

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang penuh dengan perkembangan teknologi, isu keamanan dan privasi data menjadi sangat penting karena serangan siber yang marak menimbulkan kekhawatiran masyarakat tentang privasi dan kerugian finansial (Perweij, dkk, 2021). Pada tahun 2022, isu ancaman privasi data menjadi topik hangat di Indonesia. Isu ini mendapatkan perhatian besar dari masyarakat dan menciptakan beragam pro dan kontra di berbagai platform media sosial. Dalam era Revolusi Industri 4.0, akses terhadap informasi, termasuk berita, penjualan, dan lainnya, semakin mudah melalui *platform* media sosial. YouTube, sebagai salah satu media sosial populer, memainkan peran penting dalam penyiaran video yang menarik dan informatif. Dibandingkan dengan platform lain seperti Facebook, Twitter, WhatsApp, dan Instagram, YouTube menawarkan keunggulan dalam hal format video yang lebih mendalam, serta menyediakan saluran interaksi yang memungkinkan audiens memberikan tanggapan langsung melalui kolom komentar (Nnaane, 2024). Keunggulan ini menjadikannya pilihan utama dalam pengumpulan data, khususnya dalam menganalisis opini publik dan efektivitas pesan. Oleh karena itu, penelitian ini menggunakan YouTube untuk mengumpulkan data terkait opini publik mengenai ancaman privasi data di Indonesia.

Analisis sentimen adalah proses untuk mengidentifikasi dan memahami emosi, sikap, dan opini yang terkandung dalam teks. Tujuannya adalah mengetahui kecenderungan atau pandangan seseorang terhadap topik tertentu. Ini penting dalam memantau media sosial karena membantu mendapatkan gambaran tentang opini publik mengenai suatu isu atau peristiwa (Rahman, dkk, 2020). Ini memberikan wawasan tentang bagaimana masyarakat memberikan respon terhadap berita yang dipublikasikan. Dalam analisis sentimen, berbagai metode klasifikasi digunakan untuk mengklasifikasikan sentimen dalam data. Metode-metode tersebut mencakup metode *Naïve Bayes*, *Decision Tree*, SVM, JST, *Random Forest* dan lainnya. Penelitian Aufar dkk (2020), tentang analisis sentimen pada saluran YouTube Nokia Mobile menggunakan algoritma *Decision Tree* (metode C4.5) dan *Random*

Forest menunjukkan bahwa algoritma *Decision Tree* (C4.5) mencapai akurasi 89,4%, sementara algoritma *Random Forest* mencatat akurasi sebesar 88,2%. Algoritma *Decision Tree* (C4.5) merupakan salah satu algoritma yang sering digunakan dalam analisis sentimen.

Algoritma C5.0 adalah sebuah algoritma yang merupakan pengembangan dari algoritma C4.5 (Izzah dan Pitri, 2023). Algoritma C5.0 memiliki sejumlah keunggulan dibandingkan C4.5, termasuk pemilihan atribut yang efisien dan teknik pemangkasan yang mengurangi kompleksitas pohon keputusan. Selain itu, C5.0 lebih cepat dalam membangun model karena pendekatan yang efisien dalam menghitung *gain information*. Algoritma ini juga lebih efektif menangani dataset besar dan kompleks, serta mampu mengelola atribut diskrit dan kontinu, termasuk *missing values*. Model yang dihasilkan oleh C5.0 lebih ringkas dan mudah dipahami, memudahkan interpretasi hasil dan pengambilan keputusan (Pasaribu, dkk, 2022).

Studi tentang analisis sentimen pada industri kopi melibatkan penggunaan 10.646 data ulasan Twitter dan empat algoritma *machine learning*: *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Random Forest* (RF). Hasil eksperimen menunjukkan SVM dan DT memberikan akurasi tertinggi yaitu 95%, diikuti oleh RF dengan 94%, sedangkan KNN memiliki akurasi terendah, yaitu 74% (Alsemaree, dkk, 2024).

Dalam analisis sentimen, metode seleksi fitur digunakan sebagai langkah pra-pemrosesan data. Tujuannya adalah menghilangkan fitur yang tidak relevan, mengurangi biaya komputasi, dan meningkatkan kinerja metode *machine learning* (Hu, dkk, 2023). Metode *Chi-Square* adalah salah satu metode seleksi fitur yang banyak digunakan karena waktu eksekusinya yang cepat dan kemampuannya untuk meningkatkan akurasi dengan memanfaatkan atribut yang memiliki tingkat signifikansi tertinggi (Ahakonye, dkk, 2023).

Metode *Chi-Square* digunakan untuk mengevaluasi tingkat ketergantungan antara fitur-fitur dalam dataset dan kelas (Ali, dkk, 2020). Penelitian oleh Nurhayati dkk (2019), menunjukkan bahwa penerapan metode seleksi fitur *Chi-Square* dalam analisis sentimen dengan algoritma *Naïve Bayes* menghasilkan akurasi yang lebih

baik, yaitu 93,33%. Sementara itu, analisis sentimen tanpa metode seleksi fitur memperoleh akurasi 73,33%.

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma C5.0, *Random Forest* (RF), dan *Support Vector Machine* (SVM) dengan menerapkan metode seleksi fitur *Chi-Square* dalam analisis sentimen terkait berita ancaman privasi data di Indonesia. Penggunaan metode seleksi fitur *Chi-Square* sangat penting dalam penelitian ini, karena dapat membantu mengidentifikasi fitur-fitur yang paling relevan, meningkatkan akurasi model, dan mengurangi *noise* dalam data. Dengan menerapkan *Chi-Square*, diharapkan kinerja algoritma dalam mengklasifikasikan sentimen dari data komentar dapat meningkat.

1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma C5.0, *Random Forest* (RF), dan *Support Vector Machine* (SVM) dengan metode seleksi fitur *Chi-Square* dalam analisis sentimen berita ancaman privasi data di Indonesia. Selain itu, penelitian ini juga berfokus pada pemahaman sentimen masyarakat untuk membantu penyusunan kebijakan perlindungan data yang lebih efektif dan meningkatkan kepercayaan terhadap keamanan data pribadi.

1.3 Manfaat Penelitian

1. Manfaat untuk Pendidikan:

Penelitian ini memberikan wawasan yang berharga dalam memahami penerapan metode seleksi fitur dan algoritma klasifikasi dalam analisis sentimen. Temuan ini dapat digunakan sebagai referensi dalam pengajaran teknik-teknik data mining, analisis data, dan kecerdasan buatan, yang sangat relevan dalam dunia pendidikan saat ini. Selain itu, penelitian ini juga dapat berfungsi sebagai panduan dalam pengembangan kurikulum yang mengintegrasikan teknologi terbaru, serta memberikan contoh aplikasi nyata dalam menganalisis respon publik terhadap isu-isu sosial penting, seperti ancaman privasi data.

2. Manfaat untuk Masyarakat:

Penelitian ini memberikan pemahaman yang lebih dalam mengenai bagaimana teknologi, khususnya algoritma analisis sentimen, dapat digunakan untuk mengevaluasi opini publik terkait isu-isu yang mempengaruhi kehidupan mereka, seperti privasi data. Dengan pemahaman ini, masyarakat dapat lebih kritis dalam menanggapi informasi yang beredar di media sosial dan meningkatkan kesadaran mereka terkait perlindungan data pribadi. Selain itu, penelitian ini juga dapat membangun kepercayaan masyarakat terhadap teknologi dalam menangani isu-isu sosial yang krusial, serta memberikan panduan dalam merespon berbagai opini dan berita yang berhubungan dengan privasi data.



SEKOLAH PASCASARJANA