publik yaitu terhadap individu atau institusi. Hal ini karena *tweet* mengandung sentimen yang dapat digunakan sebagai pengukur opini publik dan dapat digunakan untuk evaluasi di masa mendatang. Untuk menentukan sentimen sebuah *tweet*, diklasifikasikan menjadi dua class: class negatif dan class positif. penelitian ini dilakukan analisis sentimen Karapan sapi dengan metode *Support Vector Machine (SVM)* untuk mengklasifikasikan komentar publik menjadi negatif dan positif. [3].

Sebuah penelitian yang dilakukan untuk menganalisis sentimen warga Indonesia terhadap pandemi *Covid-19* di *Twitter*. Penelitian menggunakan tiga metode yaitu *Naive Bayes*, *SVM*, dan *KNN* untuk melakukan analisis, dan ketiga algoritma tersebut dibandingkan untuk menentukan algoritma terbaik dalam mengklasifikasikan data respon. Hasil evaluasi dengan teknik *10-Fold Cross Validation* menunjukkan bahwa metode *SVM* memiliki tingkat *accuracy* hasil rata-rata yang cukup tinggi dibandingkan dengan metode *KNN* dan *Naive Bayes*. Akurasi rata-rata algoritma *SVM* dengan kernel linier mencapai 90,01%, sedangkan *Naive Bayes* dengan *Laplace smoothing* 1 mencapai 79,20% dan *KNN* dengan nilai *K* 20 dan menggunakan kernel optimal mencapai 62,10% [4].

Sebelumnya telah dilakukan sebuah penelitian yang bertujuan klasifikasikan teks ke dalam kategori negatif serta positif dari tweet Jawa Barat calon gubernur pada periode 2018-2023, menggunakan algoritma *Naive Bayes* dan *SVM* dengan seleksi fitur menggunakan Genetic Algorithm. Pada dataset yang tidak seimbang, hasil evaluasi menunjukkan bahwa *SVM* hasil rata-rata *accuracy* 92,61% dengan AUC 0,950, sedangkan *Naive Bayes* menghasilkan *accuracy* rata-rata 93,29% dengan AUC 0,525. Hasil tersebut menunjukkan bahwa *SVM* dapat digunakan untuk mendeteksi kategori negatif dan positif pada *tweet* dengan tingkat akurasi yang tinggi, khususnya pada dataset berbahasa Indonesia. Penelitian ini menjadi inovatif karena menunjukkan potensi penggunaan *SVM* pada dataset *Twitter* berbahasa Indonesia. [5].

Support vector Machine (SVM) adalah dilatih menggunakan algoritma pembelajaran berdasarkan teori optimisasi dan machine learning menggunakan hipotesis fungsi linier dalam ruang fitur berdimensi tinggi. *SVM* pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai seperangkat konsep luar biasa yang harmonis di bidang pengenalan pola. Proses transformasi dalam *SVM* ketergantungan pada parameter serta fungsi kernel yang digunakan, yang berpengaruh pada *accuracy* model yang dihasilkan. *SVM* sangat efektif dalam melakukan klasifikasi teks dan dapat dilihat secara geometris *hyperplane* dalam fitur ruang, yang memisahkan contoh positif dari contoh negatif. *Hyperplane* yang dipilih selama

pelatihan adalah yang unik dan dapat memisahkan contoh positif dari contoh negatif. Keuntungan utama dari pendekatan klasifikasi SVM adalah mampu mengatasi *overfitting*, sehingga dapat menghasilkan kinerja yang baik [6]. Hasil tersebut masih kurang baik, yang dimana SVM memiliki sebuah permasalahan terbesar yakni menentukan kernel dan nilai dari parameter yang digunakan dan dapat berakibat dalam hasil yang buruk. Pengembangan metode SVM telah banyak dilakukan, salah satunya adalah optimasi metode svm dengan menggunakan *gridsearchCV* yang dilakukan oleh Ahmad Toha, Purwono Purwono dan Windu Gata. Penelitian tersebut menekankan optimasi pada tuning *hyperparameter* SVM. Percobaan dilakukan sebanyak 4 yaitu: kernel *polynomial* 2 derajat, kernel *linear*, kernel *polynomial* 3 derajat dan kernel *RBF*. Model terbaik adalah dengan menggunakan kernel *polynomial* 2 derajat dengan nilai akurasi sebesar 73,31%. Setelah itu, model terbaik ditambah dengan nilai C dan *Gamma* terbaik ditemukan nilai akurasi sebesar 94,8% pada nilai $C=0,1$ dan $\text{Gamma}=1$ [7]. Sehingga metode *Grid Search* dapat digunakan untuk optimasi parameter SVM.

Menghitung bobot *TF-IDF* (*Term Frequency-Inverse Document Frequency*) ialah sebuah metode digunakan untuk mengonversi teks data menjadi numerik data dengan tujuan memberikan bobot pada setiap fitur atau kata. Tujuan dari penggunaan *TF-IDF* adalah untuk mengevaluasi tingkat kepentingan sebuah kata pada sebuah dokumen dengan menghitung frekuensi muncul kata tersebut pada setiap *documents* (*TF*) dan yang mengandung kata tersebut frekuensi dokumen (*DF*), yang kemudian diubah menjadi nilai *IDF* (*inverse* dari *DF*). Setelah itu, pembobotan kata dihasilkan dari perkalian nilai *TF* dengan *IDF*, di mana kata dengan nilai bobot yang besar menunjukkan kepentingannya yang tinggi dalam dokumen, dan kata dengan nilai bobot yang kecil menunjukkan kepentingannya yang rendah karena kemunculannya yang sering dalam banyak dokumen. Namun, ekstraksi fitur dengan *TF-IDF* menghasilkan banyak *term*, sehingga perlu dilakukan pemilihan fitur yang penting dalam dokumen teks. Untuk memilih fitur yang penting digunakan beberapa algoritma *feature selection* yang antara lain *chi square*, *Information Gain*, *backward elimination* dan *forward selection*. Dalam suatu penelitian, algoritma SVM menunjukkan hasil yang memuaskan

dengan tingkat akurasi mencapai 81,10% dan *AUC (Area Under the Curve)* sebesar 0,904 setelah dibandingkan dengan beberapa algoritma *feature selection* yang berbeda. Algoritma *feature selection* yang baik adalah *Information Gain* yang menghasilkan tingkat akurasi rata-rata sebesar 84,57% dan *AUC* rata-rata 0,899. Integrasi antara algoritma klasifikasi terbaik dan algoritma *feature selection* terbaik meningkatkan tingkat *accuracy* menjadi 81,50% dan *AUC* 0,929 dibandingkan dengan eksperimen yang hanya menggunakan *SVM* tanpa *feature selection*. Pada pengujian algoritma *feature selection* untuk setiap algoritma klasifikasi, *Information Gain* dihasilkan sebagai algoritma terbaik untuk digunakan pada algoritma *NB*, *SVM*, dan *ANN* [8].

Penelitian ini menggunakan metode *SVM* dalam menganalisis sentimen publik di *Twitter* terhadap karapan sapi, dengan menggunakan *information gain* untuk mendapatkan *feature selection* terbaik. Opini masyarakat ini diambil dari komentar dan pendapat yang disampaikan melalui jejaring sosial media yaitu *Twitter* yang mana tujuannya untuk mengetahui penerapan metode *Support vector Machine* dan *Information Gain*.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Melihat latar belakang penelitian, opini publik di *Twitter* adalah sumber informasi berharga bagi pemangku kepentingan yang dapat dijadikan sebagai evaluasi terhadap tradisi karapan sapi. Informasi ulasan mengenai karapan sapi terdapat di *Twitter*. Maka dari itu, diperlukan sistem analisis untuk mendapatkan informasi mengenai persepsi publik terhadap karapan sapi berdasarkan sentimen atau ulasan masyarakat, apakah ulasan tersebut bersifat positif dan negatif. Untuk dapat mengetahui opini masyarakat diperlukan suatu metode yang dapat mengklasifikasikan opini mereka secara akurat agar mendapatkan suatu informasi. Penerapan *Support Vector Machine* menghasilkan model klasifikasi yang cukup baik meskipun dilatih dengan himpunan data yang relatif kurang atau sedikit, tetapi juga memiliki kelemahan terkait himpunan dengan jumlah dimensi yang sangat besar dikarenakan membutuhkan waktu training yang lama. Karena itu

dibutuhkan pemilihan parameter yang optimal diantaranya C dan Γ dengan hyperparameter dengan *Grid Search* dan seleksi fitur *Information Gain*.

1.2.2 Metode Usulan

Berdasarkan permasalahan yang diidentifikasi, penulis mengusulkan melakukan analisis sentimen ulasan publik di Platform X (*Twitter*) terhadap karapan sapi yang digunakan metode *Support Vector Machine* (SVM) dengan *Hyperparameter Tuning* dengan *gridsearch* dan feature selection *Information Gain*.

1.2.3 Pertanyaan Penelitian

Berapa nilai *accuracy* dan waktu komputasi terkait sentimen karapan sapi pada Platform X (*Twitter*) dengan menggunakan Algoritma *Support vector Machine* (SVM) dengan penerapan *Information Gain* dibandingkan dengan *Support Vector Machine* tanpa *Information Gain* ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Dibawah adalah tujuan dalam penelitian ini, yaitu :

1. Mengetahui pengaruh *Hyperparameter Tuning* dengan *Grid Search* untuk nilai akurasi pada proses penerapan *Information Gain*.
2. Mengetahui pengaruh *Information Gain* terhadap metode *Support Vector Machine*.
3. Mengetahui perbandingan hasil analisa model kinerja *Support Vector Machine* dengan *Information Gain* dan *Support Vector Machine* tanpa *Information Gain* untuk implementasi data karapan sapi.

1.3.2 Manfaat

Meneliti dalam hal ini diharapkan memberikan beberapa manfaat, yaitu:

1. Mampu menguji performansi dari algoritma *Support Vector Machine* sebelum dan sesudah penerapan hyperparameter *Grid Search* dan *Information Gain*.

2. Model dapat diterapkan untuk analisa opini terhadap data karapan sapi dalam memberikan edukasi lebih lanjut.

1.4 Batasan Masalah

Dibawah ini adalah beberapa batasan masalah pada penelitian ini:

1. Menggunakan data opini karapan sapi dari *Twitter* dengan kata kunci “karapan sapi”, yang tercatat mulai bulan Mei 2023 hingga bulan Juli 2024, sebanyak 646 opini.
2. Menggunakan akan diteliti ini dilakukan pelabelan dengan bantuan guru Sastra dan Bahasa Indonesia.
3. Kumpulan opini pada penelitian ini akan diklasifikasikan ke dalam kelas negatif dan positif
4. Dilakukannya klasifikasi sentimen metode *Support Vector Machine* dengan *Hyperparameter Grid Search* dan *Information Gain*.
5. *Hyperparameter Tuning* yang digunakan yaitu *C* dan *Gamma*.
6. Data yang digunakan 2 class yaitu positif dan negatif.
7. Pakar untuk validasi data dilakukan oleh guru Sastra dan Bahasa Indonesia.