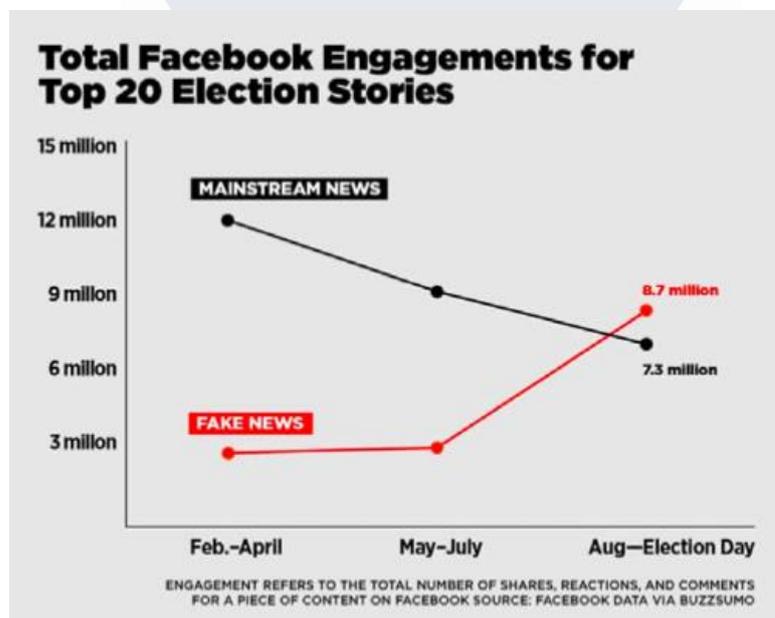


BAB I

PENDAHULUAN

1.1 Latar Belakang

Hoaks, atau yang lebih dikenal sebagai berita palsu atau disinformasi, kini dianggap sebagai salah satu ancaman terbesar terhadap demokrasi, jurnalisme, dan kebebasan berekspresi [1]. Kemampuan berita palsu untuk memengaruhi opini publik dan membentuk naratif yang mungkin tidak akurat dapat merusak integritas proses demokratis, seperti terlihat dalam kasus referendum Brexit dan pemilihan presiden AS 2016 [2]. Pada tahun 2016, Inggris menggelar referendum untuk menentukan apakah negara tersebut akan tetap menjadi bagian dari Uni Eropa atau keluar (Brexit). Selama kampanye referendum, berita palsu tersebar luas, termasuk klaim mengenai dampak finansial, imigrasi, dan klaim lainnya yang belum tentu akurat.



Gambar 1.1 Sebaran hoaks selama periode pemilihan presiden AS 2016
(Sumber: www.buzzfeednews.com)

Sebaran berita palsu sangat mencuat selama bulan-bulan krusial kampanye pemilihan presiden AS 2016. Selama periode tersebut, 20 berita palsu yang paling sering dibahas menghasilkan 8.711.000 kali *shared*, *reaction*, dan *comments* di

Facebook. Ironisnya, jumlah ini lebih banyak dibandingkan dengan 7.367.000 interaksi yang diperoleh untuk 20 berita pemilihan yang paling sering dibahas yang diposting oleh 19 situs berita utama (Gambar 1.1) [3]. Dalam kedua kasus tersebut, dampak berita palsu sulit diukur secara pasti, tetapi banyak yang percaya bahwa informasi yang tidak akurat yang tersebar melalui media sosial dapat memengaruhi pola pikir pemilih dan proses keputusan, membentuk opini yang tidak mencerminkan realitas.

Berita hoaks memiliki potensi untuk memutarbalikkan opini publik dan menciptakan polarisasi politik yang lebih dalam. Informasi palsu yang menyebar dapat memperkuat pandangan ekstrem, memperdalam perpecahan antar kelompok, dan memunculkan ketidakpercayaan terhadap lembaga-lembaga demokratis [4]. Dalam studi yang dipublikasikan di *Proceedings of the National Academy of Sciences* pada tahun 2018, Vosoughi et al. menyelidiki sebaran berita palsu di Twitter dan menemukan bahwa berita palsu memiliki kemungkinan lebih besar untuk menjadi viral daripada berita yang benar, terutama dalam konteks berita politik [5]. Polarisasi politik yang semakin meningkat dapat menghambat proses pengambilan keputusan yang efektif, merusak hubungan antarwarga, dan memperumit upaya untuk mencapai konsensus dalam masyarakat [6]. Dalam mengatasi dampak buruk dari berita hoaks, langkah-langkah efektif melibatkan pendidikan literasi media, peningkatan transparansi sumber informasi, dan implementasi teknologi deteksi penyebar berita palsu yang lebih canggih [7]. Dengan demikian, pendekatan *data mining* dan prediksi penyebar berita hoaks akan dilakukan dalam penelitian ini.

Data mining adalah proses untuk mengekstraksi informasi implisit, yang sebelumnya belum ditemukan dan mungkin bermanfaat dari kumpulan data yang besar. Algoritma *data mining* memperkirakan pola yang dapat digunakan untuk mengidentifikasi penyebar berita hoaks. Dalam konteks identifikasi penyebar berita hoaks, *Social Network Analysis* (SNA) dapat menjadi langkah awal yang kuat. SNA dapat membantu mengungkapkan pola-pola kompleks dalam data, seperti hubungan antar entitas (misalnya, pengguna media sosial atau sumber berita) dan pola penyebaran informasi. Informasi yang ditemukan melalui SNA dapat

membantu dalam pengelompokkan (*clustering*) dengan menggunakan algoritma *machine learning*.

Machine learning menjadi solusi yang menjanjikan untuk menangani kompleksitas identifikasi penyebar berita hoaks. Akun penyebar berita hoaks yang tersebar luas di media sosial dan platform online dapat dengan efisien diolah oleh *machine learning*, memungkinkan untuk penanganan cepat dan analisis menyeluruh [8]. Algoritma *machine learning* dapat digunakan untuk mendeteksi pola dan kesamaan dalam penyebaran berita. Hal tersebut dapat membantu dalam mengidentifikasi fitur atau ciri-ciri khusus yang sering terkait dengan penyebar berita hoaks. Metode klasifikasi, memungkinkan sistem untuk secara otomatis mengidentifikasi penyebar hoaks berdasarkan data latih yang dilakukan pelatihan sebelumnya.

Klasifikasi termasuk ke dalam kategori algoritma *supervised learning*. Dalam konteks klasifikasi, tujuan utamanya adalah untuk memprediksi label kelas dari akun-akun yang belum terlihat sebelumnya berdasarkan pola yang telah dipelajari data lain. Terdapat beberapa algoritma klasifikasi yang diterapkan pada prediksi berita hoaks. Pada penelitian yang berjudul *Machine Learning Technique Based Fake News Detection*, model *machine learning* terbaik untuk mendeteksi *fake news* adalah *Naïve Bayes Classifier* (NBC) dengan *F1-macro* rata-rata sebesar 32%. Model lain yang diuji melibatkan *Logistic Regression*, *Stochastic Gradient Descent*, LSTM dan AWD_LSTM. Namun, NBC memberikan hasil terbaik dalam penelitian ini [9]. Penelitian lain yang berjudul *Fake News Spreader Detection Using Naïve Bayes Classifier and Logistic Regression* menggunakan algoritma *MultinomialNB*, *GaussianNB*, *BernoulliNB*, dan *LogisticR*. Hasil penelitian menunjukkan bahwa model *Logistic Regression* memberikan akurasi tertinggi, mencapai 95%. Penelitian ini memberikan bukti bahwa pendekatan klasifikasi dengan menggunakan *Naïve Bayes* dan *Logistic Regression* mampu mencapai tingkat akurasi yang tinggi dalam mendeteksi penyebar berita hoaks [10].

Dalam penelitian ini, dataset diperoleh dari hasil klusterisasi K-Means menggunakan data *scraping* media sosial X menggunakan API. Dataset ini terdiri dari beberapa atribut, termasuk *Vertex*, *Relationship*, *Relationship Date (UTC)*,

Tweet dan beberapa atribut lainnya. Penelitian ini bertujuan untuk melakukan klusterisasi data dan membandingkan model prediktif dari algoritma *Naïve Bayes* dan *Logistic Regression*. Proses klusterisasi dapat membantu memahami struktur data dan mengidentifikasi pola atau kelompok tertentu. Setelah itu, model prediktif akan dibangun menggunakan algoritma *Naïve Bayes* dan *Logistic Regression* kemudian dikomparasikan untuk menentukan model prediksi yang terbaik. Metode ini diharapkan dapat memberikan wawasan yang lebih baik tentang bagaimana berita hoaks menyebar dan memengaruhi pemilu, serta membantu dalam upaya deteksi dini dan mitigasi dampaknya.

1.2 Rumusan Masalah

1. Algoritma mana di antara *Naïve Bayes* dan *Logistic Regression* yang menunjukkan kinerja terbaik dalam melakukan klasifikasi akun penyebar berita hoaks melalui perbandingan tingkat akurasi, sensitivitas, dan presisi?
2. Bagaimana efisiensi kinerja dari algoritma *Naïve Bayes* dan *Logistic Regression* berdasarkan *time processing*?

1.3 Batasan Masalah

1. Data yang digunakan adalah dataset yang diperoleh dari *scraping* media sosial X menggunakan API.
2. Analisis akhir yang dilakukan adalah dengan membandingkan *time processing*, akurasi, sensitivitas, dan presisi antara algoritma *Naïve Bayes* dan *Logistic Regression*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

1. Membandingkan dan menganalisis kinerja antara algoritma *Naïve Bayes* dan *Logistic Regression* melalui perbedaan tingkat akurasi, sensitivitas, dan presisi algoritma dalam memprediksi penyebaran berita hoaks, untuk menentukan algoritma yang paling efektif dalam tugas tersebut.
2. Mengevaluasi efisiensi kinerja dari algoritma *Naïve Bayes* dan *Logistic Regression* dengan mempertimbangkan *time processing*, untuk

menentukan algoritma yang tidak hanya akurat tetapi juga efisien dalam prosesnya.

1.4.2 Manfaat Penelitian

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam penanganan berita hoaks dengan mengidentifikasi dan memprediksi keberadaan berita palsu menggunakan algoritma *machine learning*.
2. Penelitian ini dapat memberikan pemahaman lebih lanjut terkait kinerja dan karakteristik algoritma *Naïve Bayes* dan *Logistic Regression* dalam konteks deteksi berita hoaks.
3. Hasil perbandingan *time processing*, akurasi, sensitivitas, dan presisi antara algoritma *Naïve Bayes* dan *Logistic Regression* dapat memberikan wawasan tentang kelebihan dan kekurangan masing-masing algoritma.
4. Hasil penelitian dapat menjadi pedoman bagi peneliti, praktisi, dan pengembang sistem dalam memilih algoritma yang sesuai untuk deteksi berita hoaks berdasarkan karakteristik yang diinginkan.
5. Penelitian ini dapat memberikan dasar untuk pengembangan teknologi deteksi berita hoaks yang lebih canggih dan efektif dalam mengatasi tantangan informasi palsu di media sosial.

1.5 Sistematika Penulisan

Penulisan skripsi ini mengikuti sistematika yang terorganisir dengan baik agar pembaca dapat memahami dengan mudah. Adapun sistematika penulisan yang digunakan adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bagian ini, memaparkan konteks penelitian dengan merinci latar belakang masalah yang menjadi dorongan utama penelitian. Selain itu, terdapat pembahasan mengenai rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan. Secara ringkas, turut dijelaskan fenomena yang menjadi fokus penelitian, dampak signifikan dari berita hoaks, serta penerapan algoritma *clustering* K-Means, juga algoritma *classification*

Naïve Bayes dan *Logistic Regression* dalam pemahaman latar belakang penelitian ini.

BAB II LANDASAN TEORI

Pada bagian ini, disajikan kajian literatur mengenai landasan teori terkait dengan penelitian ini. Terdapat pembahasan mendalam mengenai data mining, analisis jaringan, *machine learning*, algoritma *clustering* K-Means, konsep *classification*, dan algoritma *classification Naïve Bayes* dan *Logistic Regression*. Selain itu, diuraikan pula konsep-konsep terkait penelitian, seperti *time processing*, akurasi, sensitivitas, dan presisi. Keseluruhan landasan teori ini bertujuan untuk memberikan dasar pemahaman yang kokoh dalam menjalankan analisis dan interpretasi hasil penelitian.

BAB II METODOLOGI PENELITIAN

Pada bagian ini, menjelaskan secara rinci mengenai metodologi penelitian yang diterapkan, penjelasan meliputi jenis penelitian yang digunakan, sampel yang menjadi fokus, teknik pengumpulan data yang diterapkan, serta langkah-langkah analisis data yang akan diimplementasikan. Pendekatan ini dirancang untuk memberikan gambaran yang jelas mengenai prosedur penelitian dan cara pengolahan data yang mendukung validitas dan reliabilitas hasil penelitian.

BAB IV ANALISIS DAN HASIL PENELITIAN

Pada bagian ini, disajikan hasil-hasil yang telah diperoleh dari penelitian, khususnya terkait dengan performa dari dua algoritma *classification*, yaitu *Naïve Bayes* dan *Logistic Regression*. Analisis mendalam dilakukan terhadap efektivitas dan keunggulan masing-masing algoritma dalam mengidentifikasi dan memprediksi berita hoaks. Kesimpulan yang diperoleh dari hasil analisis menjadi landasan untuk memberikan rekomendasi dan interpretasi yang mendalam terhadap kontribusi penelitian ini dalam mengatasi berita hoaks.

BAB V SIMPULAN DAN SARAN

Pada bagian ini, disajikan simpulan hasil penelitian sesuai dengan tujuan yang telah ditetapkan. Analisis mendalam terhadap hasil penelitian, termasuk keunggulan dan keterbatasan dari penggunaan algoritma *classification Naïve*

Bayes dan *Logistic Regression* yang dikombinasikan dengan *clustering* algoritma K-Means. Selain itu, disajikan juga saran-saran yang dianggap perlu untuk pengembangan lebih lanjut berdasarkan temuan-temuan penelitian ini. Simpulan dan saran menjadi landasan penting untuk menyimpulkan dampak penelitian terhadap pemahaman dan penanganan berita hoaks serta memberikan arah bagi penelitian mendatang di bidang ini.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA