

BAB I

PENDAHULUAN

1.1 Latar Belakang

Platform media sosial telah memberikan dampak yang signifikan dan telah menjadi sangat penting dalam berbagai domain keberadaan individu sehari-hari dalam era digitalisasi *modern*. X sebagai situs jejaring sosial terkemuka karena *platform* yang ramah pengguna dan ketersediaan yang luas. Saat ini, X memiliki basis pengguna sekitar 330 juta orang aktif setiap bulan, dengan pertumbuhan informasi yang sangat cepat [1]. Melalui penggunaan X, individu memiliki kemampuan untuk terlibat dalam berbagai tugas seperti berkomunikasi dengan orang lain, mendokumentasikan acara harian, memfasilitasi promosi perdagangan, terlibat dalam debat, dan berbagi sudut pandang dengan menyusun pembaruan status ringkas yang dikenal sebagai tweet, yang tidak boleh melebihi 280 karakter. Tweet yang dihasilkan oleh pengguna ini berfungsi sebagai aset berharga untuk mengevaluasi sikap publik tentang subjek yang dibahas, karena mereka merangkum pendapat yang dapat dimanfaatkan untuk tujuan analitis dan pengambilan keputusan [2].

Tabungan Perumahan Rakyat (Tapera) dirancang untuk menawarkan periode pembiayaan murah guna mendukung pengembangan perumahan yang terjangkau bagi masyarakat Indonesia. Program ini merupakan salah satu upaya pemerintah dalam mengatasi krisis kepemilikan rumah, khususnya bagi kalangan masyarakat berpenghasilan rendah (MBR). Untuk mempercepat penyelesaian tantangan pembiayaan perumahan, tercatat backlog perumahan sebanyak 13 juta unit pada akhir 2015, yang berhasil ditekan menjadi 7,6 juta unit pada 2019. Pemerintah telah menjanjikan manfaat Tapera bagi Individu Berpenghasilan Rendah (MBR) yang saat ini belum memiliki properti tempat tinggal. Partisipasi dalam skema Tapera bersifat wajib bagi seluruh karyawan yang berpenghasilan di atas upah minimum, baik pekerja formal di sektor swasta maupun aparatur sipil negara. Pelaksanaan program Tapera didasarkan pada prinsip gotong royong, di mana dana yang dikumpulkan melalui potongan gaji peserta akan dikelola dan diinvestasikan untuk

membayai kepemilikan rumah bagi MBR. Individu dengan penghasilan di atas batas penerima bantuan tidak akan memperoleh subsidi pembiayaan rumah, namun tetap menerima akumulasi tabungan mereka beserta hasil pengembangannya saat memasuki masa pensiun. Dana Tapera dikelola secara profesional dan transparan oleh Badan Pengelola Tapera (BP Tapera) untuk memastikan bahwa tujuan jangka panjang program, yaitu pemerataan kepemilikan rumah dan peningkatan kesejahteraan masyarakat, dapat tercapai secara berkelanjutan. [3]. Ada banyak reaksi dari masyarakat umum mengenai adanya Tapera. Komentar yang mereka berikan menunjukkan berbagai sudut pandang. Mulai dari penolakan, mendukung, dan netral akan peraturan Tapera. Media sosial X muncul sebagai saluran untuk mengkomunikasikan reaksi individu terhadap penegakan peraturan Tapera. Selanjutnya, terdapat analisis komentar masyarakat yaitu seseorang dapat membedakan opini publik yang berlaku mengenai penerapan peraturan Tapera [4].

Dalam menghadapi berbagai reaksi masyarakat terhadap penerapan Tapera, diperlukan suatu pendekatan analitis yang mampu menangkap dan memahami opini publik secara lebih komprehensif. Salah satu pendekatan analitis berkaitan dengan penerapan metodologi dalam Natural Language Processing (NLP). Pendekatan NLP ini umumnya digunakan untuk pemrosesan data dan ekstraksi data dari platform seperti media sosial, khususnya media social X [5]. Hasil selanjutnya akan menjalani klasifikasi menggunakan metodologi model Machine Learning yang dikenal sebagai XGBoost. XGBoost dianggap sebagai teknik terkemuka dalam bidang pembelajaran mesin, dirancang khusus untuk mengatasi tantangan regresi dan klasifikasi dengan memanfaatkan prinsip-prinsip Gradient Boosting Decision Tree (GBDT) [6]. Hal ini menunjukkan bahwa penggunaan berbagai teknik optimasi seperti XGBoost dan model seperti PSO dapat meningkatkan kinerja klasifikasi secara signifikan.

Pada jurnal “Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO” [2]. Pada penelitian tersebut dilakukan analisis sentimen terhadap tweet yang membahas topik UU Cipta Kerja menggunakan metode klasifikasi teks. Data dikumpulkan melalui proses crawling

dengan menggunakan Twitter API dan sncrape, menghasilkan 1000 tweet berbahasa Indonesia yang terdiri dari 500 tweet dengan sentimen positif dan 500 tweet dengan sentimen negatif. Pengujian klasifikasi dilakukan menggunakan empat skenario dengan 10 fold cross-validation untuk membandingkan kinerja algoritma Logistic Regression, Support Vector Machine (SVM) tanpa optimasi, serta SVM yang dioptimasi dengan Particle Swarm Optimization (PSO) pada pengaturan default dan 50 iterasi. Hasil pengujian menunjukkan bahwa SVM yang dioptimasi dengan PSO menghasilkan performa yang lebih baik dibandingkan dengan Logistic Regression dan SVM tanpa optimasi, dengan peningkatan akurasi sebesar 2,01%, presisi sebesar 1,84%, dan recall sebesar 1,97% pada iterasi PSO 50 kali. Hal ini membuktikan bahwa PSO efektif dalam mengatasi kelemahan SVM dalam hal pemilihan parameter, sehingga dapat meningkatkan performa model secara keseluruhan.[2].

Berdasarkan penelitian sebelumnya yang dilakukan oleh[7] tentang analisis risiko kredit menggunakan Algoritma Extreme Gradien Boosting (XGBoost), hasil evaluasi model mengungkapkan algoritma XGBoost telah terbukti sangat efektif dalam menganalisis data kredit, khususnya dalam menangani masalah ketidakseimbangan kelas yang sering terjadi pada data kredit. Penambahan metode SMOTE (Synthetic Minority Over-sampling Technique) sebelum pemodelan terbukti meningkatkan performa model dalam memprediksi kelayakan kredit, yang ditunjukkan oleh peningkatan nilai akurasi dan AUC. Hasil evaluasi menunjukkan bahwa model dengan skenario SMOTE memiliki nilai AUC tertinggi, yaitu 0.918, yang termasuk dalam kategori "Very Good". Hal ini menandakan bahwa model yang dihasilkan mampu memisahkan kelas dengan baik, yang penting untuk pengambilan keputusan dalam proses pemberian kredit. Selain itu, penelitian ini menyoroti bahwa penggunaan SMOTE sangat membantu dalam menangani ketidakseimbangan data, yang dapat meningkatkan akurasi model secara signifikan. Penerapan algoritma XGBoost pada penelitian ini juga memperlihatkan kemampuan yang superior dibandingkan algoritma lain yang telah digunakan dalam berbagai penelitian terkait analisis risiko kredit. Dengan demikian, model XGBoost yang telah dioptimalkan dengan SMOTE memberikan

solusi yang kuat untuk meningkatkan akurasi dan keandalan prediksi dalam konteks analisis kredit, terutama dalam pengelolaan risiko kredit di sektor perbankan[7].

Penelitian ini akan berfokus pada analisis sentiment terhadap media sosial X dengan mengumpulkan dan menganalisis opini atau komentar masyarakat yang terkait dengan isu Tapera. Dengan memahami sentimen tersebut, penelitian ini bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai pandangan masyarakat terhadap Tapera. Dalam proses analisis sentiment ini, konsep Natural Language Processing (NLP) akan digunakan untuk mengekstraksi dan memproses data teks secara efektif. Untuk mencapai tujuan tersebut, penelitian ini akan menerapkan metode XGBoost, yang dikenal karena kemampuannya dalam menangani data yang kompleks dan menghasilkan performa yang optimal. Dengan demikian, penelitian ini diharapkan tidak hanya mengidentifikasi sentimen yang dominan, tetapi juga mengoptimalkan akurasi dan efisiensi dalam proses analisis sentiment menggunakan metode XGBoost.

1.2 Rumusan Masalah

Berdasarkan identifikasi masalah yang telah diuraikan dalam latar belakang, didapatkan rumusan masalah sebagai berikut :

- a. Bagaimana kinerja XGBoost dalam tugas mengkategorikan sentimen (positif, negatif, netral, dan others) dalam kaitannya dengan tweet atau posting media sosial yang berkaitan dengan Tapera?
- b. Bagaimana pengaruh algoritma PSO terhadap analisis sentiment pada studi kasus tapera?

1.3 Tujuan Penelitian

Tujuan penelitian ini untuk melakukan klasifikasi sentimen yang diekspresikan oleh pengguna media sosial X terhadap isu Tapera, dengan memanfaatkan metode pembelajaran mesin yang canggih, yaitu XGBoost yang telah dioptimalkan menggunakan teknik *Particle Swarm Optimization* (PSO). Dengan pendekatan ini, diharapkan dapat diperoleh model klasifikasi yang lebih akurat dan efisien, yang mampu memberikan detail sentimen publik secara lebih mendalam dan memberikan pemahaman yang lebih baik terhadap

persepsi masyarakat mengenai isu Tapera yang menjadi topik perdebatan.

1.4 Batasan Masalah

- a. Penelitian ini akan membatasi analisis sentimen hanya pada tweet atau opini tentang isu Tapera yang berbahasa Indonesia untuk memudahkan interpretasi hasil.
- b. Penelitian ini hanya mengimplementasikan salah satu algoritma *Machine Learning* yaitu algoritma XGBoost.

