BAB II
TINJAUAN PUSTAKA

2.1 Studi Literatur

Diastama et al. [6] dalam penelitiannya membahas mengenai analisis sentimen pada ulasan produk *women's e-commerce* menggunakan pendekatan machine learning dengan algoritma *Naïve Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN). Dataset yang digunakan berasal dari kaggle dengan atribut "*Review Text*" dari berbagai ulasan produk yang diberikan oleh konsumen, baik itu ulasan positif maupun negatif. Proses penelitian meliputi tahap preprocessing (*cleaning*, *lowercase*, *tokenization*, dan *stopword removal*), ekstraksi fitur menggunakan TF-IDF, serta penanganan ketidakseimbangan data dengan teknik *oversampling* seperti *Random Oversampling* (ROS) dan SMOTE. Dari penelitian ini didapatkan bahwa model SVM memberikan performa terbaik dibanding model lainnya dengan akurasi 0.94, dengan skor seimbang pada *precision*, recall, dan F1-*score*. Penggunaan SMOTE meningkatkan kemampuan model dalam mendeteksi kelas minoritas (seperti ulasan negatif), meskipun terkadang menurunkan akurasi keseluruhan. Sementara itu, KNN menunjukkan performa yang kurang stabil, terutama saat menggunakan SMOTE. Secara keseluruhan, penelitian ini menegaskan bahwa penanganan data tidak seimbang sangat berpengaruh terhadap performa klasifikasi, dan kombinasi SVM dengan teknik oversampling menjadi solusi paling efektif dalam kasus ini.

Nurhayati et al. [7] dalam studinya membandingkan performa algoritma *K-Nearest Neighbor*, *Naïve Bayes*, dan *Support Vector Machine* untuk mengklasifikasikan genre film berdasarkan sinopsis menggunakan dataset dari Kaggle dan IMDB dengan teknik klasifikasi teks yang memanfaatkan *natural language processing* (NLP). Penelitian ini dilakukan untuk mengatasi permasalahan ketidakefisienan klasifikasi manual sehingga diperlukan pendekatan otomatis berbasis *machine learning* untuk meningkatkan akurasi dan efisiensi. Proses yang dilakukan pada penelitian ini meliputi tahapan *pre-processing* teks, pembersihan teks, *vectorization* menggunakan *CountVectorizer* dan TF-IDF,

pembagian data menjadi data latih dan data uji, serta evaluasi dengan metrik akurasi, *precision*, *recall*, dan f1-*score*. Tiap algoritma memiliki cara kerja yang berbeda KNN bekerja berdasarkan kedekatan jarak data, Naive Bayes menggunakan pendekatan probabilitas dengan asumsi independensi fitur, sedangkan SVM bekerja dengan mencari hyperplane optimal yang dapat memisahkan kelas data.

Pada percobaan dataset pertama menunjukkan bahwa beberapa model mencapai akurasi diatas 90%, dengan akurasi tertinggi 93% dicapai oleh SVM yang menggunakan TF-IDF. Pada dataset kedua, SVM tetap menjadi algoritma terbaik dengan akurasi 65% menggunakan *count vectorizer* meskipun performanya menurun signifikan disebabkan karakteristik data yang lebih kompleks, noise data, jumlah label yang lebih banyak, serta panjang teks yang bervariasi. Sementara itu, KNN dan *Naïve Bayes* menunjukkan performa yang bervariasi tergantung metode *vectorization* yang digunakan, dengan KNN cenderung lebih baik pada TF-IDF dan *Naïve Bayes* lebih stabil pada *CountVectorizer*. Disimpulkan bahwa SVM terbukti stabil dengan menggunakan dua metode *vectorizer*, lalu penelitian ini juga menunjukkan bahwa karakteristik dataset sangat berpengaruh terhadap hasil klasifikasi sehingga model yang digunakan harus disesuaikan dengan kondisi data. Dengan demikian, penelitian ini menyimpulkan bahwa SVM adalah algoritma yang unggul dalam tugas klasifikasi teks pada genre film berbasis sinopsis.

Arsi et al. [8] menganalisis opini warganet mengenai isu vaksin Covid-19 di platform Instagram akun resmi Kementerian Kesehatan RI (kemenkes_ri) dengan menerapkan teknik *text mining* dan perbandingan berbagai model klasifikasi. Penelitian ini dilakukan karena adanya komentar negatif masyarakat mengenai vaksin Covid-19 meskipun telah didukung berbagai penelitian kesehatan. Penelitian bertujuan menguji dua algoritma klasifikasi, yaitu *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) dalam mengklasifikasi sentimen komentar. Menggunakan 2.925 data komentar yang diperoleh dari instagram melalui metode web *scraping*. Setelah proses pembersihan manual yang

menghilangkan komentar tidak relevan diperoleh 2.034 data yang dapat digunakan. Dataset diklasifikasikan dalam tiga kelas sentimen, yaitu positif, netral, dan negatif.

Penelitian Arsi et al. [8] dimulai dengan tahap preprocessing yang dilakukan menggunakan beberapa teknik NLP seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya dilakukan pembobotan data menggunakan metode TF-IDF untuk merepresentasikan kepentingan atau kemunculan tiap kata dalam sebuah dokumen. Metode SMOTE diterapkan untuk menyeimbangkan data karena terdapat ketimpangan jumlah kelas sentimen. Rasio 90:10 digunakan dalam tahap pembagian data latih dan data uji. Proses klasifikasi dilakukan menggunakan algoritma SVM dengan kernel RBF dan KNN dengan variasi nilai k. Evaluasi performa model dilakukan menggunakan *confusion matrix* dan perhitungan akurasi. Hasil eksperimen menunjukkan bahwa SVM mencapai akurasi tertinggi sebesar 98,30%, sedangkan KNN hanya mencapai 80,03%. Dari penelitian ini disimpulkan SVM memiliki performa yang lebih baik dibandingkan KNN dalam menangani klasifikasi sentimen vaksin Covid-19 di Instagram.

Sabrila et al. [9] membahas analisis sentimen terhadap tweet berbahasa Indonesia mengenai penanganan Covid-19 oleh pemerintah yang menimbulkan perdebatan di media sosial Twitter dengan menerapkan algoritma *Support Vector Machine* (SVM) dan K-*Nearest Neighbor* (KNN). Penggunaan media sosial yang meningkat selama pandemi menjadikan Twitter sebagai sumber data untuk mengetahui opini publik. Penelitian ini bertujuan mengklasifikasikan cuitan warganet menjadi dua kelas sentimen, yaitu positif dan negatif. Menggunakan 801 data tweet berbahasa Indonesia yang didapatkan dari Kaggle dan telah melalui proses pelabelan manual oleh tiga anotator. Dataset melalui tahap preprocessing meliputi *cleaning*, *tokenizing*, *case folding*, *stopword removal*, dan *stemming* menggunakan *library* Sastrawi. Ekstraksi fitur dilakukan menggunakan model *Word Embedding* Word2Vec dengan skip-gram dan 200 dimensi fitur.

Evaluasi model dilakukan menggunakan metode 10-fold *cross validation* dengan pengukuran akurasi, presisi, recall, dan Area Under Curve (AUC). Hasil pengujian menunjukkan SVM lebih unggul dibandingkan KNN. SVM memperoleh nilai akurasi terbaik 85% dengan nilai AUC 0.92 yang termasuk dalam kategori *excellent* model. Sedangkan KNN mendapatkan akurasi 76% dan AUC 0.87 yang termasuk dalam kategori *fair* model. Perbedaan performa ini menunjukkan bahwa SVM lebih mampu memanfaatkan fitur Word2Vec dibandingkan KNN. Sehingga dapat disimpulkan bahwa, Word2Vec yang dikombinasikan dengan SVM terbukti efektif untuk analisis sentimen tweet berbahasa Indonesia. Dengan demikian penelitian ini menyimpulkan kedua metode dapat digunakan namun, SVM memberikan hasil lebih unggul dibandingkan KNN untuk klasifikasi sentimen terkait topik ini.

Ridho et al. [10] melakukan penelitian terkait masalah ketidaktepatan sasaran penerimaan bantuan sosial pemerintah yang disebabkan oleh kurang optimalnya pengolahan data. Untuk mengatasi hal tersebut dilakukan analisis data Survei Sosial Ekonomi Nasional (SUSENAS) menggunakan metode klasifikasi untuk menentukan apakah suatu keluarga layak mendapatkan bantuan. Dengan begitu penelitian ini bertujuan membandingkan kinerja algoritma klasifikasi *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) dalam mengklasifikasikan status penerimaan bantuan pemerintah berdasarkan data SUSENAS. Data tersebut merupakan data sekunder yang diperoleh dari BPS Kabupaten Pekalongan dengan variabel respon berupa status penerimaan bantuan. Dataset menunjukkan kondisi ketidakseimbangan kelas data, jumlah penerima bantuan lebih sedikit dibandingkan jumlah non penerima. Oleh karena itu dilakukan penyeimbangan data menggunakan teknik SMOTE agar kedua kelas data seimbang. Pada proses membangun model, data dibagi menjadi data latih dan data uji dengan rasio 85% dan 15%. Proses klasifikasi dilakukan menggunakan KNN dan SVM ditambah optimasi parameter melalui *grid search* dan *cross validation*.

Hasil klasifikasi dengan algoritma KNN menunjukkan bahwa nilai k terbaik adalah k = 1 dengan tingkat akurasi sebesar 80,95%. Hasil evaluasi performa KNN menunjukkan error rate yang relatif kecil dengan nilai sensitivitas dan presisi yang cukup tinggi dalam mengidentifikasi kelas penerima bantuan. Sementara itu, pada algoritma SVM dilakukan pengujian menggunakan kernel polynomial dan Radial Basis Function (RBF), dimana kernel RBF menghasilkan kinerja terbaik. Hasil akurasi dengan kernel RBF pada metode SVM sebesar 78%, dengan tingkat error rate yang lebih besar dibandingkan KNN. Hasil evaluasi menunjukkan SVM lebih unggul pada nilai *specificity* dan presisi, namun kemampuan KNN dalam mengenali kelas positif menunjukkan hasil yang lebih baik. Oleh karena itu, secara keseluruhan berdasarkan keseluruhan evaluasi kinerja yang menunjukkan KNN lebih unggul pada beberapa metrik utama dapat disimpulkan bahwa metode KNN lebih optimal dalam klasifikasi status penerimaan bantuan pemerintah.

Ramadhan et al. [11] dalam penelitiannya membahas analisis sentimen pada ulasan film dari website IMDB berdasarkan aspek rating bintang dengan metode *Support Vector Machine* (SVM). Penelitian ini didasari karena banyaknya komentar film yang beragam sehingga menyulitkan pengguna dalam memahami opini secara keseluruhan. Oleh karena itu analisis sentimen digunakan untuk mengklasifikasikan opini dalam kelas positif atau negatif. Dataset diambil dengan teknik random sampling dari IMDB khususnya pada ulasan film Squid Game. penelitian dimulai dari preprocessing hingga klasifikasi. Tahap *preprocessing* meliputi tokenisasi, penghapusan karakter non-alfabet, *stopword removal*, dan *stemming*. Selanjutnya dilakukan ekstraksi fitur menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik. Label sentimen ditentukan berdasarkan rating, yaitu positif jika rating lebih dari 5 dan negatif jika kurang dari 5.

Pada tahap klasifikasi algoritma yang digunakan adalah SVM dengan kernel linear dan pembagian data 70:30. Hasil evaluasi menunjukkan bahwa model SVM menghasilkan akurasi sebesar 79%, precision 75%, dan recall 87%.

Hasil tersebut menunjukkan bahwa SVM memiliki performa yang cukup baik dalam klasifikasi teks, terutama karena kemampuannya dalam menangani dataset berdimensi tinggi. Selain itu, SVM juga dibandingkan dengan *Logistic Regression* dan terbukti lebih unggul dalam semua metrik evaluasi. Namun, nilai *precision* lebih rendah dibanding recall yang menunjukkan masih terdapat kesalahan dalam prediksi positif. Hal ini mengindikasikan bahwa model masih perlu perbaikan untuk meningkatkan ketepatan klasifikasi. Secara keseluruhan, penelitian ini menyimpulkan bahwa SVM dengan TF-IDF efektif digunakan untuk analisis sentimen ulasan film berbasis IMDB.

Sitepu et al. [12] dalam penelitiannya menganalisis masalah perundungan yang marak terjadi di sosial media seiring dengan berkembangnya teknologi. Kebebasan berekspresi yang tidak diawasi di sosial media menimbulkan munculnya komentar negatif yang berdampak buruk bagi pengguna. Oleh karena itu penelitian ini bertujuan membangun dan menganalisis model otomatis yang dapat mengidentifikasi serta menyaring komentar yang mengandung unsur perundungan. Penelitian ini menerapkan pendekatan *text mining*, untuk mengklasifikasikan komentar pada media sosial. Untuk proses pengklasifikasian teksnya dimanfaatkan algoritma Support Vector Machine (SVM) yang dipilih karena keunggulannya yaitu memiliki kinerja yang baik dalam klasifikasi teks. Dataset penelitian yang digunakan bersumber dari komentar media sosial Twitter pada pemilu 2019 yang banyak mengandung ujaran negatif. Dari komentar tersebut didapatkan sebanyak 1000 komentar yang selanjutnya diberikan label secara manual dalam dua kelas, yaitu bully dan tidak bully.

Sebelum proses klasifikasi dilakukan, data melalui tahapan preprocessing seperti case folding, cleaning, stemming, filtering, dan pembobotan TF-IDF. Penelitian menggunakan pembagian data 60% untuk data latih dan 40% untuk data uji. Kinerja algoritma diuji menggunakan confusion matrix untuk mengukur nilai akurasi, presisi, dan recall. Hasil klasifikasi menunjukkan bahwa SVM menghasilkan akurasi 87,5% pada data uji. Selain itu, nilai precision mencapai 89% dan recall 91%, yang menunjukkan performa model cukup baik dan akurat

dalam mengenali komentar bullying. Dari hasil evaluasi tersebut disimpulkan bahwa SVM efektif digunakan dalam klasifikasi komentar perundungan di media sosial.

Tabel 2.1 Keaslian Penelitian

No	Judul penelitian	Nama Penulis	Tahun Publikasi	Hasil Penelitian	Perbandingan Penelitian
1	Sentiment Analysis Classification In Women's E-Commerce Reviews With	Alfiki Diastama Afan Firdaus, Rizki Dwi Rahmawan, Yuzzar Rizky Mahendra, Hasan Dwi Cahyono	2024	Dari algoritma NB, SVM, dan KNN diperoleh bahwa SVM memberikan performa terbaik dengan akurasi 0.94 serta nilai precision, recall, dan F1-score yang seimbang.	Data diproses menggunakan text encoders TF-IDF dan peanganan ketidak simbangan data menggunakan ROS & SMOTE
2	Comparative Analysis of KNN, Naïve Bayes and SVM Algorithms for Movie Genres Classification Based on Synopsis.	Nurhayati, Lee Kyung Oh, Muhammad Hugo Athallah Hardy, Yusuf Wijaya	2022	SVM memberikan akurasi tertinggi dibanding KNN dan NB. Dengan akurasi sebesar 90,23% menggunakan CountVectorizer dan 93,05% menggunakan TF-IDF Vectorizer pada dataset pertama dan 65,26% menggunakan Count Vectorizer dan 62,94% menggunakan TF-IDF Vectorizer pada dataset kedua.	Menggunakan data genre film. Representasi text menggunakan TF-IDF.

No	Judul penelitian	Nama Penulis	Tahun Publikasi	Hasil Penelitian	Perbandingan Penelitian
3	Komparasi Model Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform Instagram	Primandani Arsi, Laili Nur Hidayati, Azizan Nurhakim	2022	Hasil penelitian menunjukan bahwa hasil klasifikasi terbaik ditunjukkan pada metode SVM yakni 98,30% dengan pembobotan TF-IDF, SMOTE, dan Kernel RBF, sementara dari metode KNN menghasilkan akurasi 80,03%.	Analisis sentimen yang mengangkat topik vaksin covid. Menggunakan representasi text TF-IDF
4	Analisis Sentimen Pada Tweet Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K-Nearest Neighbor	Trifebi Shina Sabrila, Veronica Retno Sari, Agus Eko Minarno	2021	Hasil penelitian menunjukkan bahwa SVM memiliki kinerja yang lebih baik dibandingkan KNN. Dengan kombinasi fitur Word2Vec, SVM berhasil mencapai akurasi tertinggi sebesar 85% dan nilai AUC 0,92 yang termasuk dalam kategori excellent model. Sementara itu, KNN mencapai akurasi 76% dengan nilai AUC 0,87 yang dikategorikan sebagai fair model.	Menggunakan Word2Ved sebagai ekstraksi fitur.

No	Judul penelitian	Nama Penulis	Tahun Publikasi	Hasil Penelitian	Perbandingan Penelitian
5	Perbandingan Metode K-Nearest Neighbor Dan Support Vector Machines Pada Status Penerimaan Bantuan Dari Pemerintah	Wahyu Anwar Ridho, Arief Rachman Hakim, Triastuti Wuryandari	2024	Algoritma K-Nearest Neighbor (KNN) maupun Support Vector Machine (SVM) mampu digunakan untuk mengklasifikasikan status penerimaan bantuan sosial, namun KNN memberikan kinerja yang lebih optimal secara keseluruhan. KNN dengan nilai tetangga terbaik k = 1 menghasilkan tingkat akurasi tertinggi sebesar 80,95%. Sedangkan, SVM dengan kernel Radial Basis Function (RBF) memberikan performa terbaik di antara kernel yang diuji, dengan akurasi sebesar 78%.	KNN menggunakan kernel euclidean tetapi bukan untuk data text. Menerapkan SMOTE untuk penanganan ketidakseimbangan kelas.

No	Judul penelitian	Nama Penulis	Tahun Publikasi	Hasil Penelitian	Perbandingan Penelitian
6	Analysis Sentiment based on IMDB aspects from movie reviews using SVM Metode K-Nearest Neighbor	Nur Ghaniaviyanto Ramadhan, Teguh Ikhlas Ramadhan	2022	Hasil penelitian menunjukkan SVM mencapai akurasi 79% dan lebih unggul dibanding Logistic Regression, meskipun precision lebih rendah dari recall sehingga masih terdapat kesalahan dalam prediksi positif.	Menggunakan TF-IDF dan perbedaan dataset yang digunakan
7	Analisis Kinerja Support Vector Machine dalam Mengidentifikasi Komentar Perundungan pada Jejaring Sosial	Ade Clinton Sitepu, Wanayumini, Zakarias Situmorang	2021	Hasil penelitian menunjukkan Algoritma SVM mampu melakukan klasifikasi dengan tingkat keakuratan 87,75%, presisi 89% dan Recall 91%	Menggunakan data komentar twitter pada saat pemilu tahun 2019 dan TF-IDF sebagai representasi textnya.

2.2 Dasar Teori

2.2.1 Pelecehan Seksual Verbal

Pelecehan seksual dapat didefinisikan sebagai perilaku maupun tindakan berkonotasi seksual yang mengganggu dan menjengkelkan yang dilakukan sepihak dan tidak dikehendaki oleh korbannya, bentuknya dapat berupa ucapan, tulisan, simbol, isyarat, dan tindakan berkonotasi seksual [14]. Menurut D. Khoirunisa [15], Pelecehan seksual verbal merupakan salah satu bentuk tindak pidana asusila yang kini semakin marak terjadi di ruang digital, seiring berkembangnya teknologi informasi. Jika dulu pelecehan seksual lebih banyak terjadi secara langsung di dunia nyata, kini media sosial menjadi salah satu tempat yang umum bagi pelecehan jenis ini, dengan berbagai bentuk dan modus, seperti komentar tidak pantas, pesan berbau seksual, hingga ujaran yang melecehkan secara verbal. Fenomena ini termasuk dalam kejahatan siber (*cybercrime*) karena dilakukan secara daring dan berdampak luas, baik pada laki-laki maupun perempuan meskipun perempuan lebih sering menjadi korban. Pada konteks penelitian ini dampak yang ditimbulkan antara lain :

a. Bagi publik figur sebagai korban :
   1) Tekanan psikologis seperti *anxiety* (kecemasan sosial), rasa tidak aman dan nyaman ketika tampil di publik.
   2) Objektifikasi yang terus-menerus karena candaan seksual dianggap normal, dan
   3) Ketakutan dalam berinteraksi dengan penggemar sehingga publik figur menghindari konten tertentu juga lebih menjaga jarak.

b. Bagi pengguna media sosial lainnya :
   1) Menormalisasi budaya objektifikasi seperti bebas mengomentari tubuh maupun seksualitas siapapun dan tidak bisa membedakan apresiasi dengan objektifikasi sehingga membentuk budaya penggemar *toxic*.

2) Menurunnya batas etika yang menyebabkan ruang digital menjadi tidak aman
3) Efek pelecehan verbal ini menyebabkan perilaku menyebar ketika satu akun membuat candaan seksual dan pengguna lain melihatnya sebagai sesuatu yang normal karena mendapat perhatian. Lama-kelamaan komentar dapat menjadi kebiasaan sehingga pelecehan dianggap sebagai identitas komunitas.
4) Konflik internal karena sebagian penggemar membela publik figur sementara yang lain menganggap pelecehan sebagai "humor".

2.2.2 TikTok

Aplikasi populer melalui penggunaan video pendek beberapa tahun terakhir aplikasi buatan Negeri Tiongkok yang diluncurkan tahun 2016 oleh Zhang Yiming. Tiktok menjadi sarana pengguna mengekspresikan diri, penyebaran informasi, dan sebagai media hiburan. Dari laman salah satu platform pengunduh aplikasi yaitu Google Play Store, tiktok sendiri sudah di unduh oleh lebih dari 100 juta pengguna dengan rating 4.0 dari rating maximal 5. Penggunanya terus meningkat setiap tahun yang terdiri dari berbagai kalangan terutama generasi muda, Tiktok sudah menjadi rutinitas sehari-hari sehingga mengubah kehidupan penggunanya.

Tiktok memberikan banyak kebebasan kepada penggunanya sehingga dapat berkreativitas dengan video pendek. Namun seiring populernya aplikasi Tiktok , perlu diwaspadai juga dampaknya. Banyaknya kalangan muda yang aktif terutama penggemar Kpop sering terpapar berbagai jenis konten dan melihat interaksi di kolom komentar yang tidak jarang berisi komentar-komentar tidak pantas yang seperti ujaran kebencian atau pelecehan verbal. Oleh karena itu penting bagi media sosial seperti tiktok untuk melakukan moderasi komentar yang dapat memfilter komentar yang melanggar atau berkonotasi negatif.

2.2.3 Machine Learning

Machine learning adalah salah satu cabang ilmu komputer yang mempelajari bagaimana pengembangan pola dan teori pembelajaran pada komputer dalam kecerdasan buatan [16]. Bidang ini berfokus mempelajari struktur dan analisis algoritma pembelajaran dan prediksi data. Dengan berbagai teknik pemodelan machine learning membangun model berdasarkan data input untuk mendapatkan prediksi atau keputusan yang akurat berdasarkan data yang dipelajari.

Terdapat tiga metode *machine learning* yaitu *supervised learning* (pembelajaran terbimbing), *unsupervised learning* (pembelajaran tidak terbimbing), serta *reinforcement learning* dan metode yang digunakan dalam penelitian ini adalah *supervised learning*. *Supervised* learning merupakan pembelajaran terarah dan terawasi yang membangun model berdasarkan data input yang diberikan untuk memprediksi hasil tertentu setelah mempelajari pola umum yang dapat mengubah data input menjadi output [16]. Proses tersebut terus dilakukan hingga model mencapai tingkat akurasi yang baik pada data pelatihan. Metode *supervised learning* dapat dimanfaatkan untuk proses klasifikasi dan regresi (proses prediksi nilai kontinu). Contoh penerapan klasifikasi adalah deteksi penyakit, analisis sentimen, deteksi penipuan, dan lain-lain. Sedangkan contoh penerapan regresi adalah prediksi harga rumah, peramalan cuaca, prediksi harga saham, dan lain-lain.

2.2.4 *Natural Language Processing* (NLP)

Bidang dalam ilmu komputer dan linguistik yang merupakan cabang *artificial intelligence* yang memfokuskan pada interaksi antara komputer dan bahasa manusia, memungkinkan mesin memahami, menganalisis, memanipulasi, dan merespon bahasa alami manusia [17]. Dalam konteks ini, NLP digunakan untuk memproses komentar teks dari pengguna TikTok dan mengidentifikasi pola linguistik yang mengandung

unsur pelecehan. Komponen umum dalam NLP meliputi pemrosesan teks (termasuk tokenisasi dan analisis semantik), pemahaman konteks, generasi bahasa, serta penerjemahan antarbahasa. NLP memiliki beberapa teknik dasar antara lain :

a. Tokenisasi

Proses memecah teks menjadi token yaitu unit-unit kecil seperti kata-kata atau frasa. Setiap token dianggap penting karena mewakili tiap bagian teks yang nantinya berguna dalam proses analisis dan pemahaman teks oleh mesin. Untuk mengubah teks menjadi token dapat menggunakan *WordPunctTokenizer* dalam Python.

b. Stopword Removal

Proses menghilangkan kata-kata yang kurang bermakna yang umum seperti "di", "dan", "atau", "yang", dsb. Kata-kata tersebut sering muncul dalam teks tetapi tidak memberikan informasi khusus dalam pemrosesan atau analisis teks. Proses ini penting dilakukan untuk meningkatkan kualitas analisis dengan menghilangkan kata-kata umum, mengurangi *noise*, dan pembersihan teks sehingga model yang dirancang menjadi lebih baik.

c. Stemming dan *Lemmatization*

Proses mengubah kata menjadi bentuk dasar. *Stemming* bertujuan untuk menghilangkan imbuhan serta akhiran kata, sedangkan *lemmatization* mengubah kata menjadi bentuk baku berdasarkan kamus atau aturan linguistik.

d. Word *Embeddings*

Teknik yang digunakan untuk mempresentasikan kata-kata menjadi vektor numerik sehingga memungkinkan mesin memahami makna kata dalam NLP. Hal ini bermanfaat dalam pemrosesan bahasa, analisis sentimen, kesamaan kata, dan lain-lain. Teknik yang umum digunakan

dalam *word* embeddings antara lain : *Glove*, *Word2Vec*, *Continuous Bag-of-Words*, *Skip-gram*.

2.2.5 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang umum digunakan untuk klasifikasi teks karena kemampuannya menangani data berdimensi tinggi dan memaksimalkan margin pemisah antar kelas [18]. SVM bekerja dengan mencari *hyperplane* yang memisahkan data ke dalam dua buah kelas yang berbeda yaitu +1 dan -1 serta berbagai alternatif *discrimination boundaries* atau garis pemisah. Pemisahan dilakukan demikian, sehingga jarak margin antara *hyperplane* dan titik-titik data (*support vector*) dari kedua kelas maksimum [17]. Secara garis besar cara kerja SVM dalam klasifikasi teks yang pertama adalah pra-pemrosesan teks seperti *tokenisasi*, penghapusan *stopwords*, dan vektorisasi fitur yang dalam penelitian ini menggunakan LaBSE.

Selanjutnya pemilihan *hyperplane* terbaik dan yang memiliki margin terbesar yaitu jarak terbesar antara *support vektor* (titik-titk data) dan hyperplane tersebut. Terakhir setelah mendapat *hyperplane* terbaik SVM akan mengklasifikasikan data baru kedalam kelas berdasarkan posisi terhadap garis pembatas. Dalam penelitian ini digunakan kernel *Radial Basis Function* (RBF), yang memungkinkan data dipetakan ke dalam ruang berdimensi lebih tinggi sehingga hubungan *non-linear* antar data dapat dimodelkan dengan lebih baik. Dengan demikian SVM akan mencari garis batas pemisah non-linear yang lebih sesuai dengan pola data. Secara matematis rumus RBF SVM ditulis sebagai berikut :

$$K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2) \quad (2\text{-}0)$$

Dimana :

a. x_i, x_j adalah vektor embedding

b. $K(x_i, x_j)$ adalah nilai kemiripan antara dua data

c. γ (gamma) adalah parameter sensitivitas jarak

Kernel RBF menghitung tingkat kemiripan antar vektor komentar berdasarkan jaraknya. Kemudian dengan fungsi eksponensial jarak tersebut diubah menjadi nilai kemiripan antara nol dan satu, komentar dengan jarak lebih dekat memiliki nilai kemiripan yang lebih tinggi. Parameter γ (*gamma*) mengatur sejauh mana pengaruh satu data terhadap data lainnya, kernel ini memungkinkan model menangkap pola kompleks data komentar yang sering disampaikan secara implisit.

2.2.6 *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) dapat dikatakan sebagai algoritma machine learning yang paling sederhana dan merupakan salah satu metode algoritma klasifikasi *supervised learning* yang bekerja dengan mengelompokkan data ke kategori tertentu sesuai tingkat kemiripan mayoritas yang dihitung berdasarkan jarak (*distance*) terdekat dari data latih dan jumlah tetangga atau disebut K [16]. Nilai K pada algoritma KNN sangat berpengaruh terhadap hasil klasifikasi dan nilai k *sebaiknya* ganjil agar mencegah terjadinya hasil seri [18]. Saat ada data baru yang akan diprediksi algoritma akan terlebih dahulu menentukan nilai K sebagai acuan, setelah itu jarak antara data baru dan semua data dalam dataset dihitung.

Untuk menghitung jarak antar titik ada beberapa metrik yang umum digunakan pada algoritma KNN yaitu *euclidean distance*, *manhattan distance*, *minkowski distance*, dan *hamming distance*, namun pada penelitian ini akan *digunakan cosine similarity* yang dikonversi menjadi *cosine distance* sebagai metrik kemiripan. Pada klasifikasi hasil diputuskan berdasarkan mayoritas kelas dari K tetangga. Secara umum rumus jarak *cosine distance* yang digunakan adalah sebagai berikut :

$$cosine\ distance(x,y) = 1 - \frac{x \cdot y}{||x||\ ||y||} \quad (2\text{-}1)$$

a. x = *vector data uji*

b. y = *vector data latih*

c. $x \cdot y$ = *hasil perkalian dot product vector x dan y*

d. $||x||\ ||y||$ = *panjang vector*

Rumus tersebut mengukur seberapa besar perbedaan arah kalimat, sehingga semakin kecil nilai *cosine distance*, maka semakin tinggi pula kemiripan antara dua data. Dalam konteks penelitian ini, KNN digunakan untuk *mengenali* pola kemiripan antara komentar yang dianalisis dengan komentar lain yang sudah memiliki label. Dengan menghitung kedekatan semantik antar komentar, KNN dapat menentukan apakah sebuah komentar termasuk kategori pelecehan seksual atau tidak.

2.2.7 LaBSE

Labse (*Language-agnostic BERT Sentence Embedding*) salah satu model *embedding* kalimat yang dirancang untuk merepresentasikan kalimat multibahasa menjadi vektor sehingga maknanya dapat dipahami oleh komputer. Adapun keunggulan LaBSE ini yaitu mampu memahami 109 bahasa dan dengan kombinasi *pre-training* serta translation ranking loss, model ini dapat mengenali dua kalimat yang sama meskipun dalam bahasa yang berbeda, sehingga *embedding* akan saling berdekatan di ruang vektor [21]. LaBSE dilatih dengan 17 miliar kalimat monolingual dan 6 miliar kalimat *bilingual* sehingga model tampak efektif dengan bahasa yang memiliki sumber daya terbatas yang tidak memiliki data latih.

LaBSE bekerja dengan mempelajari teks *monolingual* melalui *masked language modeling* (MLM) dan belajar memahami hubungan antar bahasa dengan membaca kalimat terjemahan melalui *translation language modeling* (TLM). Selanjutnya LaBSE disempurnakan dengan

memanfaatkan *translation ranking* yang dapat membuat model dapat mengenali terjemahan yang paling tepat. Hasilnya model menghasilkan embedding yang menempatkan kalimat antarbahasa dengan makna yang sama di posisi berdekatan dalam ruang vektor. Model LaBSE sangat relevan karena mampu merepresentasikan komentar-komentar yang bercampur menjadi representasi yang lebih rapi dan bermakna yang kemudian dapat digunakan sebagai input model agar proses klasifikasi lebih akurat.

2.2.8 Confusion Matrix

Dalam *machine learning* diperlukan alat evaluasi untuk memahami kinerja model yang disebut dengan *confusion matrix* [22]. Hasil evaluasi klasifikasi ditampilkan dalam bentuk tabel yang menunjukkan jumlah prediksi benar dan salah untuk tiap kelas . *Confusion matrix* terdiri dari empat komponen yaitu:

a. True Positive (TP) : jumlah data yang benar diprediksi positif
b. True Negative (TN) : jumlah data yang benar diprediksi negatif
c. False Positive (FP) : jumlah data negatif tetapi diprediksi positif
d. False Negative (FN) : jumlah data positif namun diprediksi negatif

Komponen-komponen tersebut selanjutnya digunakan untuk mengevaluasi kemampuan model machine *learning* dalam melakukan prediksi maupun klasifikasi menggunakan metrik kinerja model. Berikut rumus metrik kinerja yang digunakan dalam klasifikasi :

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+F)} \quad (2\text{-}2)$$

Accuracy menunjukkan seberapa sering model memprediksi dengan benar dibandingkan dengan keseluruhan jumlah data.

$$Precision = \frac{TP}{(TP+FP)} \quad (2\text{-}3)$$

Precision menunjukkan ketepatan prediksi model positif. Ketika digunakan untuk mendeteksi komentar pelecehan dan nilai precision tinggi berarti komentar yang diprediksi benar-benar pelecehan bukan komentar netral yang salah prediksi menjadi pelecehan.

$$Recall = \frac{TP}{(TP+FN)} \quad (2\text{-}4)$$

Recall mengukur seberapa baik model dapat menemukan seluruh data positif sebenarnya. Yang berarti ketika nilai *recall* tinggi model berhasil mendeteksi banyak komentar pelecehan tanpa melewatkan yang seharusnya terdeteksi.

$$F1 - Score = 2(\frac{Recall \times Precision}{Recall + Precision}) \quad (2\text{-}5)$$

F1-*Score* adalah gabungan precision dan *recall*, ketika skor tinggi berarti model seimbang dalam mendeteksi komentar pelecehan dan minimal kesalahan prediksi.