

1. Pendahuluan

Latar Belakang

Corona Virus (Covid-19) merupakan sebuah virus yang pertama kali muncul di kota Wuhan Cina pada bulan Desember 2019 yang saat ini sudah menyebar ke seluruh dunia [1], Berdasarkan data yang di disampaikan oleh pemerintah Indonesia, hingga saat ini 4,244,761 orang terkonfirmasi positif Covid-19 [2]. Berbagai upaya di lakukan pemerintah Indonesia untuk menekan angka penyebaran Covid-19, salah satunya adalah melakukan vaksinasi terhadap warga Indonesia.

Tindakan dari pemerintah tersebut menuai Pro dan Kontra dari berbagai macam pihak, adapun kelompok atau masyarakat yang tidak setuju terhadap kebijakan tersebut. Dengan adanya pihak yang tidak setuju dengan kebijakan tersebut muncul lah berita-berita bohong atau palsu yang sering kita sebut Hoax, Hoax bermunculan dimana-mana khususnya di internet, motivasi dari pihak-pihak yang menyebarkan berita tersebut adalah untuk menghasut atau menggiring opini publik ke arah yang mereka inginkan [3]. Pada masa pandemi terjadi peningkatan signifikan jumlah Hoax tentang Covid-19 dan vaksin dengan jumlah mencapai 1733 berita [4]. Sama halnya dengan Covid-19, Hoax sendiri juga dapat dikatakan sebagai wabah, karena penyebarannya yang sangat cepat di internet.

Untuk mengetahui apakah berita yang tersebar di internet merupakan benar atau palsu di perlukan sebuah sistem pendeteksi hoax, salah satunya penelitian mengenai sistem pendeteksi hoax yang dilakukan oleh Madani Y. [5], dalam penelitiannya data yang digunakan merupakan tweets twitter yang memiliki unsur hoax dan non-hoax, pendekatan yang digunakan adalah dengan Kecerdasan Buatan dengan membandingkan metode Decision Tree, Random Forest dan Logistic Regresion yang dimana hasil terbaik di dapatkan dengan metode Random Forest, penelitian sejenis yang dilakukan oleh Aldwairi M. [6], menggunakan metode yang sama dengan hasil akurasi, Recall, Precision dan F-Measure di atas 90% tetapi penelitian ini sendiri memiliki kelemahan yang dimana data yang digunakan tidak di perlihatkan seperti apa dan jumlah data yang digunakan tidak dijelaskan. Peneliti juga melakukan perbandingan terhadap 2 metode klasifikasi yang berbeda yaitu Support Vector Machine (SVM) dan Stochastic Gradient Descent (SGD) dengan pembobotan Fitur TF-IDF mendapatkan hasil terbaik dengan metode SGD namun pada penelitian ini data yang digunakan masih sangatlah kecil [7]. Penelitian lain oleh Adzlan Ishak [8], menggunakan pendekatan Distance Based dengan metode Levenshtein Distance, metode ini tergolong unik dikarenakan cara mendeteksi Hoax dengan melihat nilai similarity dari struktur kata untuk melakukan klasifikasi, penelitian ini mendapatkan hasil yang cukup baik yaitu 74% pada sensitivity, dimana penelitian ini tidak menggunakan data pada sosial media maupun berita online melainkan hanya melalui email saja.

Pada Sebagian besar penelitian permasalahan yang dialami adalah pada data yang dikumpulkan, dan pada penelitian-penelitian di atas sudah banyak membahas atau menggunakan metode klasifikasi standar seperti Decision Tree, SVM, SGD, Random Forest dan lain-lain. Oleh karena itu Pada riset ini akan membuat sistem Pendeteksi Hoax dengan menggunakan metode Levenshtein Distance karena sudah teruji pada penelitian sebelumnya, dan memiliki potensi bila dilakukan penelitian lebih lanjut mengenai pembuatan sistem pendeteksi hoax menggunakan metode tersebut. Kontribusi dari penelitian ini adalah penggunaan data yang lebih banyak dan penambahan tf-idf. data yang digunakan dalam penelitian ini adalah berita Indonesia yang dikumpulkan dari website pemerintah Indonesia terpercaya yang diberi label secara manual, dengan menambahkan Feature Extraction TF-IDF dan menggunakan metode K-Fold Cross-Validation untuk Mengukur kinerja sistem. Tujuan penggunaan metode ini adalah untuk mengetahui berita hoax yang lebih akurat berdasarkan kemiripan dan struktur kalimat pada korpus/dataset.

Topik dan Batasannya

Topik utama dalam penelitian ini adalah mendeteksi berita tentang vaksinasi Covid-19 menggunakan metode Levenshtein Distance sebagai klasifier, dengan melakukan beberapa percobaan/eksperimen, menambahkan fitur ekstraksi TF-IDF untuk melihat perbandingan hasil yang diberikan dengan adanya penambahan fitur ekstraksi TF-IDF juga menggunakan K-fold Cross validation untuk memvalidasi hasil keluaran sistem.

Batasan pada penelitian ini adalah fitur dan kelas yang digunakan jumlahnya masih kurang dan juga dataset yang persebaran datanya masih tidak merata antara tiap kelas.

Tujuan

Tujuan dari penelitian ini adalah membangun sistem pendeteksi hoax menggunakan metode Levenshtein Distance dengan menambahkan ekstraksi fitur TF-IDF agar mendapatkan keluaran sistem yang lebih efisien dibandingkan penelitian terdahulu.

Organisasi Tulisan

Struktur pada penulisan penelitian ini disusun sebagai berikut: di awali dengan pendahuluan yang berisi latar belakang tentang penelitian ini, selanjutnya ada Studi Terkait yang berisi literatur dari penelitian terdahulu yang menjadi landasan penelitian ini, bagian ketiga merupakan penjabaran model sistem yang akan dibangun juga

penjelasan mengenai metode-metode yang akan digunakan pada sistem, bagian keempat merupakan deskripsi hasil yang didapatkan pada hasil eksperimen yang sudah dilakukan, bagian terakhir merupakan kesimpulan dari penelitian.