

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Penelitian Terdahulu**

Pada bagian ini, Peneliti akan menyajikan hasil dari studi terdahulu yang membahas topik yang sama atau terkait dengan penelitian yang akan dilakukan. Peneliti akan mengevaluasi dan menganalisis studi terdahulu yang relevan dengan tujuan penelitian ini dan menjelaskan bagaimana hasil dari studi tersebut dapat digunakan sebagai dasar untuk melakukan penelitian ini. Hal ini akan membantu dalam menentukan metodologi yang tepat dan dalam mengidentifikasi celah dari penelitian sebelumnya yang dapat diisi dengan penelitian ini.

Penelitian oleh (Isa & Ashianti, 2017). Penelitian ini bertujuan untuk mengidentifikasi ciri-ciri dari *cyberbullying* pada pesan dan melakukan klasifikasi menggunakan metode SVM dan Naïve Bayes. Hasil dari penelitian ini menunjukkan bahwa kernel Poly pada metode SVM adalah yang paling optimal dalam melakukan klasifikasi *cyberbullying*. Hasil akurasi yang diperoleh adalah 97.11%.

Penelitian oleh (Dalvi et al., 2020), melakukan perbandingan pada metode Naive Bayes dan SVM. Data yang digunakan berasal dari berbagai sumber seperti Kaggle, github. Penelitian ini menunjukkan bahwa metode SVM mendapatkan akurasi sebesar 71.25% dan Naïve Bayes 52.70%. Berdasarkan penelitian ini, SVM memiliki akurasi yang lebih tinggi dalam mendeteksi kalimat *cyberbullying*.

Penelitian oleh (Atoum, 2020). Tujuan dari penelitian ini adalah melakukan identifikasi teks *cyberbullying* di media sosial Twitter. Klasifikasi dilakukan dengan menggunakan metode SVM dan Naïve Bayes. Hasil menunjukkan bahwa performa dari SVM lebih baik daripada Naïve Bayes. Hasil akurasi yang diperoleh adalah 92.02% pada SVM dan 81.1 pada Naïve Bayes.

Penelitian lain dilakukan oleh (Saputra & Rosiyadi, 2019) yang menggunakan aplikasi whatsapp sebagai objek penelitian. Data yang digunakan berasal dari teks obrolan pada grup Whatsapp “Pascasarjana 2017”. Hasil pengujian menunjukkan bahwa K-NN, Naïve Bayes, dan SVM memiliki akurasi masing-masing sebesar 81.32%, 78.95%, dan 81.58%. Penelitian ini menunjukkan bahwa metode SVM lebih baik dibandingkan metode lainnya.

Hasil penelitian yang dilakukan oleh (Andriansyah et al., 2017) menunjukkan bahwa metode *support vector machine* memiliki tingkat akurasi sebesar 79,412% dalam mendeteksi kalimat *cyberbullying* pada 1053 data komentar yang diambil dari akun selebgram di Indonesia. Data tersebut diolah menggunakan bahasa pemrograman R dengan menggunakan Rstudio.

Tabel 2.1 Penelitian Terdahulu 1

No	Judul	Dimuat	Hasil
1	<i>Cyberbullying Classification using Text Mining</i>	International Conference on Informatics and Computational Sciences (ICICoS), 2017	Hasil pengujian menunjukkan bahwa <i>SVM</i> dengan <i>Poly Kernel</i> memiliki hasil paling optimal dengan nilai 97.11%
2	<i>Detecting a Twitter Cyberbullying Using Machine Learning</i>	International Conference on Intelligent Computing and Control Systems (ICICCS), 2020	Hasil perbandingan antara <i>Naïve Bayes</i> dan <i>SVM</i> menunjukkan bahwa <i>SVM</i> memiliki akurasi yang lebih tinggi dalam mendeteksi <i>cyberbullying</i> dengan nilai 71.25%
3	<i>Cyberbullying Detection Through Sentyment Analysis</i>	International Conference on Computational Science and Computational Intelligence (CSCI), 2020	Pengujian yang dilakukan menunjukkan hasil bahwa <i>SVM</i> memiliki rata-rata akurasi 92.02% dan <i>Naïve Bayes</i> memiliki akurasi 81.1%

Tabel 2. 2 Penelitian Terdahulu 2

No	Judul	Dimuat	Hasil
1	Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp	Factor Exacta, 2019	Hasil pengujian menunjukkan akurasi <i>KNN</i> 81.32%, <i>NBC</i> 78.95%, dan <i>SVM</i> 81.58%
2	Cyberbullying comment classification on Indonesian Selebgram using support vector machine method	Second International Conference on Informatics and Computing (ICIC), 2017	Hasil pengujian menunjukkan bahwa <i>Support Vector Machine</i> memiliki akurasi sebesar 79.412%

Penelitian sebelumnya menunjukkan bahwa metode *Support Vector Machine* memiliki hasil akurasi yang tinggi dalam melakukan klasifikasi *data mining*. Namun, belum ada penelitian yang membandingkan penggunaan C-SVC dan Nu-SVC dalam mengidentifikasi kalimat *cyberbullying* pada media sosial Twitter. Hal ini menjadi penting untuk diketahui karena dapat membantu dalam meningkatkan efektivitas deteksi kalimat *cyberbullying*. Penggunaan C-SVC dan Nu-SVC dapat memberikan informasi yang lebih akurat dan dapat menghindari kemungkinan kesalahan dalam deteksi kalimat *cyberbullying*. Oleh karena itu, perlu dilakukan penelitian lebih lanjut untuk mengetahui seberapa efektif kedua metode tersebut dalam deteksi kalimat *cyberbullying* pada media sosial Twitter.

## **2.2 Teori-Teori Dasar**

Bagian ini akan membahas dasar-dasar teori yang dibutuhkan untuk memahami konsep-konsep yang akan dikembangkan lebih lanjut dalam materi selanjutnya.

### **2.2.1 Data Mining**

*Data mining* adalah proses menemukan pola yang menarik dan bermanfaat dalam sekumpulan data (*database*) dengan tujuan untuk mengekstrak informasi dan pengetahuan yang akurat dan potensial yang dapat membantu dalam pengambilan keputusan (Setiyani et al., 2020).

### **2.2.2 Analisis Sentimen**

Analisis sentimen atau penambangan opini adalah suatu bidang studi untuk menganalisis pendapat orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik (Septian et al., 2019). Analisis sentimen bertujuan untuk mengetahui sentimen yang terkandung dalam suatu teks, baik berupa teks yang terdapat di media sosial, forum, atau bahkan di buku tamu suatu website. Dengan mengetahui sentimen yang terkandung dalam suatu teks, kita dapat mengetahui bagaimana respon masyarakat terhadap suatu entitas tertentu.

Analisis sentimen memiliki dampak yang besar dan manfaat yang luas, sehingga menyebabkan perkembangan yang cepat dalam penelitian dan aplikasinya. Saat ini, analisis sentimen menjadi salah satu topik penelitian yang

sedang berkembang dengan sangat cepat, bahkan di Amerika lebih dari 20 hingga 30 perusahaan berfokus pada layanan analisis sentiment (Buntoro, 2017). Selain itu, analisis sentimen juga digunakan dalam berbagai bidang, seperti marketing, media, politik, dan teknologi. Ini menunjukkan bahwa analisis sentimen merupakan bidang yang sangat penting dan menjanjikan dalam pengambilan keputusan bisnis dan pemerintahan. Penelitian dan aplikasi analisis sentimen terus berkembang dan diharapkan dapat memberikan solusi yang efektif dalam memahami persepsi masyarakat dan opini publik.

### **2.2.3 *Cyberbullying***

*Cyberbullying* adalah tindakan perundungan yang dilakukan melalui internet, terutama melalui situs atau platform jejaring social (Riswanto & Marsinun, 2020). *Cyberbullying* merupakan tindakan yang disengaja untuk menyakiti atau merendahkan seseorang melalui internet, khususnya melalui situs atau platform jejaring sosial. Dampak dari tindakan ini sangat merugikan bagi korban, baik secara fisik maupun mental. Tindakan ini dapat berupa pembuatan atau penyebaran konten yang tidak pantas, menyebarluaskan informasi pribadi tanpa izin, atau mengirim pesan yang menakut-nakuti atau mengancam. Hal ini menunjukkan bahwa *cyberbullying* merupakan masalah yang serius yang perlu ditangani dengan cepat dan tepat agar dapat mengurangi dampak negatif yang ditimbulkan pada korban.

### **2.2.4 *Support Vector Machine***

*Support Vector Machine* merupakan suatu metode yang dapat menyelesaikan masalah linear menggunakan algoritma pembelajaran yang

didasarkan pada teori optimasi (Puspitasari et al., 2018). *Support Vector Machine* mencoba mengoptimalkan batas pemisah antara dua kelas data dengan memaksimalkan jarak antara data dan batas pemisah tersebut. *Support Vector Machine* dapat digunakan untuk klasifikasi biner maupun multi-kelas. Metode ini biasanya digunakan untuk mengatasi masalah klasifikasi yang memiliki data linier yang tidak terpisah secara jelas. Selain itu, *Support Vector Machine* juga dapat digunakan untuk mengatasi masalah klasifikasi non-linier dengan menggunakan kernel untuk mengubah data ke dalam bentuk yang dapat dipisahkan secara linier.

Metode *Support Vector Machine* memiliki beberapa algoritma yang dapat digunakan, di antaranya adalah C-SVC dan Nu-SVC. Nu-SVC menggunakan parameter Nu untuk mengatur tingkat kesalahan dalam klasifikasi. Nilai Nu membatasi jumlah *sample* yang salah diklasifikasikan dalam set data pelatihan. Sementara itu, Algoritma C-SVC tidak memiliki parameter sejenis dan lebih terfokus pada mengurangi jumlah *sample* yang salah diklasifikasikan (Książek et al., 2019). Nu-SVC digunakan untuk mengatur jumlah *support vectors* yang diinginkan. Kedua implementasi ini dapat digunakan untuk mengoptimalkan akurasi dari hasil klasifikasi yang dilakukan oleh *Support Vector Machine*.

### **2.2.5 Google Colaboratory**

Google Colaboratory, atau yang lebih dikenal sebagai Colab, adalah layanan *cloud* yang didasarkan pada Jupyter Notebook yang digunakan untuk mendukung proses belajar dan riset dalam bidang *machine learning* (Fahmi et al., 2020). Google Colaboratory adalah platform yang digunakan untuk mengeksekusi kode dalam

bahasa pemrograman Python (Alhajir et al., 2021). Platform ini juga menyediakan runtime yang telah dikonfigurasi dengan baik dan dukungan untuk *deep learning* serta akses gratis ke GPU yang kuat, sehingga mempermudah dalam melakukan proses komputasi yang intensif.