

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Perkembangan teknologi pada zaman sekarang semakin canggih, perkembangan teknologi ini juga berpengaruh pada industri media. Dalam industri media perkembangan teknologi sangat berpengaruh, dimana industri media bertransformasi dari media cetak atau media fisik bertransformasi menjadi bentuk digital atau lebih sering dikenal dengan portal media online. Media merupakan perantara atau penghubung yang terletak antara dua pihak, atau sarana komunikasi seperti koran, majalah, radio, televisi, film, poster, dan spanduk (Dagun, 2006). Menurut (Arsyad, 2002) media adalah semua bentuk perantara yang digunakan oleh manusia untuk menyampaikan atau menyebar ide, gagasan atau pendapat, sehingga ide, gagasan atau pendapat yang dikemukakan itu sampai kepada penerima yang dituju. Dalam dunia jurnalistik media dibagi menjadi tiga bentuk yaitu: Media Cetak, misalnya seperti: Majalah, Koran, Surat Kabar, Media Elektronik, misalnya seperti: Radio, TV, Film atau Video, dan Media Siber / *online*, misalnya seperti: Media Sosial, Website, Portal Berita, Blog.

Media *online* adalah media massa "generasi ketiga" setelah media cetak (*printed media*) –koran, tabloid, majalah, buku– dan media elektronik (*electronic media*) – radio, televisi, dan film/video. Media *Online* merupakan produk jurnalistik *online*. Jurnalistik online – disebut juga *cyber journalism* didefinisikan sebagai “pelaporan fakta atau peristiwa yang diproduksi dan didistribusikan melalui internet”. (Romli, 2012)

Media berita *online* atau situs berita *online* adalah situs atau media berita yang menyediakan informasi *up to date* tentang peristiwa atau kejadian yang menyangkut kehidupan sehari-hari seperti pendidikan, olahraga, teknologi, politik, dan hidup sehat (Mulya, 2008). Berita yang disampaikan dalam media berita *online* atau situs berita pada umumnya terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi, kesehatan, dan lainnya. Dalam pembagian berita ke dalam kategori dan sub-kategori tersebut saat ini masih dilakukan secara manual, yang berarti pengunggah berita harus terlebih dahulu mengetahui isi dari berita yang akan diunggah secara keseluruhan untuk selanjutnya dimasukkan ke dalam kategori atau subkategori yang tepat. (S.Lin, 2008)

PT Merah Putih Media adalah perusahaan berita *online* yang memiliki situs website yaitu *merahputih.com* dalam situs berita *merahputih.com* terdapat kategori berita yaitu Indonesiaku, di dalam kategori berita Indonesiaku terdapat tiga sub-kategori yaitu Tradisi, Kuliner, dan Travel. Dalam pembagian berita ke dalam sub-kategori berita Indonesiaku saat ini masih dilakukan secara manual yang dilakukan oleh kepala redaksi jurnalis. Hal ini tentunya menjadi membutuhkan sumber daya yang besar ketika berita yang akan diterbitkan berjumlah besar. Maka dari itu dibutuhkan sebuah sistem yang dapat membantu untuk mengklasifikasi berita untuk subkategori Indonesiaku secara otomatis.

Penelitian terkait klasifikasi yang dilakukan oleh (Primajaya & Sari, 2018) menggunakan algoritma random forest. Pada penelitian ini, algoritma random forest digunakan untuk memprediksi curah hujan yang bisa berpotensi banjir dapat menghasilkan tingkat akurasi yang baik dengan skor akurasi keberhasilan

99,4453% dalam proses klasifikasi pada data yang berukuran besar dan efektif untuk mengatasi data yang tidak terstruktur.

Penelitian terkait klasifikasi yang dilakukan oleh (Putri, et al., 2019) menggunakan algoritma random forest. Pada penelitian ini, algoritma random forest dibandingkan dengan algoritma multilayer perceptron, naïve bayes, support vector Machine, decision Tree untuk analisis konten hoax pada berita Indonesia, hasil dari perbandingan algoritma dilakukan oleh (Putri, et al., 2019) menyatakan algoritma random forest cocok untuk konten hoax pada berita Indonesia karena mendapat nilai akurasi keberhasilan sebesar 76,47%. Hasil yang didapat pada penelitian tersebut adalah nilai tertinggi dari hasil perbandingan algoritma yang diteliti.

Penelitian terkait klasifikasi yang dilakukan (Parmar, et al., 2014) menggunakan algoritma random forest. Pada penelitian ini, algoritma random forest digunakan untuk menganalisis sentimen untuk review film, dalam penelitian yang dilakukan oleh (Parmar, et al., 2014) menggunakan dua dataset dengan jumlah data yang berbeda, dataset pertama berjumlah 1400 data *review film* dan dataset kedua berjumlah 2000 data *review film*, hasil akurasi dari kedua dataset yang digunakan juga berbeda pada dataset pertama dengan jumlah data 1400 *review film* mendapat nilai akurasi 87.85% dan untuk dataset kedua dengan jumlah data 2000 *review film* mendapatkan nilai akurasi 91,00%. Algoritma random forest dapat memberikan hasil yang kompetitif dan dapat memberikan hasil yang lebih baik jika hyperparameter disetel dengan baik.

Pada penelitian yang dilakukan (Fauzi, 2018) , algoritma Random Forest terbukti dapat menangkap seluruh fitur dalam dataset dengan baik. Hal tersebut

dibuktikan dengan nilai rata rata Out of Bag atau OOB sebesar 82,9 saat model digunakan untuk permasalahan klasifikasi sentimen.

Penelitian terkait klasifikasi yang dilakukan (Akinyelu & Adewumi, 2014) menggunakan algoritma random forest. Pada penelitian ini, algoritma random forest digunakan untuk klasifikasi email phishing dalam penelitian ini menghasilkan tingkat akurasi sebesar 99,7% untuk klasifikasi email phishing.

Untuk mengembangkan model klasifikasi otomatis pada sub-kategori Indonesiaku pada website *merahputih.com* dapat menggunakan algoritma Random Forest untuk mengklasifikasi sub-kategori pada kategori berita Indonesiaku. Menurut (V. Y. Kulkarni, 2014), metode ini merupakan sebuah model ensemble (kumpulan) dari metode pembelajaran berbasis pohon keputusan sebagai *base classifier*.

Berdasarkan penelitian sebelumnya menggunakan algoritma Random Forest, maka pada penelitian ini akan dibuat Implementasi Metode Random Forest Classifier untuk Klasifikasi Sub- Kategori Artikel Media *Online*. Penelitian ini juga menggunakan fitur ekstraksi TF-IDF untuk pembobotan setiap *term* dalam sebuah dokumen (Amelia Rahman, 2017).

## **1.2.Rumusan Masalah**

Berdasarkan kebutuhan yang diuraikan pada latar belakang yang di atas, maka dibuat rumusan permasalahan sebagai berikut:

1. Bagaimana mengimplementasikan algoritma Random Forest Classifier untuk mengklasifikasi subkategori berita Indonesiaku pada situs berita merahputih.com ?

2. Berapa tingkat performa dari model Algoritma Random Forest Classifier yang diukur dengan metrik pengukuran F1-Score untuk klasifikasi subkategori Indonesiaku pada berita merahputih.com?

### **1.3. Batasan Masalah**

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Jumlah berita yang diambil sejumlah 2000 data yang diperoleh dari *website* merahputih.com. Data berita diperoleh dari *website* merahputih.com.
2. Tipe data yang digunakan berupa *text* yang berisikan judul dan isi berita.
3. *Ekstraksi* fitur dilakukan menggunakan fitur TFIDF.
4. Performa model diukur menggunakan tingkat F1 score.
5. Terdapat 3 sub-kategori pada konten kategori indonesiaku pada *website* merahputih.com yaitu tradisi, kuliner, dan travel.
6. Data pada penelitian ini diperoleh dari situs berita online milik PT. Merah Putih Media yaitu *merahputih.com* hasil data proses *crawling* atas seizin dari pihak perusahaan terkait.

### **1.4. Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah.

1. Mengimplementasikan algoritma Random Forest Classifier untuk mengklasifikasi subkategori berita Indonesiaku pada situs berita merahputih.com

2. Mengetahui tingkat performa dari model Algoritma Random Forest Classifier yang diukur dengan metrik pengukuran F1-Score untuk klasifikasi sub-kategori Indonesiaku pada berita merahputih.com

### **1.5. Manfaat Penelitian**

Dengan mengetahui performa algoritma random forest menggunakan fitur ekstraksi TF-IDF, dapat diketahui apakah metode ini cocok digunakan untuk klasifikasi sub-kategori pada kategori berita Indonesiaku. Jika model yang dibangun berhasil maka model tersebut dapat digunakan oleh PT Merah Putih Media atau perusahaan lain yang bergerak di bidang media berita online.