

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pekan Olahraga Nasional (PON) merupakan ajang olahraga terbesar di Indonesia yang diselenggarakan setiap empat tahun sekali. PON menjadi momentum penting bagi para atlet dari seluruh provinsi di Indonesia untuk berlaga dan menunjukkan kemampuan terbaik mereka di berbagai cabang olahraga. Pada tahun 2021, PON ke-20 diadakan di Provinsi Papua, yang secara khusus mendapat perhatian besar dari masyarakat, media, dan pemerintah. PON Papua 2021 tidak hanya penting dari sisi olahraga, tetapi juga karena diadakan di wilayah yang memiliki berbagai tantangan geografis dan infrastruktur, menjadikannya peristiwa bersejarah dalam penyelenggaraan PON di Indonesia. PON Papua membawa dampak luas terhadap perkembangan daerah, baik dalam hal infrastruktur, ekonomi, maupun sosial budaya. Di tengah perkembangan teknologi, penyelenggaraan PON juga dipantau secara intensif melalui berbagai media sosial. Media sosial, sebagai platform komunikasi massa, memainkan peran penting dalam membentuk opini publik terhadap penyelenggaraan PON. Publik dapat dengan cepat menyampaikan apresiasi, kritik, serta masukan terhadap berbagai aspek dari PON, seperti pelaksanaan acara, fasilitas. Dengan tingginya volume interaksi di media sosial, penting untuk melakukan analisis sentimen untuk memahami persepsi masyarakat terhadap PON Papua. Melalui analisis ini, diharapkan dapat diperoleh wawasan yang berguna bagi penyelenggara dalam mengevaluasi dan meningkatkan kualitas PON di masa depan. Penelitian ini akan menggunakan pendekatan machine learning untuk menganalisis komentar masyarakat dan mengidentifikasi sentimen yang berkembang, sehingga dapat memberikan informasi yang relevan bagi pengembangan ajang olahraga di Indonesia[1].

Penelitian sebelumnya tentang analisis sentimen terhadap PON diplatform instagram menggunakan *Naive Bayes Classifier*. Data yang diperoleh sebanyak 2000 komentar dengan 965 negatif, 256 netral, dan 770 positif menghasilkan tingkat akurasi hingga 75% dari 4 kondisi percobaan[1]. Dalam penelitian yang dilakukan oleh Manullang, Prianto, dan Harani tentang Analisis Sentimen Untuk

Memprediksi Hasil Calon Pemilu Presiden Menggunakan *Lexicon Based* dan *Random Forest* menunjukkan hasil akurasi model sebesar 88% untuk metode *Lexicon Based* dan 94% untuk metode *Random Forest*[2]. Dalam penelitian yang dilakukan oleh Pamuji dan Ramadhan tentang *Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy* menunjukkan bahwa metode *Random Forest* memiliki akurasi lebih tinggi yaitu 85,5%, dibandingkan dengan *Decision Tree* yang memiliki akurasi sebesar 84,4% dalam memprediksi keberhasilan pengobatan *immunotherapy* pada penyakit kutil [3].

Berdasarkan penelitian terdahulu, penggunaan metode *Random Forest* dalam analisis sentimen menunjukkan hasil yang lebih unggul dibandingkan dengan algoritma *Support Vector Machine* dan *Naive Bayes*. Oleh karena itu, pada penelitian ini peneliti menggunakan metode *Random Forest* untuk analisis sentimen PON Papua XX pada platform X. Dalam konteks ini, penggunaan metode *Random Forest* muncul sebagai pendekatan yang potensial dalam mengolah data sentimen yang beragam.[1].

Dalam penelitian ini, metode *Random Forest* akan didukung oleh penerapan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan data yang mungkin terjadi dalam klasifikasi sentimen. SMOTE digunakan untuk menghasilkan sampel sintesis bagi kelas minoritas (misalnya, sentimen negatif atau netral yang mungkin memiliki jumlah lebih sedikit dibanding sentimen positif), sehingga distribusi data menjadi lebih seimbang. Dengan mengurangi risiko bias yang disebabkan oleh ketidakseimbangan data, diharapkan model *Random Forest* dapat bekerja lebih efektif dalam mengklasifikasikan sentimen, yang pada akhirnya meningkatkan akurasi dan generalisasi hasil.[13]

Random Forest adalah suatu teknik dalam *machine learning* yang memanfaatkan kumpulan pohon keputusan (*Decision Trees*) untuk mengatasi keberagaman dan kompleksitas data. Pendekatan ini didasarkan pada konsep *ensemble learning*, beberapa model digabungkan untuk meningkatkan kinerja dan keakuratan prediksi. Penerapan metode *Random Forest* pada analisis sentimen

memungkinkan identifikasi pola-pola yang lebih kompleks dalam data teks, yang mungkin sulit untuk dipahami oleh algoritma tunggal. Salah satu keunggulan utama dari *Random Forest* adalah kemampuannya untuk menangani berbagai jenis data, baik data kategorikal maupun data numerik, tanpa memerlukan pra-pemrosesan yang rumit. Hal ini membuatnya sangat cocok untuk digunakan dalam analisis sentimen, di mana data teks sering kali memiliki sifat-sifat yang beragam dan kompleks. Dalam analisis sentimen untuk ulasan aplikasi, penggunaan metode *Random Forest* membutuhkan ekstraksi fitur TF-IDF untuk mengonversi teks ulasan menjadi data nilai numerik yang dapat dipahami oleh algoritma.

Term Frequency- Inverse Document Frequency (TF-IDF) adalah metode yang dilakukan untuk menghitung pembobotan pada setiap kata dalam suatu dokumen atau kalimat. Dengan penerapan TF-IDF sebagai ekstraksi fitur dan metode *Random Forest* sebagai metode klasifikasi dalam sistem analisis sentimen terhadap PON Papua XX

1.2 Perumusan Masalah

1.2.1 Permasalahan

Permasalahan pada klasifikasi *Random Forest* dengan data PON adalah terjadi ketidakseimbangan kelas, sehingga membutuhkan proses sampling untuk menyeimbangkan data.

1.2.2 Metode Usulan

Peneliti mengusulkan untuk melakukan analisis sentimen menggunakan metode *Random Forest* yang memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode penelitian-penelitian sebelumnya.

1.2.3 Pertanyaan Penelitian

1. Berapakah nilai akurasi terbaik pada perubahan parameter dengan metode *Random Forest* dalam analisis sentimen menggunakan data dari komentar dari platform x

2. Berapakah nilai akurasi dan terbaik pada data balance dan imbalance dengan metode *Random Forest* dalam analisis sentimen menggunakan data dari komentar dari platform x

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan Penelitian

1. Setelah pertanyaan penelitian, tujuan yang akan dihasilkan dari penelitian ini adalah nilai akurasi terbaik pada perubahan parameter metode *Random Forest* dalam analisis sentimen terhadap PON Papua dalam platform X
2. Dapat mengatasi imbalance data untuk klasifikasi sentimen.

1.3.2 Manfaat Penelitian

Menggunakan metode *Random Forest* dalam analisis sentimen dapat berkontribusi pada pengembangan model-model pembelajaran mesin yang lebih akurat dalam memahami sentimen secara umum, serta memberikan referensi bagi penelitian sejenis di masa depan.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini sebagai berikut:

1. Data yang digunakan bersumber dari sosial media X
2. Data *tweets* dengan kata kunci PON Papua sebanyak 2000 *tweets*.
3. Komentar diambil pada tanggal 11 November 2024 dengan timeline tweet tanggal 1 September 2021 sampai 30 Oktober 2021
4. Komentar menggunakan Bahasa Indonesia
5. Komentar pada penelitian ini akan diklasifikasikan ke dalam kelas negatif, positif, dan netral.
6. Menggunakan metode klasifikasi *Random Forest* untuk membangun pohon keputusan.
7. Dalam melakukan analisis menggunakan bahasa pemrograman *Python*.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dari penelitian ini yaitu:

1. **BAB I Pendahuluan**

Pada bab I berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

2. **BAB II KAJIAN PUSTAKA**

Pada bab II berisi tentang teori-teori penunjang dan penelitian terdahulu yang berkaitan dengan penelitian ini.

3. **BAB III METODE PENELITIAN**

Bab III berisi tentang proses atau langkah-langkah dalam memecahkan masalah dengan menggunakan metode yang telah ditetapkan.

4. **BAB IV HASIL DAN PEMBAHASAN**

Bab IV berisi hasil implementasi sistem dengan metode yang digunakan, yang terdiri dari lingkungan uji coba dan hasil uji coba.

5. **BAB V PENUTUP**

Bab V berisi kesimpulan dari penelitian yang disesuaikan dengan pernyataan penelitian dan saran terkait pengembangan aplikasi selanjutnya.