

BAB II

TINJAUAN PUSTAKA

Banyak penelitian yang berfokus pada kecerdasan buatan khususnya dalam bidang *machine learning* akibat dari peningkatan teknologi di bidang kecerdasan buatan. Media sosial sekarang menjadi salah satu sumber informasi utama yang menyampaikan pandangan, opini, dan tanggapan masyarakat terhadap berbagai masalah aktual secara real-time. Teknik ini memudahkan analisis data kompleks seperti klasifikasi, prediksi, dan visualisasi data. Semakin banyak peneliti di seluruh dunia didorong oleh kemajuan ini untuk mengembangkan teknik yang lebih efektif dan akurat. Akibatnya, penting untuk menggunakan pendekatan berbasis *machine learning* untuk menganalisis tren dan sentimen publik terhadap masalah aktual seperti Tapera. Penelitian ini mengangkat topik analisis sentimen media sosial X terhadap masalah Tapera menggunakan metode XGBoost berbasis PSO dalam upaya untuk mendapatkan pemahaman yang lebih baik tentang persepsi masyarakat secara online.

2.1 Studi Literatur

Berbagai penelitian telah dilakukan yang berkaitan dengan analisis sentimen media sosial terhadap berbagai isu menggunakan berbagai metode. Studi-studi tersebut menyediakan fondasi teoretis dan empiris yang kokoh bagi penelitian saat ini, dengan tujuan untuk mengeksplorasi keunggulan XGBoost berbasis Particle Swarm Optimization (PSO) dalam mengklasifikasikan sentimen dibandingkan metode lainnya.

Penelitian yang dilakukan oleh T. S. Sabrila, Y. Azhar, dan C. S. K. Aditya pada tahun 2022 mengkaji optimasi algoritma Support Vector Machine (SVM) dengan menggunakan metode Particle Swarm Optimization (PSO). Hasil penelitian mereka menunjukkan bahwa optimasi ini dapat secara signifikan meningkatkan performa klasifikasi sentimen tweet jika dibandingkan dengan SVM tanpa optimasi maupun dengan metode Logistic Regression. Model SVM yang telah dioptimasi dengan PSO selama 50 iterasi berhasil

meningkatkan akurasi sebesar 2,01%, presisi sebesar 1,84%, dan recall sebesar 1,97%. Peningkatan ini mengindikasikan bahwa pendekatan berbasis PSO mampu memberikan perbaikan signifikan dalam klasifikasi sentiment [2].

Penelitian lainnya yang dilakukan oleh B. N. S. Aldo Andika Saputra pada tahun 2024 membuktikan bahwa penggunaan algoritma XGBoost yang dioptimalkan dengan metode Synthetic Minority Over-sampling Technique (SMOTE) dapat meningkatkan akurasi dan Area Under Curve (AUC) dalam analisis risiko kredit. Dalam penelitian ini, ketidakseimbangan data kredit menjadi salah satu tantangan utama dalam proses klasifikasi, dan penggunaan SMOTE berhasil mengatasi permasalahan tersebut. Model Extreme Gradient Boosting (XGBoost) yang dioptimasi dengan metode SMOTE berhasil mencapai nilai AUC sebesar 0,918. Hasil ini menunjukkan bahwa metode yang digunakan mampu meningkatkan performa prediksi dalam menentukan kelayakan kredit serta menangani distribusi data yang tidak seimbang dengan lebih efektif [7].

2.2 Implementasi Metode dalam Analisis Sentimen

Implementasi metode pembelajaran mesin dalam analisis sentimen telah berkembang pesat dengan berbagai pendekatan optimasi yang diterapkan. Support Vector Machine (SVM) dengan optimasi berbasis PSO telah terbukti mampu meningkatkan performa klasifikasi dalam analisis sentimen di media sosial. Penerapan teknik ini memberikan keuntungan dalam meningkatkan akurasi serta efisiensi pemrosesan data, terutama dalam menghadapi dataset yang kompleks dan beragam.

Selain itu, pendekatan menggunakan Extreme Gradient Boosting (XGBoost) yang dioptimalkan dengan metode SMOTE telah memberikan hasil yang menjanjikan dalam menangani permasalahan ketidakseimbangan data. Teknik ini memungkinkan model untuk belajar secara lebih efektif dari data yang memiliki distribusi tidak merata, sehingga meningkatkan kemampuan

klasifikasi dalam berbagai domain, termasuk analisis sentimen dan analisis risiko kredit.

Penelitian sebelumnya menunjukkan bahwa optimalisasi algoritma pembelajaran mesin dapat memberikan peningkatan yang signifikan dalam kinerja klasifikasi. Dengan memanfaatkan metode seperti PSO, model dapat lebih adaptif terhadap variasi data yang kompleks serta menghasilkan prediksi yang lebih akurat. Oleh karena itu, dalam penelitian ini, penerapan XGBoost berbasis PSO akan dieksplorasi lebih lanjut untuk mengevaluasi keunggulannya dalam analisis sentimen dibandingkan dengan metode lain yang telah digunakan sebelumnya.

2.3 Media Sosial X

Media sosial X digunakan sebagai sumber data karena platform ini sedang kembali populer setelah sebelumnya sempat kalah bersaing dengan media sosial lain. X juga merupakan media sosial yang mengutamakan komunikasi berbasis teks, sehingga sangat sesuai untuk pengambilan data sentimen masyarakat secara daring. Keunikan dari tweet yang dibatasi dalam jumlah karakter (maksimal 280 karakter) membuat setiap unggahan cenderung padat, langsung, dan mengandung opini atau reaksi pengguna terhadap isu tertentu secara real-time. Hal ini menjadikan X sebagai platform yang sangat relevan untuk mengamati dinamika opini publik dan menganalisis sentimen masyarakat terhadap suatu kebijakan atau peristiwa terkini. Hal ini disebabkan karena X sering digunakan masyarakat untuk menyampaikan pendapat pribadi mengenai berbagai topik yang sedang hangat dibicarakan. Dengan sifatnya yang terbuka dan real-time, Media sosial X memberikan gambaran langsung mengenai opini publik terhadap isu-isu aktual seperti Tapera[8].

2.4 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (AI) yang berfokus pada interaksi antara manusia dan komputer

menggunakan bahasa alami. NLP digunakan dalam berbagai proses *preprocessing* data teks, seperti tokenisasi, penghapusan stopword, dan stemming, untuk mempersiapkan data sebelum dianalisis lebih lanjut. Tahapan ini menjadi fondasi awal yang penting agar data teks dapat dipahami oleh algoritma machine learning dan menghasilkan analisis yang akurat[9].

2.5 TF - IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu metode pembobotan yang digunakan dalam pemrosesan teks untuk menentukan seberapa penting suatu kata dalam sebuah dokumen dibandingkan dengan seluruh kumpulan dokumen. Teknik ini menggabungkan dua konsep utama, yaitu **TF (Term Frequency)** yang menghitung frekuensi kemunculan kata dalam satu dokumen, dan **IDF (Inverse Document Frequency)** yang mengukur seberapa jarang kata tersebut muncul dalam seluruh dokumen. Kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memperoleh bobot yang lebih tinggi, karena dianggap lebih spesifik dan relevan terhadap isi dokumen tersebut. Melalui pendekatan ini, TF-IDF mampu menyeleksi kata-kata penting dari teks sehingga hasil ekstraksi fiturnya lebih bermakna untuk proses analisis lebih lanjut. Oleh karena itu, bobot yang dihasilkan dari TF-IDF sangat cocok dijadikan sebagai fitur input dalam algoritma machine learning seperti SVM atau XGBoost untuk klasifikasi teks[10].

2.6 Pemodelan XGBoost

Extreme Gradient Boosting (XGBoost) merupakan sebuah algoritma yang dikembangkan sebagai peningkatan dari metode Gradient Boosting Decision Tree (GBDT). Algoritma ini termasuk dalam kelompok teknik pembelajaran mesin (machine learning) yang sangat populer dan efektif, terutama untuk menyelesaikan berbagai permasalahan dalam domain klasifikasi dan regresi. Dibandingkan dengan metode gradient boosting konvensional, XGBoost menawarkan sejumlah keunggulan seperti kecepatan komputasi yang lebih

tinggi, efisiensi memori yang lebih baik, serta kemampuan untuk mengatasi overfitting melalui regularisasi [11].

2.7 PSO (Particle Swarm Optimization)

Particle Swarm Optimization (PSO) merupakan salah satu algoritma optimasi yang berbasis populasi dan terinspirasi dari perilaku sosial makhluk hidup yang bergerak dalam kelompok, seperti burung yang terbang dalam kawanan atau ikan yang berenang dalam gerombolan. Konsep dasar PSO didasarkan pada cara individu dalam suatu kelompok berinteraksi dan berbagi informasi untuk mencapai tujuan secara kolektif. Setiap individu atau "partikel" dalam populasi mewakili suatu solusi potensial dan bergerak dalam ruang pencarian berdasarkan pengalaman terbaiknya sendiri (personal best) dan pengalaman terbaik seluruh kawanan (global best). Mekanisme ini membuat PSO mampu menjelajahi ruang solusi secara efektif dan efisien. Dalam konteks penelitian dan pengolahan data, PSO banyak digunakan untuk menyelesaikan berbagai permasalahan optimasi, termasuk dalam proses pencarian nilai parameter terbaik atau yang dikenal dengan istilah hyperparameter tuning. Pemilihan parameter yang tepat sangat penting dalam membangun model machine learning yang optimal. Oleh karena itu, PSO sangat relevan digunakan untuk mengoptimalkan performa algoritma klasifikasi seperti XGBoost. Dengan bantuan PSO, nilai-nilai parameter seperti learning rate, max depth, dan jumlah estimators dapat disesuaikan secara otomatis untuk meningkatkan akurasi model dan menurunkan error. Dalam penelitian ini, PSO diimplementasikan sebagai metode optimasi guna menyempurnakan hasil klasifikasi sentimen oleh XGBoost, sehingga model yang dihasilkan mampu memberikan performa prediksi yang lebih baik dibandingkan tanpa optimasi[12].