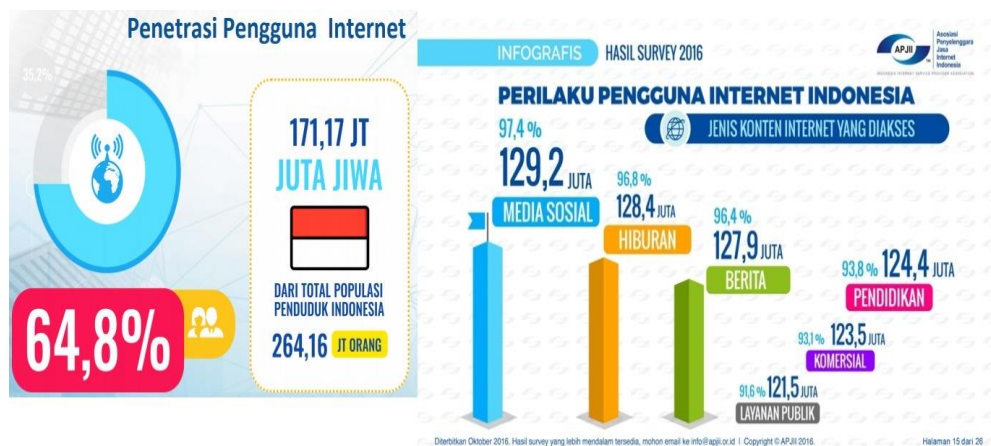


# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Perkembangan teknologi ke arah digital saat ini semakin pesat, termasuk penggunaan internet yang bisa diakses dimana saja dan kapan saja [1]. Berdasarkan buletin Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) edisi 40 Mei 2019, jumlah pengguna internet di Indonesia mengalami peningkatan setiap tahunnya dari 2016 hingga 2019 dari survei yang telah dilakukan. Data survei mencatat, jumlah pengguna internet bertambah 10,12% pada tahun 2018 dibandingkan dengan tahun sebelumnya. Dan secara total, pengguna internet yang tercatat mencapai 171,17 juta pengguna dari populasi 264,16 juta jiwa pada tahun 2018. Berdasarkan data survei yang dilakukan pada tahun 2016, ada 3 besar jenis konten internet yang sering diakses oleh pengguna, yaitu sosial media di urutan pertama (97,4%), hiburan di urutan kedua (96,8%) dan berita di urutan ketiga (96,4%) [2].



**Gambar 1.1.** Infografis Jumlah Pengguna dan Jenis Konten yang diakses [2]

Berita merupakan sebuah sumber informasi terhadap suatu kejadian yang ada dalam lingkungan sosial masyarakat. Berita berisi laporan mengenai peristiwa terbaru dan penting untuk disampaikan ke masyarakat yang dikemas dalam bentuk surat kabar, siaran televisi, maupun media *online* [3]. Keberadaan berita yang beredar juga memiliki ancaman sebagai sebuah berita

yang bohong dan meresahkan masyarakat atau sering disebut hoaks [4]. Perlu diingat juga bahwa penyebaran berita atau informasi palsu terdapat konsekuensi hukum seperti yang diatur dalam UU ITE Nomor 11 Tahun 2008 [5]. Sebagai bentuk ketegasan pemerintah dalam upaya memerangi dan mengurangi penyebaran hoaks ini adalah dengan memblokir situs internet yang dinilai sebagai sumber informasi palsu [6].

Penerapan *machine learning* dengan sebuah pemodelan algoritma juga dapat digunakan untuk menentukan atau mengklasifikasikan sebuah berita tersebut benar atau hoaks. Klasifikasi berita hoaks merupakan salah satu aplikasi atau implementasi dari klasifikasi teks [7]. Klasifikasi teks adalah sebuah sistem yang menggunakan pendekatan *machine learning* yang terdiri dari *preprocessing*, ekstraksi fitur dan klasifikasi menggunakan pemodelan sebuah algoritma seperti *Support Vector Machine (SVM)* atau *Recurrent Neural Network (RNN)* [8].

Dalam *machine learning*, *Support Vector Machine (SVM)* adalah model pembelajaran yang menganalisis data yang digunakan untuk analisis klasifikasi dan regresi. *SVM* dapat membuat sebuah model yang akan digunakan untuk melakukan klasifikasi maupun prediksi [9]. Sedangkan, algoritma *Neural Network* adalah model yang terinspirasi oleh bagaimana *neuron* dalam otak manusia bekerja. Tiap *neuron* yang ada pada otak manusia saling berhubungan satu sama lain dan informasi mengalir dari setiap *neuron* tersebut. *Neural Network* merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi yang mengalir melalui jaringan tersebut dan dapat digunakan untuk memodelkan hubungan yang kompleks untuk menemukan pola-pola dalam pengklasifikasian dalam data [10]. *Recurrent Neural Network (RNN)* adalah bagian dari *Neural Network* yang lebih difokuskan untuk memproses data yang berupa *sequence* seperti kata atau kalimat [11]. Pada *RNN* terdapat memori atau ingatan (*feedback loop*) yang dapat mengenali informasi dari masa lalu dalam proses pembelajaran modelnya dan pemrosesannya berulang kali untuk memproses *input* data sekuensial [12].

Penelitian mengenai klasifikasi berita hoaks dilakukan oleh Dina Maulina dan Rofie Sagara dengan menggunakan algoritma klasifikasi *Support Vector Machine (SVM)* yang dikombinasikan dengan pembobotan kata dengan metode *Term Frequency – Inverse Document Frequency (TF-IDF)*. Dan hasilnya akurasi dari *SVM Linear* adalah 95.83% dengan data 108 artikel hoaks dan 132 artikel tidak hoaks dari pengukuran yang dilakukan dengan metode *10-Fold Cross Validation*, *Support Vector* yang terbentuk dari pemodelan *SVM Linear* adalah *14 Support Vector*, serta memiliki kecepatan dalam training data dengan waktu 1.37 detik untuk 240 vektor data yang terbentuk [9].

Penelitian lainnya untuk klasifikasi berita hoaks dilakukan oleh Aini Hanifa dan lainnya, pada penelitian tersebut dilakukan perbandingan metode algoritma *Recurrent Neural Network (RNN)* antara *Long Short Term Memory (LSTM)* dan *Gated Recurrent Unit (GRU)*. Nilai akurasi yang dihasilkan oleh model yaitu 73% untuk metode *LSTM* dan 64% dengan menggunakan metode *GRU* [12].

Berdasarkan referensi penelitian yang telah dilakukan, baik algoritma *SVM* dan *RNN* memiliki akurasi yang baik dan cocok untuk melakukan pengklasifikasian. Berdasarkan permasalahan yang telah dijelaskan, penelitian ini akan melakukan komparasi antara kedua algoritma tersebut yaitu *SVM* dan *RNN* untuk mengklasifikasikan data berita atau informasi hoaks, dikarenakan belum adanya penelitian yang membandingkan algoritma *SVM* dan *RNN* untuk klasifikasi data berita atau informasi hoaks.

## 1.2. Rumusan Masalah

Berdasarkan uraian yang dijelaskan pada latar belakang, maka rumusan masalah yang ada di penelitian ini yaitu :

1. Keberadaan berita yang beredar memiliki ancaman sebagai sebuah berita yang bohong dan meresahkan masyarakat. Maka dari itu, dilakukan klasifikasi dengan model algoritma *SVM* dan *RNN*

2. Belum diketahui performa algoritma yang terbaik dalam melakukan klasifikasi berita hoaks. Maka dari itu, dilakukan perbandingan antara algoritma *SVM* dan *RNN*.

### 1.3. Pertanyaan Penelitian

Berdasarkan uraian diatas, maka peneliti merumuskan beberapa pertanyaan yang akan dibahas pada penelitian ini yaitu :

1. Bagaimana membuat model algoritma *SVM* dan *RNN* untuk klasifikasi berita hoaks bahasa indonesia pada media *online*?
2. Bagaimana hasil komparasi akurasi algoritma *SVM* dan *RNN* dalam menentukan algoritma terbaik untuk klasifikasi berita hoaks?

### 1.4. Batasan Masalah

Berdasarkan uraian diatas, maka untuk membuat penelitian yang sesuai dengan masalah yang ada, diperoleh batasan-batasan masalah penelitian sebagai berikut :

1. Berfokus pada penerapan dan komparasi model algoritma *SVM* dan *RNN* untuk klasifikasi sebuah berita hoaks.
2. Data berita yang digunakan dalam klasifikasi menggunakan bahasa indonesia.
3. Berita yang diteliti adalah berita hasil publikasi media massa *online* pada situs *website turnbackhoax.id* yang berjumlah 500 berita.

### 1.5. Tujuan Penelitian

Berdasarkan uraian pertanyaan penelitian yang ada, maka dapat diketahui tujuan dari penelitian ini yaitu :

1. Membuat dan menerapkan model algoritma *SVM* dan *RNN* untuk klasifikasi sebuah berita hoaks bahasa indonesia pada media massa *online*.
2. Mengetahui hasil komparasi akurasi algoritma *SVM* dan *RNN* untuk klasifikasi berita hoaks.

## **1.6. Manfaat Penelitian**

Berdasarkan uraian tujuan penelitian yang ada, maka manfaat dari penelitian ini adalah untuk mengetahui teknik dan algoritma yang paling baik dan tepat antara algoritma *SVM* dengan *RNN* untuk klasifikasi berita hoaks.