



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen adalah salah satu bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap suatu entitas tertentu seperti, jasa, organisasi, individu, masalah, peristiwa, topik dan atribut lainnya (Liu, 2012). Analisis sentimen dapat dimanfaatkan untuk melakukan ekstraksi opini-opini dari dokumen, komentar, sosial media, *review blog*, dan data-data lainnya. Analisis sentimen memiliki beberapa metode, salah satu diantaranya adalah metode pendekatan *machine learning*. Metode pendekatan *machine learning* menghasilkan model yang digunakan untuk proses klasifikasi, namun menggunakan sumber atau *resource* untuk melakukan pemrosesan data sampai dengan membangun model klasifikasi. Semakin besar *dataset* yang digunakan, maka akan semakin besar pula sumber atau *resource* yang diperlukan untuk melakukan proses data tersebut, maka dari itu diperlukan informasi pengetahuan dari model yang telah dibangun sebelumnya untuk digunakan pada *data train* untuk model yang akan dibangun berikutnya.

Transfer learning adalah metode Deep Learning yang menerapkan pengetahuan atau *knowledge* dari domain yang berbeda namun terkait ke domain tujuan (Di Zhang dkk, 2019). *Transfer learning* dianalogikan dengan seseorang belajar untuk mengendarai sepeda, maka informasi berupa pengetahuan atau *knowledge* untuk menstabilkan sepeda tersebut bisa diterapkan untuk menstabilkan motor. Tujuan dari transfer learning adalah untuk meningkatkan proses

pembelajaran sesuai target domain melalui transfer pengetahuan pada data *train* (Di Zhang dkk, 2019). Dengan menerapkan informasi berupa *knowledge* dari model yang telah dibangun sebelumnya, diharapkan dapat mengurangi penggunaan sumber atau *resource* untuk melakukan proses *dataset* tersebut dalam membangun model berikutnya.

Algoritma Random Forest adalah salah satu metode berbasis klasifikasi dan regresi dimana terdapat proses agregasi pohon keputusan (Dhawangkhar, dan Riksakomara 2017). Penelitian terkait mengenai klasifikasi teks dengan algoritma Random Forest dan TF-IDF telah banyak dilakukan salah satunya adalah penelitian oleh M AliFauzi. (2018) yang mengklasifikasikan teks Bahasa Indonesia dengan algoritma Random Forest menggunakan fitur Bag-of-word dan metode pembobotan Term Frequency Inverse Document Frequency (TF-IDF). Disimpulkan bahwa penelitian tersebut memberikan hasil akurasi yang baik, yaitu kinerja dengan skor *out-of-bag* (OOB) 0,829. Penelitian terkait dengan penggunaan TF-IDF selanjutnya adalah penelitian yang dilakukan oleh Yulius Denny dkk (2019) yang menerapkan algoritma TF-IDF untuk *text mining*. Penelitian tersebut menyimpulkan bahwa penggunaan TF-IDF merupakan metode yang tepat untuk digunakan dalam pencarian kata di tiap dokumen dan dapat membantu pengguna mendapatkan dokumen terkait sesuai dengan *query* yang telah dimasukkan.

Penelitian terkait mengenai analisis sentimen dengan menggunakan teknik N-gram telah banyak dilakukan, salah satunya adalah penelitian oleh Wahyu Candra Inddhiarta (2017) yang melakukan analisis sentimen pemilihan kepala daerah Jakarta dengan menggunakan N-gram dan algoritma Naïve Bayes. Dari

penelitian tersebut, disimpulkan bahwa berdasarkan ketiga jenis token N-gram, yaitu *unigram*, *bigram*, dan *trigram* dengan metode Naïve Bayes, nilai akurasi tertinggi terdapat pada penggunaan *bigram* yaitu 0,823, menunjukkan bahwa dengan menggunakan *bigram* ketepatan akurasi dari sistem lebih baik daripada *unigram* dan *trigram*. Nilai presisi tertinggi juga terdapat pada penggunaan *bigram* yaitu 0,76. Namun pada nilai *recall* nilai tertinggi terdapat pada penggunaan *trigram* yaitu 0,898, sehingga disimpulkan penggunaan *bigram* dalam pengklasifikasian data lebih baik daripada menggunakan *unigram* atau *trigram*. Berdasarkan penelitian yang telah ada sebelumnya, penelitian ini mencoba untuk menganalisis seberapa besar pengaruh penerapan *transfer learning* yang bertujuan untuk mengurangi penggunaan *resource* atau sumber daya terhadap hasil uji performa pada model klasifikasi yang dibangun dengan menggunakan algoritma Random Forest dengan menggunakan N-gram dan TF-IDF.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana mengimplementasikan algoritma Random Forest dengan N-gram dan TF-IDF pada analisis sentimen?
2. Seberapa besar tingkat performa model yang dirancang dengan menggunakan metode *transfer learning* ?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dapat dijabarkan menjadi beberapa poin sebagai berikut.

1. *Datasets* terdiri dari kumpulan *review* dari pengguna pada produk yang dimiliki oleh perusahaan Amazon, dan Yelp.
2. *Datasets review* yang digunakan berupa teks Bahasa Inggris.
3. *Transfer learning* akan dilakukan dengan mengirimkan pengetahuan dari satu model Random Forest Classifier berupa fitur-fitur dan nilai kepentingan setiap fitur-fitur tersebut kepada model baru yang akan dibangun, sehingga model tersebut tidak menggunakan fitur yang dianggap tidak penting.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang dijelaskan sebelumnya, penelitian ini memiliki tujuan sebagai berikut.

1. Menerapkan algoritma Random Forest dengan menggunakan metode Bigrams dan TF-IDF untuk klasifikasi analisis sentimen *review* oleh *user*.
2. Mengukur dan mengetahui seberapa besar hasil performa model jika diterapkan metode *transfer learning* pada analisis sentimen dengan algoritma Random Forest.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Memberikan fasilitas kepada masyarakat untuk mengetahui sentimen dari suatu kalimat. Sehingga dapat mengetahui apakah suatu sentimen bersifat positif dan negatif untuk diaplikasikan pada analisis reputasi seorang tokoh, perusahaan, dan pelayanan.
2. Memberikan fasilitas kepada masyarakat yang memiliki penelitian seputar analisis sentimen untuk menggunakan informasi pengetahuan dari model yang telah dilatih sebelumnya yang bertujuan untuk lebih menghemat penggunaan sumber atau *resource* jika *transfer learning* memberikan hasil yang positif terhadap model yang akan dilatih selanjutnya.