

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Naturalisasi adalah proses peralihan status hukum seseorang dari warga negara asing menjadi warga negara suatu negara melalui mekanisme yang telah ditetapkan. Proses ini dapat dilakukan oleh orang asing yang ingin menjadi warga negara Indonesia [1]. Naturalisasi dalam sepak bola merupakan suatu proses legal yang dilalui oleh seorang individu untuk memperoleh kewarganegaraan baru dengan tujuan utama mewakili tim nasional negara tersebut dalam ajang kompetisi internasional. Dalam konteks sepak bola, naturalisasi telah menjadi salah satu strategi yang digunakan untuk meningkatkan kualitas tim nasional [2].

Program naturalisasi pemain telah dilakukan oleh Persatuan Sepak Bola Seluruh Indonesia (PSSI) sejak 2010. Contoh pemain yang melakukan naturalisasi biasa saat itu ialah Christian Gonzales dan Marc Klok yang telah tinggal selama 5 tahun di Indonesia [3, 4]. Selain naturalisasi jalur biasa, Indonesia juga melakukan naturalisasi pemain dengan jalur istimewa. PSSI mulai gencar melakukan naturalisasi pemain keturunan yang bermain di Eropa seperti Sandy Walsh, Justin Hubner, Ivar Jenner dan lain-lain [5]. Pemain tersebut direkomendasikan untuk dinaturalisasi oleh PSSI karena pemain tersebut memiliki keturunan darah Indonesia dan bermain di liga Eropa, yang mana PSSI menganggap bahwa pemain tersebut memiliki keahlian yang dapat memperkuat timnas sepak bola Indonesia [6].

Kehadiran pemain naturalisasi di dalam tim nasional Indonesia banyak menuai pro dan kontra. Salah satu contoh yang mendukung adanya pemain naturalisasi, Justinus Lhaksana dan Binder Singh yang merupakan pengamat sepak bola menyatakan selama membela Indonesia, memiliki alasan bahwa tidak peduli latar belakang pemain, tetap harus bangga karena membela Merah Putih saat bertanding. Lalu pemain naturalisasi yang dicari merupakan pemain yang berkualitas dan bermain di *level* Eropa dengan tujuan permainan tim nasional Indonesia menjadi lebih baik [7]. Adapun kontra pada pemain naturalisasi yang dikemukakan oleh Alex Pulalo dan Rully Nere yang merupakan mantan pesepak bola Indonesia menolak hadirnya pemain naturalisasi karena kualitas pemain naturalisasi tidak jauh beda dengan pemain lokal dan masih banyak pemain lokal

yang lebih bagus ketimbang pemain naturalisasi yang ada dan hadirnya pemain naturalisasi dapat memotivasi para pemain muda untuk menjadi pemain masa depan Indonesia [7, 8]. Peforma timnas sepak bola Indonesia pada 2010 hingga 2020 saat itu memiliki prestasi yang kurang baik. Pemain yang dinaturalisasi oleh PSSI saat itu dinilai kurang baik karena usia yang sudah cukup berumur dan performa yang tidak maksimal karena sudah tidak di usia produktif [9].

Analisis sentimen adalah suatu teknik yang digunakan untuk menganalisis teks secara komputasional guna mengekstraksi informasi mengenai polaritas emosional yang terkandung di dalamnya [10]. Analisis sentimen berfokus pada mempelajari opini, perasaan, evaluasi, sikap, dan emosi orang-orang terhadap berbagai entitas. Opini ini dapat berupa produk, layanan, organisasi, individu, isu, topik, peristiwa, dan atributnya [11]. Analisis sentimen dapat digunakan untuk menilai sentimen masyarakat terhadap program naturalisasi pemain timnas sepak bola Indonesia. Data yang didapat akan diklasifikasikan menjadi dua kategori yaitu positif dan negatif.

Dataset yang tidak seimbang merupakan kendala umum dalam pengambilan data dan dapat memengaruhi akurasi klasifikasi data secara signifikan. Dalam *dataset* yang tidak seimbang, model klasifikasi sering kali bias terhadap kelas mayoritas, sehingga menghasilkan prediksi yang tidak akurat untuk kelas minoritas [12]. *Oversampling* adalah salah satu teknik dalam mengatasi masalah ketidakseimbangan *dataset*. Teknik ini meningkatkan jumlah data di kelas minoritas, sehingga distribusi data menjadi lebih seimbang dan model klasifikasi dapat belajar dengan lebih baik dari kedua kelas [13].

Penelitian yang telah dilakukan terkait analisis sentimen membahas sentimen ulasan terhadap aplikasi *e-government* yang dikembangkan oleh pemerintah Indonesia menggunakan *Naive Bayes*. Hasil penelitian menunjukkan bahwa 407 ulasan memiliki sentimen positif, sedangkan 235 ulasan memiliki sentimen negatif. Algoritma *Naive Bayes* mencapai akurasi sebesar 89%, presisi sebesar 83%, dan *recall* sebesar 87% [14].

Penelitian lainnya juga yang membahas terkait sentimen program kampus merdeka pada Twitter menggunakan algoritma *Naive Bayes* [15]. Pada penelitian terkait sentimen kampus merdeka, data dibagi dengan rasio 90:10. Hasil dari penelitian menggunakan metode *Naive Bayes* menghasilkan akurasi sebesar 97.92% .

Penelitian lainnya yang membahas perbandingan model *Naive Bayes* dalam sentimen kenaikan harga BBM pada media sosial Twitter [16]. Model yang diuji

yakni *Multinomial Naive Bayes* dan *Complement Naive Bayes*. Model *Complement Naive Bayes* yang dikembangkan berhasil mencapai akurasi keseluruhan sebesar 82%. Evaluasi lebih lanjut menunjukkan nilai *precision* sebesar 89%, *recall* sebesar 88%, dan *F1-score* sebesar 89%, mengindikasikan kinerja model yang cukup baik dalam mengklasifikasikan data.

Dalam penelitian yang membahas terkait *Support Vector Machine* dengan *kernel linear*, *polynomial*, *rbf* dengan sentimen batalnya piala dunia u-20 di Indonesia [10]. Pembagian data pada penelitian ini menggunakan rasio 80:20 sebagai yang terbaik. *Kernel rbf* mendapatkan hasil yang lebih baik dibanding *kernel linear* dan *polynomial* yaitu akurasi sebesar 78.15%, *F1-score* sebesar 76.30%, *precision* sebesar 77.37% dan *recall* sebesar 75.58%.

Penelitian lainnya juga membahas perbandingan antara *Support Vector Machine* dengan *Naive Bayes* pada analisis sentimen terkait *cyberbullying* dan menggunakan fitur ekstraksi [17]. Hasil yang didapat menggunakan algoritma *Naive Bayes* dengan akurasi sebesar 83.85% tanpa menggunakan fitur ekstraksi dan 90.77% dengan menggunakan fitur ekstraksi. Hasil lainnya menggunakan algoritma *Support Vector Machine* memiliki akurasi sebesar 82.31% tanpa menggunakan fitur ekstraksi dan 90% dengan menggunakan fitur ekstraksi.

Pada penelitian lainnya membahas sentimen terkait *e-wallet* yang digunakan di Indonesia yaitu OVO dan DANA menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* [18]. Hasil yang didapat menggunakan algoritma *Naive Bayes* untuk *e-wallet* OVO memiliki akurasi sebesar 94.90% dan *e-wallet* DANA sebesar 94.70%. Hasil lainnya menggunakan algoritma *Support Vector Machine* untuk *e-wallet* OVO memiliki akurasi sebesar 91.00% dan *e-wallet* DANA sebesar 86.60%.

Penelitian lain yang membahas perbandingan antara algoritma *Naive Bayes* dan *Support Vector Machine* dalam analisis sentimen pada ulasan *e-commerce* Shopee [19] menunjukkan bahwa algoritma *Naive Bayes* menghasilkan akurasi sebesar 85%, sementara *Support Vector Machine* memperoleh akurasi 81%. Dalam hal hasil sentimen, penelitian ini menemukan bahwa 54.5% ulasan bersifat positif, sementara 45.5% bersifat negatif. Hasil ini mengindikasikan bahwa meskipun *Naive Bayes* unggul dalam hal akurasi, kedua algoritma memiliki performa yang relatif seimbang dalam analisis sentimen. Pilihan algoritma yang tepat dapat secara signifikan mempengaruhi hasil analisis dan kesimpulan yang diperoleh dari penelitian. Oleh karena itu, penelitian ini menunjukkan pentingnya tahap pemilihan algoritma yang cermat.

Kedua algoritma yaitu *Support Vector Machine* dan *Naive Bayes* memiliki kelebihan dan kekurangan [20]. Kelebihan dari *Support Vector Machine* yaitu dapat melakukan klasifikasi dengan data yang kompleks dan akurasi dapat dicapai tinggi dengan setelan parameter yang sesuai. Kelebihan dari *Naive Bayes* yaitu performa yang cukup cepat dalam klasifikasi dan efisien dalam melakukan *training* model. Kekurangan pada *Support Vector Machine* adalah untuk mencari akurasi yang terbaik, penyetelan parameter cukup kompleks dan performa yang cukup lambat dibandingkan dengan *Naive Bayes*. Kekurangan *Naive Bayes* terletak pada jika data yang dimiliki memiliki *noise* sehingga mengacaukan hasil prediksi data.

Penelitian yang membahas bagaimana cara menangani *imbalance dataset* pada klasifikasi komentar program kampus merdeka pada aplikasi Twitter melakukan metode *balancing data* pada metodenya [12]. Metode *balancing data* yang dibahas yaitu SMOTE, ADASYN dan *random combination sampling*. Hasil dari penelitian menunjukkan dari sekian metode *balancing data*, ADASYN unggul dari segi *F1-score* dan akurasi. Maka dari itu metode *balancing* ADASYN dipilih sebagai metode *balancing data* pada penelitian ini.

Penelitian yang serupa terkait analisis sentimen pemain naturalisasi tim nasional Indonesia u-23 pada era Shin Tae Young menggunakan *Naive Bayes* dan *K-nearest neighborhood* mendapatkan akurasi sebesar 97.84% pada *Naive Bayes* dan 99.87% pada *K-Nearest Neighborhood* [21]. Penelitian lainnya yaitu analisis sentimen terhadap naturalisasi pemain Timnas Indonesia berdasarkan opini Twitter menggunakan *Naive Bayes* mendapatkan total akurasi sebesar 72% dengan *precision* data positif sebesar 28%, data negatif sebesar 82%, *recall* dengan data positif sebesar 26% dan data negatif sebesar 83% [22]. Perbedaan dari penelitian yang akan dilakukan kali ini ada pada data dilakukan metode *balancing*. Data yang sudah diseimbangkan dapat meningkatkan performa dari model yang dilatih.

Berdasarkan permasalahan yang ada, dilakukan penelitian analisis sentimen pada Twitter terkait program naturalisasi pemain Timnas sepak bola Indonesia. Tujuan penelitian ini adalah untuk mengetahui sentimen opini masyarakat terhadap pro dan kontra terkait program naturalisasi pemain Timnas sepak bola Indonesia. Lalu permasalahan yang sudah lebih dari 10 mengenai program naturalisasi membuat penelitian cukup penting dilakukan. Algoritma yang digunakan dalam penelitian ini adalah *Naive Bayes* karena memiliki akurasi yang lebih baik dibandingkan dengan *Support Vector Machine* [17, 18, 19]. Dengan penerapan algoritma *Naive Bayes* juga dapat mempersingkat waktu pengklasifikasian dalam analisis sentimen [14].

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan sebelumnya, permasalahan penelitian ini dapat dirumuskan sebagai berikut ;

1. Bagaimana cara mengimplementasikan algoritma *Naive Bayes* untuk menganalisa sentimen mengenai program naturalisasi pemain di timnas Indonesia pada media sosial Twitter?
2. Berapa tingkat akurasi algoritma *Naive Bayes* dalam melakukan analisis sentimen mengenai program naturalisasi pemain di timnas Indonesia?

1.3 Batasan Permasalahan

Batasan-batasan masalah yang akan menjadi fokus penelitian ini adalah sebagai berikut ;

1. Data yang digunakan dalam analisis sentimen penelitian ini berbahasa Indonesia dan berupa teks terkait program naturalisasi pemain.
2. Data yang diambil bersumber dari Twitter pada awal tahun 2023 hingga akhir Maret tahun 2024.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, penelitian ini bertujuan untuk melakukan analisis sentimen program naturalisasi pemain timnas sepak bola Indonesia pada media sosial Twitter adalah;

1. Mengimplementasikan algoritma *Naive Bayes* dalam analisis sentimen mengenai program naturalisasi pemain di timnas Indonesia.
2. Menguji keakuratan dan kinerja algoritma *Naive Bayes* dalam analisis sentimen mengenai mengenai program naturalisasi pemain di timnas Indonesia.
3. Mengetahui sentimen masyarakat mengenai program naturalisasi pemain di timnas Indonesia.

1.5 Manfaat Penelitian

Penelitian analisis sentimen pada media sosial Twitter terkait program naturalisasi pemain timnas sepak bola Indonesia menggunakan algoritma *Naive Bayes* ini diharapkan dapat memberikan beberapa manfaat ke berbagai pihak sebagai berikut:

1. Memberikan informasi kepada masyarakat tentang sentimen masyarakat terhadap program naturalisasi pemain timnas sepak bola Indonesia.
2. Penelitian ini dapat dijadikan referensi opini masyarakat sebagai bahan evaluasi program naturalisasi pemain yang dilakukan oleh PSSI.
3. Mengetahui performa dan akurasi terkait algoritma *Naive Bayes* dan dapat digunakan sebagai bahan referensi penelitian selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab 1 membahas mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian.
- Bab 2 LANDASAN TEORI
Bab 2 membahas landasan teori terkait dengan naturalisasi, analisa sentimen, *text-preprocessing*, TF-IDF, ADASYN, *Naive Bayes* dan *Confusion Matrix*.
- Bab 3 METODOLOGI PENELITIAN
Bab 3 membahas terkait dengan alur dan metode penelitian.
- Bab 4 HASIL DAN DISKUSI
Bab 4 membahas hasil penelitian yang telah dilakukan.
- Bab 5 KESIMPULAN DAN SARAN
Bab 5 membahas kesimpulan dan saran yang bisa diambil pada penelitian.