

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Hoaks merupakan informasi yang menyesatkan dan dapat berbahaya bagi setiap orang yang membaca dan menyebarkannya. Berita hoaks menjadi berbahaya karena dianggap mampu menyampaikan informasi yang salah atau palsu sebagai kebenaran (Rasywir dan Purwarianti, 2015). Indikator awal dari hoaks dapat berupa kesalah pahaman, manifestasi dari ketakutan, harapan, dan bias seseorang atau sekelompok orang (Silverman, 2015). Pada zaman dahulu sebelum adanya internet, hoaks jauh lebih berbahaya karena lebih sulit untuk diidentifikasi (Aditiawarman dkk., 2019). Setelah adanya internet, hoaks lebih mudah untuk divalidasi kebenarannya.

Salah satu cara untuk mengklasifikasikan berita hoaks dapat dilakukan dengan bantuan pembelajaran mesin. Beberapa model pembelajaran mesin diketahui dapat menghasilkan performa yang sangat baik dalam melakukan klasifikasi teks, contohnya model dengan menggunakan algoritma Naïve Bayes yang terbukti mampu menghasilkan akurasi sebesar 82.6% berdasarkan hasil uji coba secara statis (Rahutomo dkk., 2019). Berdasarkan beberapa penelitian terdahulu, klasifikasi teks menggunakan algoritma Naïve Bayes secara umum memiliki hasil yang lebih baik dibanding metode lainnya. Hal ini dapat dibuktikan dari adanya penelitian terdahulu mengenai klasifikasi teks yang menggunakan Naïve Bayes dengan hasil akurasi yang lebih baik dibandingkan SVM dan C4.5 dengan akurasi 91.35% (Rasywir dan Purwarianti, 2015). Naïve Bayes *classifier* pada klasifikasi data dari halaman web

menghasilkan hasil akurasi 95.20% dan *f-measure* 97.26% sedangkan Decision Tree *classifier* menghasilkan nilai akurasi 94.85% dan *f-measure* 97.09% (Xhemali dkk., 2009). Selain itu, penelitian terdahulu mengenai klasifikasi berita hoaks berbahasa Indonesia dengan algoritma Naïve Bayes menghasilkan *f1-score* sebesar 0.93 dan hasil *f1-score* 0.92 untuk Multinomial Naïve Bayes (Athailah dkk., 2020).

Hingga saat ini, terdapat lima jenis model Naïve Bayes yang dapat diimplementasikan, antara lain Gaussian Naïve Bayes, Multinomial Naïve Bayes, Complement Naïve Bayes, Bernoulli Naïve Bayes, dan Categorical Naïve Bayes. Pada penelitian klasifikasi teks dengan metode ekstraksi fitur TF-IDF yang datanya kontinu, model Naïve Bayes yang dapat digunakan yaitu Gaussian Naïve Bayes, Multinomial Naïve Bayes, dan Complement Naïve Bayes. Meski terbukti dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan algoritma lainnya, pembelajaran mesin terus berkembang. Pembelajaran mesin dibuat secara *ensemble* dengan beberapa teknik dengan tujuan meningkatkan performa model dasar.

Dalam pembuatan *ensemble learning*, terdapat beberapa teknik. Diantaranya yaitu *bagging*, *boosting*, dan *stacking*. Pemilihan model dasar harus koheren dengan teknik yang digunakan. Model dasar dengan bias rendah namun varians tinggi seharusnya menggunakan teknik yang cenderung mengurangi varians. Sebaliknya jika menggunakan model dasar dengan varians rendah tetapi bias tinggi, menggunakan teknik yang cenderung mengurangi bias (Rocca, 2019). Menurut Paul (2019), teknik *bagging* digunakan untuk mengurangi varians, *boosting* untuk mengurangi bias, sementara *stacking* biasa digunakan untuk meningkatkan performa prediksi model. *Ensemble stacking* menggunakan algoritma *meta learning* untuk mempelajari cara terbaik menggabungkan prediksi dari beberapa

algoritma pembelajaran mesin (Brownlee, 2020). Penggunaan teknik *stacking* dalam *ensemble learning* memanfaatkan kemampuan berbagai model yang menghasilkan performa baik pada klasifikasi atau regresi dan dapat membuat prediksi yang memiliki kinerja lebih baik dibandingkan dengan model tunggal apapun dalam *ensemble* tersebut (Brownlee, 2020). Melanjutkan dari penelitian terdahulu mengenai klasifikasi berita hoaks berbahasa Indonesia menggunakan model Naïve Bayes yang menghasilkan performa baik, dibuat penelitian mengenai klasifikasi berita hoaks dengan *ensemble learning* menggunakan teknik *stacking* beberapa model Naïve Bayes sebagai model dasar atau *base learners*.

Pada *ensemble learning* menggunakan teknik *stacking*, *meta learner* menerima hasil prediksi dari *base learners* sebagai input data latih. *Meta learner* biasanya sederhana dan memberikan interpretasi dari prediksi yang dibuat oleh model dasar (Brownlee, 2020). Maka, model linier seringkali digunakan sebagai *meta learner*. Model linier yang sering digunakan untuk klasifikasi yaitu Logistic Regression. Algoritma Logistic Regression dapat digunakan untuk menemukan hubungan antara fitur dan probabilitas hasil tertentu (Agrawal, 2017).

Berdasarkan data yang diperoleh dan hasil analisa penelitian terdahulu, maka penelitian ini mengimplementasikan model Naïve Bayes yang dinyatakan sebagai model yang lebih baik dibandingkan model dengan algoritma lainnya untuk menyelesaikan permasalahan klasifikasi teks termasuk berita hoaks. Tiga jenis model Naïve Bayes diimplementasikan dalam *ensemble learning* menggunakan teknik *stacking* sebagai *base learners* dan Logistic Regression sebagai *meta learner*. Penelitian klasifikasi berita hoaks dengan *ensemble learning* ini dibuat dengan harapan dapat menghasilkan performa yang lebih baik dari model dasar

yang diketahui menghasilkan performa baik dalam mengklasifikasikan berita hoaks.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, rumusan masalah yang akan dibahas dalam penelitian adalah sebagai berikut.

1. Bagaimana cara mengklasifikasikan berita hoaks dengan *ensemble model* menggunakan teknik *stacking* dan *multiple Naïve Bayes model*?
2. Bagaimana *f1-score* dari klasifikasi berita hoaks dengan *ensemble model* menggunakan teknik *stacking* dan *multiple Naïve Bayes model*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut.

1. *Dataset* berita hoaks berbahasa Indonesia menggunakan *dataset* dari Mendeley dengan judul “Indonesian Hoax News Detection Dataset” (Rahutomo dkk., 2019).
2. *Dataset* terdiri dari 228 kelas berita hoaks dan 372 kelas berita valid. Sumber *dataset* dari Mendeley (Rahutomo dkk., 2019).
3. Jenis Naïve Bayes yang diimplementasikan adalah Gaussian Naïve Bayes, Multinomial Naïve Bayes, dan Complement Naïve Bayes (Scikit-learn, 2019).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, tujuan penelitian dapat dirumuskan sebagai berikut.

1. Mengetahui bagaimana cara mengklasifikasikan berita hoaks dengan *ensemble model* menggunakan teknik *stacking* dan *multiple Naïve Bayes model*.
2. Mengetahui *f1-score* dari *ensemble model* menggunakan teknik *stacking* dan *multiple Naïve Bayes model* untuk mengklasifikasikan berita hoaks.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut.

1. Memberikan kontribusi ilmiah mengenai cara melakukan klasifikasi berita hoaks dengan *ensemble model* menggunakan teknik *stacking* dan *multiple Naïve Bayes model*.
2. Penelitian ini diharapkan dapat memberi inspirasi bagi pembaca untuk melakukan penelitian terhadap klasifikasi berita hoaks berbahasa Indonesia.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi dijabarkan sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, batasan masalah, perumusan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan skripsi.

BAB II LANDASAN TEORI

Bab ini berisi teori-teori yang mendasari penelitian. Teori yang dibahas berasal literatur seperti buku dan jurnal serta konsep secara tertulis yang berkaitan dengan permasalahan yang dikemukakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metodologi penelitian dan perancangan sistem.

BAB IV HASIL DAN DISKUSI

Bab ini berisi pembahasan mengenai implementasi dari sistem yang dibuat serta uji coba dari skenario uji.

BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan hasil penelitian dan uji coba yang telah dilakukan beserta saran untuk penelitian selanjutnya.