BAB I
PENDAHULUAN

1.1. Latar Belakang

Di era digital, interaksi antara masyarakat dengan lembaga pemerintah semakin sering dilakukan melalui platform online, termasuk website. Salah satu platform utama adalah website Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Sidoarjo, yang berfungsi sebagai media informasi dan komunikasi antara pemerintah dan masyarakat. Website ini menyediakan fitur komentar pada setiap postingan, yang memungkinkan masyarakat untuk menyampaikan kritik, saran, atau pertanyaan terkait pelayanan. Namun, dengan meningkatnya jumlah komentar dari masyarakat, tidak semua dapat dikontrol dengan baik. Sering kali ditemukan komentar yang tergolong *spam* pada postingan(Syam et al., 2020). Komentar *spam* biasanya mengandung konten yang tidak relevan, seperti promosi bisnis, tautan, atau informasi lain yang bersifat komersial. Komentar semacam ini sering kali bersifat berulang dan mengganggu(Zena Lusi et al., 2024). Komentar yang relevan atau *non-spam* sangat penting karena dapat menjadi sarana evaluasi bagi peningkatan kualitas pelayanan. Dengan memisahkan komentar yang relevan, petugas dapat lebih fokus dalam menangani keluhan atau pertanyaan yang membutuhkan perhatian, seperti masalah keterlambatan pengurusan izin. Kondisi ini tidak hanya menurunkan kualitas pelayanan, tetapi juga menghambat identifikasi komentar penting yang seharusnya menjadi prioritas. Oleh karena itu, diperlukan metode klasifikasi *text* yang lebih akurat untuk membedakan komentar *spam* dan *non-spam* secara otomatis.

Klasifikasi *text* adalah Proses klasifikasi dokumen atau *text* melibatkan pengelompokan ke dalam kelas atau kategori tertentu berdasarkan karakteristik unik yang terdapat dalam *text* tersebut. Metode ini sering dimanfaatkan dalam analisis *text* berskala besar, seperti klasifikasi sentimen, pengelompokan topik, atau pendeteksian *Spam* dalam email. Salah satu algoritma yang sering digunakan dalam *text mining* adalah *Support Vector Machine (SVM)*, yang bertujuan untuk

memprediksi kategori atau label dokumen berdasarkan fitur-fitur yang diperoleh dari *text*. Dalam penelitian ini, algoritma *Support Vector Machine (SVM)* dipilih karena kemampuannya yang unggul dalam menangani berbagai jenis data, termasuk data *text*. Algoritma ini bekerja dengan mencari *hyperplane* terbaik yang mampu memisahkan dua kelas, yaitu *spam* dan *non-spam.*(Setiawan & Suryono, 2024)

Metode *Support Vector Machine* (*SVM*) memiliki keunggulan dalam menangani data berdimensi tinggi dan performa yang baik, tetapi algoritma ini juga memiliki keterbatasan ketika dihadapkan pada dataset yang mengandung banyak *noise* dan tidak seimbang. Tantangan besar muncul ketika jumlah komentar yang bersifat *spam* (tidak relevan) jauh lebih dominan dibandingkan dengan komentar *non-spam* (relevan dan bermakna). Kondisi ini dapat mengganggu klasifikasi dalam *SVM*. Ketidakseimbangan data antara komentar *spam* dan *non-spam* menjadi masalah yang signifikan dalam proses klasifikasi. Ketidakseimbangan dataset adalah masalah umum dalam tugas klasifikasi, di mana jumlah sampel pada satu kelas jauh lebih banyak dibandingkan dengan kelas lainnya. Masalah ini perlu diatasi karena akurasi tinggi tidak hanya ditentukan oleh algoritma yang diterapkan, tetapi juga dipengaruhi oleh karakteristik *dataset* itu sendiri. Ketidakseimbangan ini dapat mengurangi performa model klasifikasi, karena model cenderung lebih fokus mempelajari data dari kelas mayoritas dibandingkan kelas minoritas(Nurhopipah & Magnolia, 2022).Oleh sebab itu, penelitian yang dilakukan perlu beberapa metode untuk memodifikasi *imbalanced dataset* dengan harapan dapat meningkatkan kinerja model klasifikasi

Berbagai teknik telah dikembangkan untuk mengatasi masalah ketidakseimbangan data (*imbalanced dataset*), salah satunya *SMOTE.* Metode *Synthetic Minority Oversampling Technique* (*SMOTE*) adalah teknik *oversampling* yang dirancang untuk menangani ketidakseimbangan kelas dengan cara memperbanyak data pada kelas minoritas menggunakan data sintetis. *SMOTE* menghasilkan data sintetis melalui proses interpolasi antara data asli dari kelas minoritas (Azy Mushofy Anwary et al., 2021)

Berdasarkan data komentar yang harus dibedakan antara *spam* dan *non-spam* serta ketidakseimbangan data komentar *spam* dan *non-spam*, penelitian ini melakukan klasifikasi *spam* dan *non-spam* komentar web DPMPTSP menggunakan metode *SVM* dan *SMOTE*. Penerapan *SMOTE* dalam penelitian yang dilakukan dapat meningkatkan performa model *Support Vector Machine* **(***SVM***)**, terutama dalam mendeteksi kelas *non-spam*, tanpa mengurangi akurasi dalam memprediksi kelas *spam*. Dengan kombinasi metode *SVM* dan *SMOTE*, penelitian ini bertujuan untuk menghasilkan model klasifikasi yang dalam membedakan komentar *spam* dan *non-spam* pada komentar website DPMPTSP. Hasil dari klasifikasi ini diharapkan dapat membantu pihak terkait dalam menyaring komentar yang relevan sehingga meningkatkan kualitas pelayanan kepada masyarakat.

1.2. Rumusan Masalah

a. Bagaimana penerepan metode *SVM* dan *SMOTE* pada klasifikasi *spam* dan *non-spam* komentar web DPMPTSP
b. Bagaimana hasil akurasi penggunaan metode *SVM* dan *SMOTE* pada klasifikasi *spam* dan *non-spam* komentar web DPMPTSP

.

1.3. Tujuan Penelitian

a. Menerapkan metode *Support Vector Machine* (*SVM*) dan *Synthetic Minority Oversampling Technique* (*SMOTE*) untuk melakukan klasifikasi komentar *spam* dan *non-spam* pada website DPMPTSP kabupaten sidoarjo..
b. Mengevaluasi tingkat akurasi metode *SVM* dan *SMOTE* dalam proses klasifikasi komentar *spam* dan *non-spam* pada website DPMPTSP kabupaten sidoarjo.

1.4. Batasan Masalah

a. Data komentar yang digunakan dalam penelitian hanya berasal dari website DPMPTSP Kabupaten Sidoarjo.

b. Penelitian yang dilakukan hanya menggunakan algoritma *SVM* sebagai metode klasifikasi utama, dengan penerapan *SMOTE* untuk menangani ketidakseimbangan data.

1.5. Manfaat Penelitian

a. Membantu DPMPTSP Kabupaten Sidoarjo dalam menyaring komentar relevan sehingga meningkatkan kualitas pelayanan kepada masyarakat.
b. Memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang klasifikasi *text* dengan metode *SVM* dan *SMOTE*.

.

1.6. Sistematika Penulisan

a. BAB I PENDAHULUAN

Bab ini merupakan penjelasan secara umum dan ringkas yang menggambarkan dengan tepat isi penelitian. Isi bab ini meliputi: Latar belakang penelitian, Rumusan Masalah, Tujuan Penelitian, Batasan Dan Asumsi Penelitian, Sistematika Penulisan dan Jadwal Penelitian.

b. BAB II TINJAUAN PUSTAKA

Bab ini berisi teori umum serta penelitian terdahulu dan dilanjutkan dengan alasan pemilihan teori

c. BAB III METEDOLOGI PENELITIAN

Bab ini menjelaskan pendekatan, metode, dan teknik yang diterapkan dalam proses pengumpulan dan analisis data untuk menjawab pertanyaan penelitian. Bagian ini mencakup penjabaran mengenai: Proses Penelitian, Pengumpulan Data, *Pre-Processing*, *SMOTE*, *TF-IDF*, Klasifikasi *SVM* dan Pengujian

d. BAB IV HASIL DAN PEMBAHASAN

Bab ini menegaskan pendekatan, metode, dan teknik yang digunakan untuk mengumpulkan dan menganalisis temuan yang dapat menjawab masalah penelitian. Bab ini terdiri dari dua bagian: bagian pertama memaparkan hasil penelitian, sementara bagian kedua menyajikan pembahasan atau analisis atas hasil tersebut. Setiap aspek pembahasan dianjurkan dimulai dari hasil analisis data, kemudian diinterpretasikan, dan diakhiri dengan

penarikan kesimpulan. Pembahasan juga sebaiknya dibandingkan dengan penelitian sebelumnya atau teori-teori yang relevan.

e. BAB V KESIMPULAN DAN SARAN

Kesimpulan merupakan jawaban atas pertanyaan penelitian dan menjadi dasar untuk memberikan saran yang relevan dengan manfaat penelitian.