

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Apple Vision Pro**

Apple Vision Pro adalah gadget revolusioner yang mengintegrasikan teknologi canggih untuk menghadirkan pengalaman komputasi spasial yang inovatif. Perangkat ini memungkinkan integrasi tanpa batas antara media digital dan dunia fisik, memungkinkan pengguna bernavigasi menggunakan mata, tangan, dan suara. Perangkat ini menggabungkan kamera dan sensor canggih untuk memungkinkan pengenalan visual, lingkungan, dan tangan yang akurat. Pita audio dan speaker memberikan pengalaman "audio spasial" premium yang menyatu dengan lingkungan sekitar [4].

Apple memperkenalkan headset realitas virtualnya dengan iklan Apple Vision Pro bertajuk "Memperkenalkan Apple Vision Pro" di platform YouTube pada 6 Juni 2023, yang memperoleh total 55.528.858 penayangan. Hal ini menunjukkan bahwa minat terhadap kemajuan teknologi berbasis kecerdasan buatan dengan konsep komputasi spasial mulai menarik perhatian masyarakat. Dalam video promosi berdurasi 9 menit 20 detik, Apple memperkenalkan produknya secara serius dan mendalam kepada pemirsa [1].



Gambar 2.1. Apple Vision Pro

Apple telah melengkapi Vision Pro dengan layar mikro-OLED dengan kepadatan piksel yang luar biasa. Ukuran layar perangkat ini 1,4 inci per mata dengan total 23 juta piksel. Layarnya menawarkan resolusi 4K yang menakjubkan untuk semua mata, membenamkan pengguna dalam lingkungan visual yang sangat detail dan realistis. Sistem kamera yang menggunakan lampu sorot inframerah,

pemindai LiDAR, dan sensor TrueDepth juga memungkinkan pelacakan kepala dan tangan secara akurat serta pemetaan 3D waktu nyata sekaligus mendeteksi gerakan tangan dari berbagai posisi [3]. Perangkat memproyeksikan output ke mata pengguna melalui dua layar kecil beresolusi tinggi. Antarmuka pengguna utama didasarkan pada pelacakan mata, tangan, gerakan, kamera, dan sensor sehingga tidak memerlukan pengontrol fisik seperti keyboard atau layar sentuh [4].

Potensi integrasi teknologi layar Vision Pro Apple yang canggih dengan augmented reality merupakan cara yang menjanjikan untuk memulihkan penglihatan. Augmented reality telah berhasil memulihkan penglihatan dengan mengubah teks, memperluas bidang pandang, dan mengurangi ketajaman visual [8]. Apple Vision Pro dilengkapi dengan rangkaian kompleks yang terdiri dari 12 kamera dan lima sensor yang menyediakan pelacakan ekspresi wajah yang mendetail dan kerangka kerja mendetail untuk pengenalan emosi. Kerangka penginderaan yang kuat ini tidak hanya memastikan pengenalan wajah dan transmisi ekspresi yang akurat dan personal, namun juga mewakili kemajuan signifikan dalam pengenalan emosi wajah untuk VR [21]. Uniknya, tidak seperti teknologi simulasi wajah yang umum digunakan pada perangkat ternama lainnya seperti Meta dan Vivo, Apple Vision Pro menggunakan pemindaian wajah individu setiap pengguna [22].

## **2.2 X (Twitter)**

X adalah jaringan sosial yang memungkinkan pengguna untuk berbagi cerita teks hingga 280 karakter, serta berkomentar dan me-retweet (dengan membagikan kiriman pengguna lain). Twitter telah terbukti menjadi salah satu situs mikroblog terpadat di dunia, dengan ratusan juta pengguna, dan merupakan platform luar biasa untuk menyebarkan informasi secara real-time [23]. Twitter adalah SNS yang banyak digunakan yang memungkinkan orang untuk bertukar pesan tertulis (disebut "tweet"). Lebih dari 300 juta orang terdaftar di Twitter, dan lebih dari 500 juta pembaruan dihasilkan setiap hari. Kemudahan berbagi di Twitter menjadikannya salah satu sumber data yang penting bagi pengguna [24].



Gambar 2. 2 Logo baru dan lama X (Twitter)

Seorang pengguna harus terdaftar pada platform untuk mengirim tweet. Pengguna memilih nama samaran (nama pengguna) selama pendaftaran, yang kemudian akan digunakan untuk mengirim pesan. Sebutan dalam tweet menunjukkan bahwa postingan tersebut menyebut pengguna lain ditulis sebagai "@namapengguna" dan simbol @ digunakan untuk merujuk pada orang atau organisasi yang terkait dengan Tweet tersebut. Tagar adalah jenis tag metadata yang digunakan oleh berbagai situs jejaring sosial yang menggunakan tag dinamis yang dibuat oleh pengguna untuk membantu orang lain menemukan tweet tentang topik tertentu [25].

Follow adalah aktivitas di mana pengguna terdaftar dapat mengikuti orang, perusahaan, atau orang yang mereka minati dan menerima tweet terbaru secara real time. Twitter lebih dari sekedar alat untuk tetap terhubung dengan teman dan berbagi aktivitas sehari-hari Anda; kekuatan sebenarnya terletak pada menyebarkan berita dan mengikuti orang lain [26]. Pencarian adalah fitur canggih yang memungkinkan pengguna mencari kata kunci dan frasa di Twitter dan menemukan Tweet terbaru yang terkait dengan minat pengguna secara real-time. Fitur pencarian ini memudahkan untuk menemukan dan mendistribusikan konten yang relevan, sehingga orang lebih cenderung berinteraksi dengan Twitter [27]. Retweet adalah ketika sebuah tweet dibagikan lagi. Jika pengguna menganggap suatu tweet menarik, mereka dapat memposting ulang menggunakan fitur retweet. Retweet dianggap sebagai alat yang efektif untuk menyebarkan informasi. Tweet yang Dibagikan tidak berubah dan biasanya menyertakan singkatan RT diikuti dengan nama pengguna pengirim (RT @namapengguna). Retweet juga dapat menyertakan komentar singkat. Twitter memberikan opsi kepada pengguna untuk memutuskan

apakah tweet-nya akan terlihat kepada semua orang atau hanya kepada pengikut Twitter yang telah disetujui[11].

Situs jejaring sosial (SNS) yaitu Twitter, memungkinkan data pengguna diunduh dengan beberapa izin terbatas melalui antarmuka pemrograman aplikasi “*Application Programming Interfaces*” (API). API merupakan perpanjangan dari perangkat lunak yang bertindak sebagai antarmuka *back end* untuk mengumpulkan data dari setiap SNS. Beberapa SNS memberikan wawasan tentang pengalaman pengguna melalui API. Data pengguna yang tersedia untuk umum atau melalui pengunduhan dengan bantuan API dapat ditelusuri secara digital [28]. Twitter dilaporkan menjadi subjek sebagian besar penelitian yang berpusat pada API karena fleksibilitasnya, yang memberikan kebebasan dalam mengakses database. Namun, Twitter juga memperketat pembatasannya seiring berjalannya waktu [26].

Mengevaluasi percakapan Twitter telah muncul sebagai bidang studi yang menjanjikan. Percakapan memberikan informasi yang kaya dan berbeda terkait berbagai topik, sehingga memudahkan untuk memahami emosi orang lain. Sentimen optimis dan pesimis yang diungkapkan dalam percakapan Twitter dianalisis menggunakan pendekatan pembelajaran mendalam yang baru [29]. Dalam beberapa tahun terakhir, para peneliti semakin tertarik untuk menganalisis tweet berdasarkan sentimen yang diungkapkan oleh pengguna Twitter. Ketertarikan ini muncul dari banyaknya tweet yang diposting di Twitter yang memberikan informasi penting mengenai sentimen publik terhadap berbagai topik [27].

Setiap pengguna memiliki informasi pribadi (nama, lokasi, biografi). Karakteristik tersebut telah meningkatkan minat komunitas data mining dalam menganalisis tweet. Sebagai bagian dari data mining, mining teks berdasarkan analisis tweet adalah salah satu tugas dasar yang dapat dipertimbangkan terkait dengan data Twitter [30]. Perusahaan besar dapat memanfaatkan sosial media Twitter untuk menggunakan umpan balik dari pelanggan untuk menganalisis variabel mana yang memengaruhi persepsi nilai produk oleh pembeli. Kepuasan atau ketidakpuasan produk secara keseluruhan dapat menjadi mekanisme yang layak untuk menganalisis umpan balik yang dikumpulkan [31].

### 2.3 Crawling Data X

Crawling adalah metode mengambil konten informasi di sebuah situs web. Teknik pengumpulan informasi dapat mengumpulkan seluruh data atau data tertentu. Awalnya, metode crawler digunakan oleh browser web untuk membuat indeks. Data dikumpulkan secara otomatis melalui pengambilan informasi menggunakan teknologi crawler. Crawler yang dibutuhkan menyesuaikan dengan spesifikasi media sosial masing-masing [32]. Tujuan web crawling adalah untuk membuat kumpulan halaman web yang berguna secepat dan seefisien mungkin, dengan tetap menghormati batasan yang diberlakukan oleh penerbit, menghindari jebakan, dan setidaknya menyediakan fungsionalitas yang diperlukan dalam mengumpulkan informasi [33].

Web crawling memainkan peran yang berbeda dari web scraping, namun keduanya digunakan di situs web. Tugas web crawling adalah beralih secara otomatis antar situs web saat web scraping. Crawling adalah pembuatan salinan dari bagian-bagian yang relevan di World Wide Web dan web crawling adalah langkah terpenting dalam menemukan informasi web [34]. Kehati-hatian khusus harus diberikan saat membangun web crawler agar tidak menemui jalan buntu atau memakan waktu yang sangat lama. Hal yang perlu diperhatikan adalah mengetahui situs web mana yang pernah dikunjungi sebelumnya, data apa yang ingin diakses, dan bagaimana data tersebut disimpan [35].

Proses crawler mengunjungi suatu website dimulai dengan mengumpulkan data dan informasi tentang seluruh URL website, melacaknya satu per satu dan menambahkannya ke daftar halaman di indeks mesin pencari. Jika melakukan perubahan pada situs web, maka akan diperbarui secara otomatis. Aplikasi web crawler juga melakukan proses mengambil dan mengumpulkan hyperlink pada situs web yang dikunjungi dan menyimpannya di disk besar. Untuk memudahkan kegiatan pencarian oleh mesin pencari, dilakukan pengindeksan data berupa daftar hyperlink [36].

Web crawler biasanya digunakan untuk mengambil data dari situs web secara otomatis dan memproses data tersebut sesuai dengan kebutuhan yang diperlukan. Misalnya saja web crawler yang digunakan untuk mengambil data

review dari aplikasi di Play Store dan menggunakannya sebagai bahan data mining. Web crawler juga digunakan untuk mengumpulkan tweet dari Twitter dan memproses data terlebih dahulu menjadi teks untuk tujuan kecerdasan buatan. Selain itu, web crawler juga digunakan untuk mengindeks situs web di mesin pencari dan menghasilkan dokumen teks yang terkait dengan domain atau situs web tertentu [37].

Di antara teknologi komputasi ini, seluruh kategori bergantung pada robot berkode untuk merayapi Internet (bot crawler) dan mengumpulkan data (bot scraper). Alat-alat ini sangat berguna dalam mengambil data yang relevan secara sistematis dan otomatis serta membantu dalam membangun database yang terstruktur [38]. API atau biasa disebut dengan Application Programming Interface adalah program/aplikasi yang disediakan oleh pengembang untuk memudahkan pihak tertentu mengakses atau mengambil dan mengolah data dari Twitter. API membantu untuk memperoleh informasi akun dan melihat sebutan, retweet, favorit, pengikut, dan teman untuk akun Twitter tertentu [39].

Twitter memberi peneliti dan pengembang cara mudah untuk mengakses dan mengumpulkan data melalui dua API Twitter: REST dan Streaming. REST API menyediakan konektivitas jangka pendek terbatas dan memungkinkan Anda meminta dan mengunduh tweet dalam jumlah berapa pun. Akses REST API terbatas pada tweet yang berumur kurang dari seminggu, karena Twitter tidak menyediakan akses ke tweet yang berumur lebih dari seminggu. API Streaming, di sisi lain, mendukung koneksi jangka panjang melalui berbagai titik akhir HTTP dan mengunduh data hampir secara real-time. Artinya API ini dapat menerima tweet terbaru yang berisi istilah tertentu, atau tweet yang diposting oleh pengguna tertentu [11].

Beberapa alat berbasis Python dapat mengumpulkan data dari server Twitter. Perpustakaan bahasa pemrograman Python mengumpulkan postingan Twitter dengan hashtag [40]. Alat lain untuk merayapi data Twitter adalah Tweepy Streaming API, yang memungkinkan untuk mengumpulkan sampel acak secara real time. Terlepas dari fungsinya, API ini terbatas untuk melacak sejumlah kata kunci

dalam satu waktu. Untuk melacak beberapa kata kunci, API harus dijalankan secara terpisah dari instalasi terpisah[16].

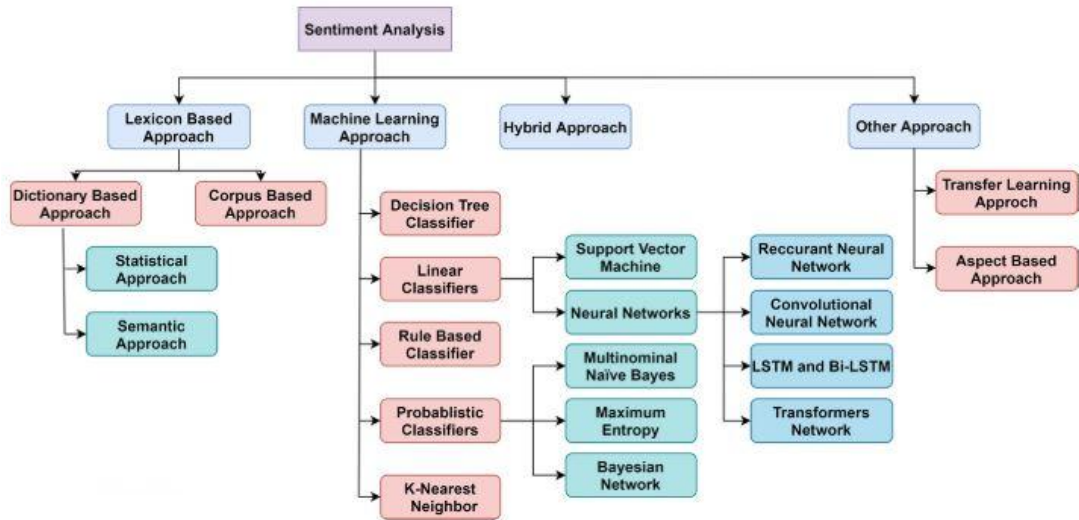
## **2.4 Analisis Sentimen**

Analisis sentimen (AS) atau opini mining (OM) adalah studi dengan bantuan komputasi mengenai opini, sikap, dan perasaan masyarakat terhadap suatu entitas. Entitas ini dapat mewakili individu, peristiwa, atau topik. Topik-topik ini kemungkinan besar akan dibahas dalam ulasan. Kedua ekspresi analisis sentimen atau opini mining dapat dipertukarkan. Keduanya mempunyai arti yang sama. Namun beberapa peneliti mengatakan bahwa analisis sentimen dan opini mining memiliki arti yang sedikit berbeda. Mining opini mengekstrak dan menganalisis opini masyarakat tentang suatu entitas, sementara analisis sentimen mengidentifikasi dan menganalisis sentimen yang diungkapkan dalam teks. Oleh karena itu, tujuan analisis sentimen adalah mengidentifikasi opini, mengidentifikasi emosi yang diungkapkan, dan mengklasifikasikan polaritasnya [41].

Analisis sentimen telah diterima secara luas dalam beberapa tahun terakhir, tidak hanya di kalangan peneliti tetapi juga di kalangan dunia usaha, pemerintah, dan organisasi [12]. Analisis sentimen secara tradisional telah diterapkan di berbagai sektor, mulai dari hotel, maskapai penerbangan, layanan kesehatan hingga pasar saham [42]. Analisis sentimen diterapkan pada ulasan hotel untuk lebih memahami kesukaan dan ketidaksukaan pelanggan. Sebagai perbandingan, analisis sentimen juga digunakan untuk menentukan tren pasar saham dan mata uang kripto berdasarkan sentimen pasar [43]. Penulis, menganalisis tweet yang terkait dengan domain berbeda dan menganalisis sentimen dari tweet tersebut [44]. Aplikasi analisis sentimen baru-baru ini meningkat di bidang kesehatan. Analisis opini pelanggan dan analisis kepuasan pelanggan adalah beberapa penerapannya dalam bidang kesehatan [45]. Departemen perusahaan terus-menerus menggunakan analisis sentimen untuk perbaikan. Analisis sentimen digunakan untuk berbagai aplikasi, termasuk manajemen reputasi, riset pasar dan analisis persaingan, analisis produk, dan umpan balik pelanggan [46].

Analisis sentimen dapat dipandang sebagai proses klasifikasi. Analisis sentimen memiliki tiga tingkat klasifikasi utama: tingkat dokumen, tingkat kalimat, dan tingkat aspek. Analisis sentimen tingkat dokumen bertujuan untuk mengklasifikasikan dokumen opini sebagai yang mengungkapkan opini atau sentimen positif atau negatif. Analisis sentimen ini menganggap keseluruhan dokumen sebagai unit dasar informasi (terkait dengan suatu topik). Analisis sentimen tingkat kalimat bertujuan untuk mengklasifikasikan emosi yang diungkapkan dalam setiap kalimat. Analisis sentimen tingkat aspek bertujuan untuk mengklasifikasikan sentimen yang terkait dengan aspek tertentu dari suatu entitas [41].

Selama dua dekade terakhir, banyak karya yang mengeksplorasi analisis sentimen dalam berbagai aplikasi dan menggunakan pendekatan yang berbeda. Pendekatan-pendekatan ini dapat dikelompokkan ke dalam berbasis leksikon, pembelajaran mesin, dan hibrida [47]. Dari ketiganya, teknik pembelajaran mesin dan pendekatan hibrid tampaknya lebih menjanjikan, dan teknik leksikon biasanya diintegrasikan dengan pendekatan pembelajaran mesin untuk meningkatkan hasil [48]. Pendekatan pembelajaran mesin menerapkan prediktor (pengklasifikasi atau estimator) pada serangkaian fitur yang mewakili masukan. Kumpulan variabel prediktor ini tidak berbeda secara signifikan dengan prediktor yang digunakan di bidang lain. Kompleksitas pendekatan ini terletak pada mengekstraksi fitur-fitur kompleks dari teks, memfilter fitur-fitur yang relevan, dan memilih prediktor yang sesuai [12].



Gambar 2.3. Pendekatan analisis sentiment [46]

## 2.5 Support Vector Machine (SVM)

Metode ini menggunakan pendekatan SVM dengan hyperplanes untuk menganalisis data dan menentukan batasan keputusan. SVM merupakan salah satu jenis teknik pembelajaran non-probabilistik yang biasa digunakan untuk tugas klasifikasi. Tujuan utama SVM adalah untuk menentukan hyperplane yang secara optimal mempartisi data ke dalam kelas yang berbeda [46].

SVM adalah algoritma pembelajaran terawasi yang membangun sebuah hyper-plane atau sekumpulan hyper-planes dalam ruang dimensi tinggi atau tak terbatas, dengan menggunakan fungsi kernel. SVM berusaha memaksimalkan jarak hyper-plane dari contoh pelatihan terdekat, dengan mendapatkan contoh pelatihan yang paling dekat dengan margin maksimum hyper-plane yang disebut sebagai support vector. SVM dapat digunakan untuk masalah klasifikasi atau regresi, di mana SVM mentransformasikan input ke dalam ruang fitur berdimensi tinggi dengan menggunakan fungsi kernel. Fungsi keputusannya adalah:

$$y = \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i K(x_i, x) + \rho \right) \quad [1]$$

Di mana  $y$  adalah label klasifikasi (1 atau -1),  $n$  jumlah vektor pelatihan,  $\alpha$  pengali Lagrange,  $K(x_i, x)$  fungsi Kernel, dan  $\rho$  intersep untuk batas keputusan

margin maksimum. Rumus ini digunakan dalam Support Vector Machine (SVM), di mana fungsi tanda ( $\text{sgn}$ ) menentukan kelas data baru berdasarkan hubungan dengan vektor pelatihan melalui fungsi kernel. [43].

Penelitian yang dilakukan oleh Borg et al. menggunakan Linear SVM dan VADER untuk memprediksi sentimen ulasan pelanggan dari sebuah perusahaan telekomunikasi besar di Swedia. Dataset yang digunakan berisi 168.010 email untuk pelatihan, dengan pelabelan awal berdasarkan sentimen dalam bahasa Swedia dan [49]. Model SVM linier mereka (Li et al., 2019a) menunjukkan performa tinggi dengan skor F1 83,4% dan AUC rata-rata 0,896. Model ini juga mengungkap pola prediksi dalam percakapan email dengan mempertimbangkan sentimen email yang tidak terlihat. Berbeda dari klasifikasi biner biasa, penelitian terdahulu menekankan pentingnya subjektivitas opini dan kredibilitas pengungkap. [50]. Sebuah kerangka kerja diusulkan untuk meringkas opini di mikroblog dengan mengekstrak topik terkait pertanyaan pengguna dan mengkategorikannya menggunakan SVM [51]. Beberapa peneliti juga menggunakan data tweet Twitter untuk percobaan [16] dan menemukan bahwa pengumpulan opini bermanfaat dalam analisis tujuan perjalanan, produk, dan film. Mereka mengusulkan model klasifikasi sentimen berbasis SVM dengan set istilah linguistik probabilistik [50].

Alqaryouti et al. meneliti cara membantu entitas pemerintah memahami ekspektasi pelanggan melalui ulasan. Mereka menemukan bahwa penggunaan leksikon dan aturan sebagai fitur input pada model SVM menghasilkan akurasi lebih tinggi dibandingkan model SVM lainnya. [52]. Hassonah et al. mengusulkan pendekatan pembelajaran mesin hibrida untuk analisis sentimen dengan mengintegrasikan dua teknik pemilihan fitur, yaitu ReliefF dan Multi-Verse Optimizer (MVO), ke dalam SVM guna meningkatkan performa. [53].

SVM adalah pengklasifikasi yang sukses dikembangkan oleh Vapnik (1995). SVM juga telah diterapkan secara luas pada area terkait dalam klasifikasi sentimen. Sebagai contoh, Akhtar dkk. menggunakan Maximum Entropy (ME), Conditional Random Field (CRF), dan Support Vector Machine (SVM) untuk analisis sentimen berbasis aspek [54]. Liu dkk. menunjukkan bahwa support vector machine memiliki kinerja terbaik dibandingkan dengan naïve Bayes, decision tree,

neural network dan K-nearest neighbor dalam klasifikasi sentimen. SVM telah digunakan untuk mengklasifikasikan sentimen komentar online [55][56].

Menemukan hyperplane yang mengkategorikan dataset ke dalam kelas-kelas yang paling efektif adalah tujuan dari metode pembelajaran yang diawasi SVM. Rangkaian algoritma Support Vector Machine (SVM) termasuk SV. Algoritma ini mencari hyperplane yang optimal untuk membagi data ke dalam beberapa kelas. Hyperplane adalah garis lurus dalam pengaturan dua dimensi. Ruang multidimensi membagi kelas-kelas ke dalam dimensi yang lebih tinggi. Menemukan hyperplane terbaik untuk memaksimalkan margin antara dua kelas adalah tujuan SV. Jarak antara vektor pendukung dua kelas, atau titik data, yang paling dekat dengan hyperplane adalah margin. Rumusan masalah ini adalah masalah optimasi. Memaksimalkan margin sambil menjamin klasifikasi titik data yang akurat adalah tujuannya. Penggunaan pemrograman kuadratik dapat mengatasi hal ini [50].

SVM memungkinkan konversi ruang input menjadi ruang dimensi yang lebih tinggi dengan menggunakan berbagai fungsi kernel. Ketika data tidak dapat dipisahkan secara linear dalam ruang aslinya, hal ini sangat membantu. Kernel yang populer termasuk sigmoid, linear, polinomial, dan radial basis function (RBF). Parameter regularisasi (C) mengatur pertukaran antara maksimalisasi margin dan minimalisasi kesalahan klasifikasi. Ini akan mengkategorikan ulasan ke dalam beberapa kategori sentimen, seperti positif atau negatif, dalam konteks analisis sentimen. SVM telah menunjukkan keampuannya dalam tugas klasifikasi teks, dan kemampuannya untuk mengidentifikasi batas-batas kelas yang rumit dapat membantu untuk analisis sentimen. Model ini dapat mendeteksi korelasi linier dan non-linier dalam data karena dapat menggunakan fungsi kernel yang berbeda. Model ini bekerja dengan baik di ruang dimensi tinggi dan oleh karena itu sesuai untuk data teks, di mana mungkin ada sejumlah besar karakteristik [57].

Pembelajaran mesin memungkinkan sistem memperoleh kemampuan baru tanpa diprogram secara eksplisit. Algoritme analisis sentimen dapat dilatih untuk melampaui definisi sederhana dan memahami informasi kontekstual, sarkasme, dan kata-kata yang disalahgunakan. Algoritme yang umum digunakan meliputi: Naive

Bayes (NB), Support vector machine (SVM), Regresi logistik (LR), Decision tree (DT), Maximum entropy (ME), K-nearest neighbours (KNN), Semi-supervised learning [46].

## 2.6 SMOTE Oversampling

SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi ketidakseimbangan data dalam penambahan data. Ketidakseimbangan data terjadi ketika satu atau lebih kelas memiliki jumlah data yang sangat sedikit, sementara kelas lainnya memiliki jumlah data yang lebih banyak. Pendekatan ini menyesuaikan algoritma klasifikasi untuk memperhitungkan masalah ketidakseimbangan data. Dataset yang tidak seimbang, dengan rasio kelas target yang tidak merata, dapat menyebabkan masalah dalam tugas klasifikasi karena model mungkin terlalu menyesuaikan diri dengan kelas mayoritas. Untuk mengatasi masalah ini, teknik pengambilan sampel ulang data seperti SMOTE oversampling diterapkan. SMOTE adalah metode oversampling yang populer dan dirancang untuk meningkatkan oversampling acak dengan menghasilkan sampel sintetis. Terdapat beberapa variasi SMOTE yang dikembangkan untuk mengatasi kelemahan algoritma asli. Dalam penelitian ini, dilakukan analisis karakteristik distribusi sampel sintetis yang dihasilkan oleh SMOTE untuk mengevaluasi kualitas data. Analisis ini bertujuan untuk menilai seberapa baik data sintetis meniru distribusi dasar yang sebenarnya, sehingga meningkatkan kualitas dan keseimbangan dataset [58].