

1. Pendahuluan

1.1 Latar Belakang

Twitter ialah salah satu media sosial yang kerap digunakan para pengguna untuk menuliskan apa yang sedang terjadi dan apa yang sedang banyak orang bicarakan dalam waktu terkini. Selain itu, pengguna juga dapat berkomunikasi, dan bertukar informasi seperti berita, foto dan juga video [1]. Kemudahan masyarakat dalam menyebarkan berita di twitter mengakibatkan timbul kebiasaan baru di masyarakat yaitu menerima berita tersebut tanpa mengetahui kebenaran dari berita yang mereka terima [2].

Dalam penyebaran berita di media sosial, penting hal nya untuk menelaah keaslian dari berita tersebut. Salah satu berita yang gemar dibicarakan yaitu berita mengenai Covid-19. Beberapa orang yang tidak bertanggung jawab kerap menyebarkan berita *hoax* dengan sengaja, yang mengakibatkan dampak buruk bagi masyarakat. Sebagian individu atau kelompok akan mulai mudah percaya pada berita yang tersebar di media sosial, perubahan karakter serta menjadikan media sosial sebagai alat provokasi untuk mengarahkan masyarakat atau para pembaca kepada opini negatif dari berita *hoax*. Masalah ini menunjukkan bahwa berita *hoax* memiliki dampak yang sangat buruk terhadap kasus Covid-19 ini. Justru seharusnya masyarakat membantu pemerintah dalam menekan penyebaran berita *hoax* di media sosial. Oleh karena itu, penulis melakukan penelitian mengenai deteksi berita *hoax* pada media sosial Twitter menggunakan dua metode, yaitu metode CNN dan SVM. Kedua metode ini merupakan metode *Deep Learning* yang kerap digunakan dalam pengklasifikasian teks.

Deep learning merupakan metode yang populer di kerap digunakan dengan digabungkan dengan *word embeddings* serta dalam pengklasifikasian berita *hoax* [3]. Dalam penelitian Rath menggunakan metode Inductive Representation Learning pada tahun 2020 [4] mengungkapkan bahwa pentingnya mendeteksi penyebar berita *hoax* akan membantu menekan penyebaran berita *hoax* yang cepat di media sosial. Pada penelitian yang dilakukan oleh Yuliani pada tahun 2019 [5] dengan menggunakan metode Machine Learning menjelaskan bahwa meski berita *hoax* bukan suatu ancaman, namun persepsi yang muncul dari tersebarnya berita *hoax* akan mempengaruhi kondisi sosial masyarakat dan politik suatu negara. Penelitian yang dilakukan oleh Widaretna menggunakan metode Doc2Vec pada tahun 2021 [6] mengungkapkan bahwa semenjak pandemi COVID-19 penyebaran berita *hoax* terus meningkat, hal ini berdasarkan pernyataan Direktur Pengelola Informasi Kementerian Komunikasi dan Informatika [7] menyebutkan bahwa sejak 23 Januari 2020 hingga 15 Juni 2020 terdapat 850 berita *hoax* terkait COVID-19 yang beredar di kalangan masyarakat melalui berbagai media sosial. Bahkan hingga 23 September 2020, berita *hoax* di Indonesia terkait COVID 19 sudah mencapai 1984 berita.

Berdasarkan permasalahan tersebut, penelitian mengimplementasikan metode CNN dan SVM guna mendeteksi berita *hoax* pada media sosial Twitter. Penulis menggunakan dua metode ini karena masih jarang ditemukan pengembangan dua metode ini pada penelitian-penelitian sebelumnya. Sistem ini melibatkan beberapa proses dan skenario seperti *crawling data*, *preprocessing data*, *word embedding*, *splitting data*, kemudian proses *modelling* menggunakan metode *Convolutional Neural Network* (CNN) dan *Support Vector Machine* (SVM) juga meningkatkan akurasi dengan menggunakan ekstraksi fitur TF-IDF (*Term Frequency Inverse Document Frequency*) dan pembobotan N-Gram.

Topik yang diangkat pada penelitian ini adalah kasus Covid-19. Penulis melakukan *crawling* dataset di media sosial Twitter dengan data Covid-19 pada tahun 2020. Berawal dari laporan pertama virus Covid-19 yang masuk ke Indonesia pada 2 Maret 2020 bertepatan di kota Depok, Jawa Barat, menyebabkan masyarakat panik dan melakukan cuitan-cuitan di media sosial Twitter. Dari hari itu, kasus Covid-19 bertambah banyak hingga tahun 2022. Maraknya berita *hoax* yang tersebar di media sosial Twitter menjadi tujuan penelitian ini dilakukan.

Agar penyebaran berita *hoax* tersebut dapat diketahui dan juga dapat terkontrol dengan baik, maka diperlukan suatu metode yaitu dengan menggunakan metode CNN dan SVM. CNN menggunakan sebuah filter untuk melakukan kinerjanya untuk menerima lokal data satu atau dua dimensi yang kemudian diproses dengan sangat baik pada tugas-tugas computer *vision*. Sedangkan, SVM merupakan suatu metode yang menggunakan ruang hipotesis yang memiliki fungsi linear dalam fitur dimensi tinggi, kemudian dilatih menggunakan algoritma pembelajaran berdasarkan teori komputasi [8].

1.2 Topik dan Batasannya

Berdasarkan latar belakang yang telah disampaikan sebelumnya, maka topik permasalahan yang diambil ialah mengenai sistem deteksi *hoax* berita Covid-19 dengan implementasi metode CNN dan SVM pada media sosial Twitter, bagaimana penerapan ekstraksi fitur dalam mendeteksi berita *hoax* dengan CNN dan svm, kemudian sejauh mana performansi kedua model dapat memberikan prediksi akurat dalam penyebaran berita *hoax* tersebut. Batasan masalah diambil dari dataset dengan topik Covid-19 dengan menggunakan Bahasa Indonesia, dataset diambil sebanyak 2020 dengan periode waktu 3 bulan, dari tanggal 2 Februari 2020 hingga tanggal 4 April 2020.

1.3 Tujuan

Pada penelitian ini memiliki tujuan yaitu mendeteksi berita *hoax* Covid-19 dengan menggunakan metode CNN dan SVM, mengimplementasikan fitur ekstraksi untuk deteksi *hoax* dan membandingkan performa metode

CNN dan SVM dalam mendeteksi berita hoax di Twitter agar mendapatkan hasil model dengan performansi yang terbaik dalam mengklasifikasikan berita sebagai hoax atau fakta. Berita yang digunakan dalam penelitian ini dibagi menjadi berita fakta dan berita *hoax*. Dengan penelitian ini, diharap dapat menekan tingkat penyebaran berita *hoax* pada media sosial terutama di Twitter.

Tabel 1. Tujuan, Pengujian dan Kesimpulan

No	Tujuan	Pengujian	Kesimpulan
1	Mendeteksi berita hoax yang beredar di Twitter terkait dengan topik pandemi Covid-19 pada tahun 2020	Dilakukan implementasi klasifikasi model CNN dan SVM dengan 3 perbandingan pembagian data agar dapat mengetahui tingkat akurasi kedua model tersebut	Mengetahui sejauh mana kedua model dapat memberikan akurasi yang terbaik dari topik tersebut, dan juga membantu menekan penyebaran informasi hoax yang dapat mempengaruhi persepsi dan tindakan masyarakat terkait dengan pandemi Covid-19.
2	Mengimplementasikan fitur ekstraksi pada pemodelan CNN dan SVM	Melakukan peningkatan akurasi terhadap hasil dari kedua model dengan mengimplementasikan ekstraksi fitur dengan bantuan TF-IDF dan pembobotan N-Gram	Proses ini dilakukan untuk meningkatkan hasil akurasi dari model CNN dan SVM
3	Membandingkan performa dua metode yang berbeda yaitu CNN dan SVM dalam mendeteksi berita hoax Covid-19 di Twitter	Melakukan perbandingan hasil akurasi dari beberapa proses dan skenario yang telah dilakukan. Skenario yang dilakukan antara lain; pengujian baseline, pengujian bigram dan trigram, pengujian model kombinasi N-Gram	Didapatkan hasil akurasi dari proses dan skenario yang dilakukan untuk kedua metode yakni CNN dan SVM yang kemudian dibandingkan untuk mengetahui akurasi terbaik

1.4 Organisasi Tulisan

Rencana kegiatan yang akan dilakukan dalam penelitian ini, yakni sebagai berikut.

1. Pada Bab 1 disampaikan pendahuluan, yang berisi latar belakang, topik dan batasan masalah, kemudian tujuan.
2. Bab 2 berisi landasan teori yang mendukung penelitian ini. Pada bagian akan ini dijelaskan teori dari jurnal-jurnal dengan pembahasan *Deep Learning*.
3. Bab 3 berisi penjelasan mengenai sistem penelitian yang dibangun, berawal dari alur sistem yang digunakan kemudian penjelasan dari setiap alur yang dikerjakan.
4. Pada Bab 4 berisi hasil pengujian dan analisis dari hasil penelitian yang telah dilakukan.
5. Bab 5 berisi kesimpulan hasil penelitian yang telah dibangun, berikut sedikit saran bagi peneliti selanjutnya.