

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Perkembangan teknologi saat ini semakin maju pesat, khususnya pada penggunaan media sosial. Media sosial dapat memberikan kontribusi secara tidak langsung, baik kontribusi berdampak positif maupun negatif. Namun, penggunaan media sosial kini terganggu sering dengan panasnya isu-isu palsu baik terkait resesi perekonomian diseluruh dunia, isu palsu terkait politik, maupun isu palsu tentang kesehatan dan banyak lainnya, yang disebabkan oleh beredarnya informasi *hoax* yang belum dikonfirmasi kebenarannya. Berita palsu (*hoax*) memiliki banyak definisi seperti solusi masalah yang digunakan sebagai senjata politik dan publik, berdasarkan kebenaran yang tidak relevan (pasca-kebenaran). *Hoax* dalam bentuk teks merupakan bentuk dari kumpulan kata yang belum teridentifikasi kebenarannya (Ahmed dkk., 2014), salah satu contohnya seperti informasi teks *hoax* yang ada khususnya pada penyebaran informasi *hoax* pada penggunaan media sosial *Twitter*.

*Twitter* merupakan jejaring media sosial (*microblog*) yang mana memungkinkan penggunanya untuk membaca dan mengirim pesan berbasis teks hingga 140 karakter yang sering disebut dengan kicauan (*tweet*). *Twitter* memiliki fitur yang memungkinkan para pengguna dapat berkomunikasi secara *realtime*. Waktu nyata yang ditawarkan aplikasi media sosial *Twitter* mengakibatkan peredaran informasi *hoax* berkembang dan meluas dengan sangat cepat. Penyebaran *hoax* membuat banyak kalangan pengguna aplikasi *Twitter* menjadi kesulitan mendapatkan informasi aktual dalam kehidupan sehari-hari (Nur, 2015). Hal ini menjadi masalah serius yang patut dicari solusinya.

Sistem pendeteksi *hoax* yang ada saat ini dan berkembang di Indonesia adalah salah satunya aplikasi *Hoax Buster Tools* yang merupakan aplikasi filter dan pendeteksi untuk informasi *hoax*. Aplikasi *Hoax Buster Tools* dikelola oleh CSO (*Civil Society Organization*) yang di bawah naungan MAFINDO, didirikan pada tanggal 19 November 2016. Aplikasi tersebut berfokus pada penyebaran

*hoax* pada media sosial dan situs-situs di internet. Sistem yang berjalan pada aplikasi *Hoax Buster Tools* adalah sistem pencarian (*search engine*), informasi yang ditampilkan sesuai dengan *keyword* yang pengguna ketikkan dan langsung merekomendasikan situs terkait sesuai dengan *keyword* yang pengguna cari (Djalante dkk., 2020). Namun, di antara kelebihan dari aplikasi *Hoax Buster Tools* ternyata masih terdapat kekurangan dimana belum mampu mendeteksi informasi *hoax* secara singkat pada waktu nyata (*real time*) pada satu *platform* khusus, oleh karena itu penelitian ini berfokus pada membangun sistem pendeteksi informasi *hoax* pada aplikasi *Twitter*.

Berdasarkan penjelasan sebelumnya, informasi *hoax* dalam bentuk teks memiliki kecenderungan tumbuh dengan cepat dan kompleks setiap waktunya. Kumpulan informasi dalam bentuk teks, dapat menangkap penggunaan bahasa dalam bentuk tertulis atau lisan (*Corpus*) secara *real time*, ditambah lagi minimalitas pengetahuan dari belum adanya kata karakteristik *hoax* pada sebuah berita. Hal ini yang menyebabkan pengguna sosial media kesulitan membedakan antara informasi yang benar atau *hoax*.

Penelitian yang berkaitan dengan proses penyaringan spam atau berita *hoax* dilakukan oleh Ahmed dkk. (2014) menjelaskan mengenai proses menggabungkan dua algoritma dalam proses penyaringan spam dengan menggunakan *dataset* berdasarkan pada data yang ada pada pesan SMS. Hasil dari proses menggabungkan algoritma Apriori dan algoritma Naïve Bayes, terdapat peningkatan akurasi sistem dalam mengklasifikasi data pesan. Penerapan algoritma Apriori dalam memproses data memiliki kelebihan, yang mana algoritma tersebut cocok untuk menanggulangi permasalahan dari banyaknya transaksi *dataset*. Pada proses klasifikasi data pesan, untuk setiap katanya dianggap sebagai data dari transaksi, ini menjadikan volume dari data sangat besar, serta mampu menambang *itemset* pada banyak *database* yang digunakan sehingga mampu memberikan rekomendasi sebuah kata kunci sebagai parameter utama dari banyaknya *database* (Ahmed dkk., 2014).

Akhir-akhir ini para peneliti dalam komunitas ilmuwan, menemukan cara cepat mengklasifikasikan *dataset*, dalam merespon tren tersebut para ilmuwan memastikan bahwa metode *Machine Learning* mampu mempercepat proses

klasifikasi dari banyaknya *dataset*, selain itu kelebihan lainnya dari metode *Machine Learning* yaitu menjadi metode penunjang yang mampu membangun sistem dengan tujuan untuk sebuah *problem case* pendeteksi (Zhang & Ghorbani, 2020). Algoritma *Random Forest* merupakan metode *machine learning averaging ensemble*, yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi *dataset*. *Dataset* yang digunakan dalam penelitian ini, menggunakan *dataset* dalam bentuk data teks informasi, karena fokus utama penelitian ini untuk mengklasifikasikan data teks pada saat proses klasifikasi (Amir dkk., 2019).

Penerapan algoritma *Random Forest* dalam membangun sistem pendeteksi spam pada pesan SMS pun dilakukan oleh Sjarif dkk. (2019), dengan menerapkan algoritma *Random Forest* dan metode *Term Frequency-Inverse Document Frequency*. Pada penelitiannya menyajikan sistem yang memiliki kinerja yang lebih baik dibandingkan algoritma lainnya dalam hal presentasi akurasi, yang mana algoritma *Random Forest* mampu memberikan presisi dan *f-measure* yang baik dengan tingkat presisi 98% dan dapat mengekstrak dari bentuk *decision tree* menjadi sebuah keputusan yang terlihat jelas dengan kata lain mampu mengklasifikasi dengan cepat dan mampu menghasilkan solusi keputusan yang paling optimal (Sjarif dkk., 2019).

Kompleksitas dari data yang digunakan yaitu data bentuk informasi teks *hoax* berserta dengan kelebihan dari algoritma Apriori yaitu algoritma yang salah satu sub prosesnya adalah menerapkan *frequent itemset*, dimana algoritma ini mampu menambang data *itemset* yang sering muncul dari banyaknya *database* yang digunakan. Selanjutnya, yaitu penjelasan dari Algoritma *Random Forest* yang mana memiliki kelebihan dalam hal presentasi akurasi dengan tingkat presisi 98% dan algoritma *Random Forest* mampu mengekstrak dari bentuk *decision tree* menjadi sebuah keputusan yang terlihat jelas dengan kata lain mampu mengklasifikasi dengan cepat dan mampu menghasilkan solusi keputusan yang paling optimal, mengingat terlalu banyak informasi yang beredar di sosial media *Twitter*.

Berdasarkan pertimbangan dan analisis di atas, maka rumusan masalah pada penelitian ini adalah mengembangkan algoritma Apriori dan algoritma *Random Forest* dalam membangun sistem pendeteksi informasi *hoax* pada

aplikasi *Twitter*. Mengintegrasikan algoritma Apriori pada proses *training data*, dapat menyelesaikan permasalahan banyaknya *dataset* pada teks informasi *hoax* (*tweet*) dalam menambang *itemset* yang mana hasil dari proses algoritma Apriori dijadikan parameter utama untuk proses klasifikasi, dan proses algoritma Apriori mampu menghasilkan data lebih terstruktur dan mudah dipahami untuk program bekerja. Sehingga pada proses *testing data* dilakukan, algoritma *Random Forest* mampu membangun dan mengklasifikasikan beberapa model *Random Forest* dengan cepat dan lebih optimal. Model *Random Forest* yang terpilih nantinya dapat membentuk model sistem yang paling optimal berdasarkan performa sistem yang dibangun, dalam mendeteksi informasi *hoax* pada aplikasi *Twitter*.

## **1.2. Tujuan**

Tujuan dari penelitian ini adalah mengembangkan algoritma Apriori dan algoritma *Random Forest* untuk membangun sistem pendeteksi informasi *hoax* pada aplikasi *Twitter*. Penggunaan dari dua algoritma pada penelitian ini bertujuan untuk mengintegrasikan algoritma Apriori dengan algoritma *Random Forest* pada saat memproses *dataset*. Kedua metodologi ini, dapat memudahkan peneliti menganalisis dan mengevaluasi hasil dari sistem yang dibangun dalam mendeteksi informasi *hoax* dengan tingkat akurasi yang paling optimal.

## **1.3. Manfaat Penelitian**

Manfaat dari mengintegrasikan algoritma Apriori dan algoritma *Random Forest* pada saat memproses *dataset* mampu meminimalisir kegagalan dalam proses mengolah data dan mampu mengklasifikasi informasi *hoax* dengan tingkat akurasi yang optimal. Selain itu, manfaat lainnya adalah dapat mengatasi segala permasalahan algoritma dengan logika yang lebih terstruktur. Hasil implementasi dari sistem yang dibangun, dapat dimanfaatkan menjadi sarana publik khususnya pada pengguna aplikasi *Twitter* untuk meminimalisir penyebaran informasi *hoax* di Indonesia.