

BAB I PENDAHULUAN

1.1 Latar Belakang

Berita merupakan sumber penyebaran informasi yang penting dalam kehidupan sehari-hari. Berita membuat suatu informasi cepat tersebar dari satu sumber ke seluruh khalayak masyarakat. Dengan semakin berkembangnya teknologi pada zaman globalisasi seperti saat ini, cara yang efektif dilakukan untuk melakukan penyebaran berita secara cepat adalah melalui platform media *online*, seperti Detik.com, Kompas.com, dan lain-lain. Saat ini, penggunaan internet di Indonesia mencapai angka 63 juta pengguna dengan persentase sebanyak 95% menggunakan internet untuk mengakses informasi (Sahid, 2023).

Keberadaan media *online* cukup membantu dalam segi pencarian informasi yang dapat dilakukan secara cepat tanpa terhambat batasan ruang dan waktu, tetapi dalam beberapa kasus berita pada media *online* memuat berita hoax menyesatkan. Berita hoax merupakan berita yang tidak diketahui kebenarannya dan berpotensi dalam menyesatkan manusia (Batoebara et al., 2020). Jumlah persebaran berita hoax yang ditemukan oleh Kementerian Komunikasi dan Informatika (Kominfo) hingga awal tahun 2022 sebanyak 9546 berita yang telah tersebar di berbagai media sosial (Muttaqin et al., 2023).

Topik berita hoax yang banyak ditemukan di Indonesia biasanya merupakan topik yang sedang hangat diperbincangkan oleh masyarakat, salah satu contoh topik tersebut ialah konflik Palestina-Israel. Salah satu berita hoax yang muncul terkait konflik Palestina-Israel adalah pemberitaan mengenai perang Hamas dan Israel yang diunggah melalui platform media sosial Twitter dan terbukti jika video tersebut mengandung berita bohong. Faktanya, video tersebut merupakan video terjun payung pasukan Mesir di Akademi Militer Mesir di Kairo (Wulandari & Candra, 2024). Telah banyak portal berita yang memberitakan topik konflik ini, tetapi dibandingkan media di dalam negeri, portal media internasional lebih banyak menyoroti mengenai konflik ini. Salah satunya adalah British Broadcasting Corporation (BBC).

BBC merupakan salah satu portal media internasional yang memiliki lingkup ke seluruh negara di dunia karena jaringan wartawan BBC World tersebar di banyak negara. Namun, dalam portal media tersebut belum ada rubrik tersendiri dalam pendeteksian berita hoax atau non hoax terkait konflik Palestina dan Israel ini. Oleh karena banyaknya berita hoax mengenai konflik Palestina-Israel yang muncul di publik, maka diperlukan suatu pengklasifikasian kategori hoax dan non hoax untuk berita-berita yang beredar. Salah satu algoritma klasifikasi yang cocok dalam klasifikasi teks berita hoax dalam penelitian ini, yaitu *Naïve Bayes Classifier*.

Algoritma *Naïve Bayes* seringkali digunakan untuk mengklasifikasikan data dalam bentuk teks ulasan ataupun berita. *Multinomial Naïve Bayes* merupakan salah satu tipe dari algoritma *Naïve Bayes* yang efektif jika digunakan dalam klasifikasi teks. Prinsip kerja dari algoritma ini berkaitan dengan konsep *term frequency* yang didasarkan pada frekuensi kemunculan dari kata-kata dalam sebuah dokumen (Yuyun et al., 2021). Selain itu, *Multinomial Naïve Bayes* juga memiliki kelebihan dalam menghasilkan tingkat hasil akurasi yang cukup tinggi dalam hal klasifikasi teks mengenai komentar atau permasalahan.

Pada penelitian deteksi hoax menggunakan perbandingan *Naïve Bayes* dan *Support Vector Machine* (SVM) yang dilakukan oleh (Febriyanty et al., 2023), dataset diambil dari portal berita Indonesia, seperti Tribunnews, Detik, dan CNN dari bulan Januari 2020 hingga Juli 2023. Setelah dataset berhasil dikumpulkan oleh peneliti, hal yang dilakukan selanjutnya adalah melakukan pra proses data dan klasifikasi dengan menggunakan perbandingan metode SVM dan *Naïve Bayes*. Dari hasil penelitian tersebut dihasilkan akurasi dengan menggunakan metode SVM sebesar 75,5%, sedangkan untuk percobaan menggunakan *Naïve Bayes* sebesar 85%. Penelitian ini cukup membuktikan jika *Naïve Bayes* dapat menghasilkan akurasi yang lebih tinggi dalam klasifikasi teks.

Penelitian lainnya yang membahas mengenai perbandingan *Naïve Bayes* dengan algoritma K-NN terkait analisis sentimen vaksinasi Astra Zeneca pada Twitter. Penelitian tersebut menggunakan data sebanyak 407 data yang dibagi menjadi sentimen positif, netral, dan negatif. Hasil akurasi dalam penelitian tersebut menunjukkan jika percobaan dengan K-NN menghasilkan akurasi 74,78 dan

dengan *Naïve Bayes* sebesar 88,56% (Jtik et al., 2022). Dengan kedua penelitian tersebut menunjukkan jika *Naïve Bayes* unggul dalam klasifikasi teks.

Dalam literatur lain menyebutkan jika *Naïve Bayes* masih memiliki kelemahan dalam hasil probabilitasnya yang kurang optimal (Sukron et al., 2022). Untuk mengatasi hal tersebut, beberapa penelitian memanfaatkan metode optimasi, salah satunya adalah *Particle Swarm Optimization* (PSO). PSO memiliki kelebihan dalam hal parameter dan konvergensinya yang sederhana, serta mudah diimplementasikan. Dalam beberapa penelitian sebelumnya, PSO telah diterapkan menjadi metode optimasi, salah satunya pada penelitian yang dilakukan oleh (Syafira et al., 2021) mengenai analisis kecacatan dalam produk menggunakan algoritma *Naïve Bayes* dan *Particle Swarm Optimization*. Terdapat 800 data yang digunakan dalam penelitian analisis kecacatan ini. Penelitian tersebut menghasilkan tingkat akurasi pada persentase 84,38% untuk percobaan menggunakan *Naïve Bayes* dan 88,62% untuk percobaan yang ditambah dengan optimasi dari *Particle Swarm Optimization*. Penelitian ini membuktikan jika PSO membantu algoritma *Naïve Bayes* dalam hal menaikkan tingkat hasil akurasi.

Penelitian terdahulu lainnya membahas mengenai klasifikasi adanya berita hoax dengan menggunakan metode *Naïve Bayes* ditambah dengan optimasi PSO. Penelitian yang dilakukan oleh Risa ini menggunakan dataset sebanyak 150 data dengan 75 data hoax dan 75 lainnya non hoax. Dataset diambil dari situs turnbackhoax.com lengkap dengan pelabelan secara otomatis yang telah dilakukan situs tersebut. Dari penelitian tersebut, tingkat akurasi yang dihasilkan sebesar 85,19% untuk percobaan menggunakan algoritma *Naïve Bayes* dan optimasi dari PSO, sedangkan percobaan dengan menggunakan *Naïve Bayes* saja hanya sebesar 74,67% (Wati, 2020).

Adapun hasil tingkat akurasi dari suatu metode juga didasarkan atas pembagian data latih dan data uji yang digunakan di dalam penelitian. Penelitian sebelumnya dilakukan oleh (Musu et al., 2021) mengenai pengaruh komposisi pembagian data latih dan data uji menggunakan algoritma C4.5 menggunakan empat buah dataset, yaitu gender, titanic, deals, dan iris. Dari keempat dataset tersebut ditemukan hasil jika tingkat akurasi tertinggi rata-rata terdapat pada pembagian data latih sebesar 90% dan data uji sebesar 10%.

Dari pemaparan latar belakang tersebut, maka pada penelitian skripsi ini penulis berencana untuk melakukan penelitian mengenai Klasifikasi Berita Hoax Konflik Palestina-Israel Menggunakan *Multinomial Naïve Bayes* dengan Optimasi *Particle Swarm Optimization*. Penelitian ini difokuskan pada topik mengenai konflik Palestina-Israel dengan memanfaatkan teknik *web scraping* dari website berita *online* internasional BBC World yang kemudian akan dikategorikan menjadi berita hoax dan non hoax. Algoritma yang digunakan ialah algoritma *Multinomial Naïve Bayes* dan *Particle Swarm Optimization*, serta terdapat beberapa pengujian untuk masing-masing pembagian data latih serta data uji agar memperoleh hasil tingkat akurasi tertinggi. Diharapkan penelitian ini dapat memberikan kontribusi yang signifikan bagi masyarakat dalam melakukan klasifikasi berita antara hoax dan non hoax, khususnya terkait topik berita konflik Palestina-Israel

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah yang dapat diambil, yaitu:

1. Bagaimana implementasi algoritma *Multinomial Naïve Bayes* dengan metode optimasi *Particle Swarm Optimization* dalam mengklasifikasikan berita hoax dan non hoax terkait konflik Palestina-Israel dalam media *online*?
2. Bagaimana pengaruh *Particle Swarm Optimization* pada hasil akurasi algoritma *Multinomial Naïve Bayes* pada pengklasifikasian berita hoax terkait konflik Palestina-Israel pada media *online*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka terdapat beberapa tujuan dari penelitian ini, yaitu:

1. Mengetahui implementasi algoritma *Multinomial Naïve Bayes* dengan metode optimasi *Particle Swarm Optimization* dalam mengklasifikasikan berita hoax dan non hoax terkait konflik Palestina-Israel dalam media *online*.

2. Mengetahui pengaruh *Particle Swarm Optimization* pada hasil akurasi algoritma *Multinomial Naïve Bayes* pada pengklasifikasian berita hoax terkait konflik Palestina-Israel pada media *online*?

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian ini, yaitu:

1. Dapat memberikan pemahaman lebih terkait implementasi dari algoritma *Multinomial Naïve Bayes* dengan atau tanpa metode optimasi *Particle Swarm Optimization*.
2. Memudahkan masyarakat dalam mengidentifikasi suatu berita, terutama berita konflik Palestina-Israel ke dalam kategori berita hoax dan berita yang faktanya telah kredibel.
3. Penelitian ini dapat dijadikan sebagai referensi acuan untuk penelitian selanjutnya yang memiliki topik serupa, yaitu klasifikasi hoax dengan menggunakan algoritma *Multinomial Naïve Bayes* dan optimasi *Particle Swarm Optimization*.

1.5 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini, yaitu:

1. Pengambilan data menggunakan teknik *web scraping* dari portal media *online* internasional BBC World. Kata kunci yang diambil ialah ‘gaza’, ‘israel’, dan ‘hamas’ dalam rentang waktu bulan Oktober 2023 – Januari 2024.
2. Penelitian ini dibagi menjadi dua kategori, yaitu berita hoax dan non hoax.
3. Penelitian ini berfokus dalam penilaian hasil akurasi, *precision*, *recall*, *f1-Score* terhadap klasifikasi berita hoax terkait konflik Palestina-Israel menggunakan *Multinomial Naïve Bayes* dan optimasi dari *Particle Swarm Optimization*.
4. Penelitian ini diimplementasikan menggunakan Python sebagai bahasa pemrogramannya.
5. Seluruh judul berita *online* yang hendak diklasifikasikan dalam bentuk bahasa Inggris.