

BAB I

PENDAHULUAN

1.1 Latar Belakang

Informasi merupakan sekumpulan data yang sudah diproses dan dikelola sehingga dapat dipahami dan bermanfaat bagi penerimanya. Informasi juga merupakan sarana komunikasi antara sesama manusia. Informasi adalah hasil pengolahan data yang memberikan arti dan manfaat tertentu bagi orang yang menerimanya (Susanto, 2017). Pada jaman dahulu penyebaran informasi sangat dibatasi dengan jarak dan waktu karena teknologi pada saat itu masih terbatas. Di era teknologi modern saat ini kendala jarak dan waktu dalam penyebaran informasi sudah teratasi. Apalagi pada era teknologi modern saat ini informasi dapat disebarkan dengan mudah melalui berbagai macam media. Salah satunya yang paling populer adalah media sosial. Media sosial sendiri merupakan wadah atau sarana yang dapat menampung aspirasi dari setiap manusia di berbagai belahan dunia. Media sosial juga merupakan media komunikasi yang juga berperan sebagai media penyebaran informasi. Informasi yang dikirimkan melalui media sosial akan sangat cepat diterima dan dikonsumsi oleh setiap orang.

Berbagi informasi bagi setiap orang merupakan hal yang diperlukan, namun tidak semua informasi yang disebarkan melalui media benar adanya. Telah banyak terjadi kasus penyebaran berita yang tidak benar atau biasa disebut dengan berita *hoax*. Berita *hoax* sendiri merupakan informasi berbahaya yang menyesatkan persepsi manusia dengan menyebarkan informasi yang salah namun dianggap sebagai kebenaran (Rasywir & Purwarianti, 2015). Penyebaran berita *hoax* bisa disebabkan oleh individu atau kelompok organisasi yang memiliki tujuan tertentu yang kemudian berita tersebut disebarkan kepada masyarakat luas. Penyebaran berita *hoax* memiliki dampak buruk bagi masyarakat. Telah dilansir oleh situs KEMENKOMINFO (Kementerian Komunikasi & Informatika) pada 13 Desember 2017 yang menyebutkan bahwa ada 800.000 situs di Indonesia yang telah terindikasi sebagai penyebar informasi palsu. Disebutkan juga internet telah salah dimanfaatkan oleh oknum tertentu untuk keuntungan pribadi dan kelompoknya

dengan cara menyebarkan konten-konten negatif yang menimbulkan keresahan di masyarakat.

Begitu maraknya penyebaran berita *hoax* yang telah terjadi pada media pemberitaan terutama media sosial, maka banyak ide gagasan untuk melakukan pencegahan dalam penyebaran berita *hoax*. Telah banyak muncul saran dan tips untuk menghindari berita-berita *hoax* yang banyak tersebar di dalam media sosial. Adapun sebagian dari media sosial yang memberikan layanan tambahan untuk melaporkan konten yang mengandung unsur berita yang tidak benar atau *hoax*. Telah dikembangkan juga beberapa sistem penangkal yang memprediksi konten berita *hoax* dengan berbagai macam metode yang digunakan dalam pengimplementasiannya dengan hasil akurasi yang bermacam-macam. Untuk mengetahui informasi yang tersebar masuk ke dalam berita *hoax* atau benar akan dilakukan proses klasifikasi. Dari hasil klasifikasi akan didapatkan nilai persentase akurasi. Pada penelitian Rahman, Wiranto & Doewes (2017) yang berjudul “*Online News Classification Using Multinomial Naïve Bayes*” membandingkan algoritma *Multinomial Naïve Bayes* dengan seleksi fitur TF-IDF dan *Document Frequency Thresholding* dalam kasus *text mining* klasifikasi dan didapatkan hasil akurasi terbaik yaitu algoritma *Multinomial Naïve Bayes* menggunakan seleksi fitur TF-IDF dengan akurasi sebesar 86,62% sedangkan pada seleksi fitur *Document Frequency Thresholding* mendapatkan akurasi sebesar 85,98%. Sedangkan pada penelitian Wisaksono & Mujiyatna (2017) yang berjudul “Klasifikasi Berita Berkategori Olahraga dengan Algoritma *Multivariate Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes*” membandingkan antara algoritma *Multinomial Naïve Bayes (MNB)* dan *Multivariate Bernoulli* dalam pengklasifikasian berita olahraga dan didapatkan hasil akurasi yaitu algoritma *Multivariate Bernoulli* dengan akurasi sebesar 97,1% sedangkan *Multinomial Naïve Bayes (MNB)* mendapatkan akurasi sebesar 94%. Dan pada penelitian Pantouw (2017) yang berjudul “Perbandingan Klasifikasi *Rocchio* dan *Multinomial Naïve Bayes* pada Analisis Sentimen Data *Twitter* Bahasa Indonesia” membandingkan antara algoritma *Rocchio* dan *Multinomial Naïve Bayes (MNB)* dalam pengklasifikasian sentiment data *twitter* dan didapatkan hasil akurasi terbaik yaitu algoritma *Rocchio* dengan akurasi sebesar 96,283% sedangkan

Multinomial Naïve Bayes (MNB) mendapatkan akurasi sebesar 85,399%. Dari beberapa sumber penelitian yang didapat, maka dari itu akan dibandingkan antara akurasi terbaik dari referensi diatas sehingga didapatkan ketiga algoritma yang akan dibandingkan yaitu algoritma yaitu *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* untuk menemukan akurasi yang terbaik di antara metode tersebut pada kasus konten berita *hoax* berbahasa Indonesia pada media sosial.

Berdasarkan latar belakang yang telah dijelaskan tentang banyaknya penyebaran berita *hoax* pada media sosial, maka penulis akan membuat sebuah penelitian dengan judul “Perbandingan Kinerja Algoritma *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* dalam Klasifikasi Konten Berita *Hoax* Berbahasa Indonesia pada Media Sosial”.

1.2 Rumusan Masalah

Dari latar belakang yang sudah dijelaskan, maka diambil perumusan masalah dalam penelitian yaitu:

1. Berapa persentase tingkat akurasi yang didapatkan dari hasil perhitungan *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* dalam mengklasifikasikan kasus berita *hoax* berbahasa Indonesia?
2. Berapa persentase tingkat presisi yang didapatkan dari hasil perhitungan *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* dalam mengklasifikasikan kasus berita *hoax* berbahasa Indonesia?
3. Berapa persentase tingkat *recall* yang didapatkan dari hasil perhitungan *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* dalam mengklasifikasikan kasus berita *hoax* berbahasa Indonesia?

1.3 Batasan Masalah

Agar permasalahan dalam penelitian yang dibahas tidak terlalu luas, maka diberikan batasan masalah sebagai berikut:

1. Data kasus yang digunakan dalam penelitian diambil dari situs *online turnbackhoax.id* yang telah memverifikasi berita hoax dari media sosial sejak 2015 sampai dengan saat ini.
2. Data yang diambil secara manual dari situs *turnbackhoax.id* adalah berupa narasi dari berbagai media sosial berbahasa Indonesia yang telah dikonfirmasi kebenarannya oleh tim *turnbackhoax.id* yang terdiri dari Muhammad Khairil Haesy M.Hum, Dedy Helsyanto S.IP, Bentang Febrylian S.IKOM dan Syarief Ramaputra S.IKOM.
3. Penelitian yang dilakukan ini hanya pada berita berbahasa Indonesia saja.
4. Dataset yang digunakan berjumlah 200 data yang terdiri dari 110 berita *hoax* dan 90 berita benar dari 12 Desember 2019 sampai 20 Januari 2020. Data yang diambil ada 200 karena pada situs *turnbackhoax.id* diambil dari berbagai macam media sosial bisa berupa gambar, teks narasi, video, website, dan lain sebagainya. Pada penelitian ini berkaitan dengan *text mining*, maka peneliti hanya menggunakan konten berita yang menggunakan teks narasi saja.
5. Penelitian ini dilakukan dengan menggunakan Bahasa Pemrograman *Python* dengan menggunakan *tool jupyter notebook*.
6. Hasil dari penelitian berupa pengklasifikasian berita benar atau *hoax* dengan akurasi dari masing-masing algoritma.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui tingkat presisi dari klasifikasi berita *hoax* berbahasa Indonesia dengan algoritma *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm*.

2. Untuk mengetahui tingkat akurasi dari klasifikasi berita *hoax* berbahasa Indonesia dengan algoritma *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm*.
3. Untuk mengetahui tingkat *recall* dari klasifikasi berita *hoax* berbahasa Indonesia dengan algoritma *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm*.

1.5 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian adalah:

1. Dapat memberikan informasi tentang cara kerja algoritma *Multinomial Naïve Bayess (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* dalam mengklasifikasikan berita *hoax* berbahasa Indonesia.
2. Dapat memberikan informasi yang dapat digunakan sebagai acuan serta referensi yang dapat dikembangkan untuk penelitian lebih lanjut.

