

## BAB II

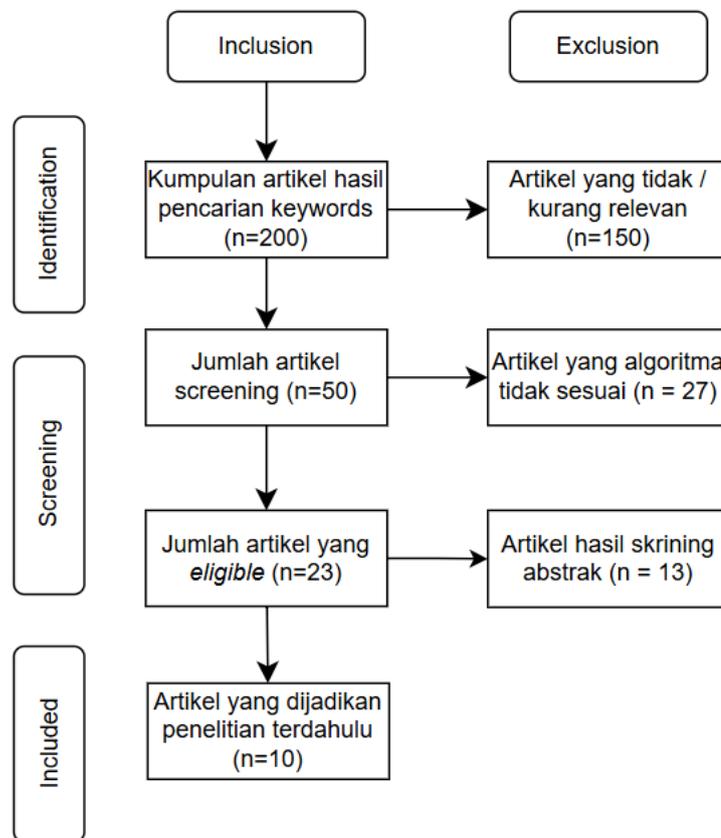
### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Penelitian terdahulu diperoleh melalui metode *Systematic Literature Review* (SLR), yang merupakan pendekatan terstruktur untuk mengumpulkan, mengevaluasi, dan menganalisis penelitian yang telah dipublikasikan terkait suatu topik tertentu [13]. Proses SLR dilakukan secara sistematis mulai dari perumusan pertanyaan penelitian hingga pencarian literatur menggunakan kata kunci yang relevan [13]. Selanjutnya, dilakukan seleksi berdasarkan kriteria inklusi dan eksklusi, serta analisis mendalam terhadap studi yang terpilih.

Dalam proses seleksi penelitian terdahulu, digunakan kriteria inklusi dan eksklusi untuk memastikan hanya studi yang relevan dan memenuhi standar tertentu yang dimasukkan dalam analisis. Proses ini dikendalikan melalui *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA), yaitu suatu alur yang memandu tahap-tahap penyaringan literatur. PRISMA terdiri dari beberapa tahap, yaitu identifikasi sumber, penyaringan awal, seleksi berdasarkan kriteria eksklusi dan inklusi, serta pemilihan akhir penelitian yang akan dianalisis lebih lanjut [14]. Dengan mengikuti metode PRISMA, penelitian dapat lebih transparan dan sistematis dalam menyajikan kajian literatur yang berkualitas. Penelitian terdahulu di dapat dari hasil *screening* dengan menggunakan kriteria eksklusi dan inklusi seperti yang ada pada *flow* Prisma pada Gambar 2.1.

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Gambar 2.1 Flow Prisma

Gambar 2.1 menggambarkan tahap pertama yaitu identifikasi artikel dimulai dengan mengidentifikasi sejumlah besar artikel melalui pencarian kata kunci. Kemudian, artikel-artikel tersebut disaring untuk relevansi dan kualitas (*inclusion*), dan yang tidak memenuhi kriteria dikecualikan (*exclusion*). Akhirnya, artikel-artikel yang memenuhi syarat dimasukkan ke dalam penelitian. Pada tahap ‘*Identification*’, terdapat 200 artikel yang ditemukan melalui pencarian kata kunci ‘*sentiment analysis*’ dan ‘*paylater*’ dengan bantuan *tools* Publish or Perish. Pada tahap ‘*Screening*’, setelah penyaringan awal yaitu dengan mengecualikan artikel yang kurang relevan sebanyak 150 berdasarkan judul dan kata kunci pada abstrak, tersisa 50 artikel tetap. Pada tahap ‘*Eligibility*’, 23 artikel dianggap memenuhi syarat untuk dimasukkan ke dalam penelitian setelah mengecualikan artikel yang menggunakan algoritma tidak sesuai atau relevan berjumlah 27 artikel. Lalu melewati tahap *screening exclusion* lagi setelah melakukan skrining hasil abstrak

dan isi pada bagian metode penelitian dengan jumlah 13 artikel. Pada tahap 'Included', 10 artikel dipilih sebagai referensi untuk penelitian.

Tabel 2.1 Penelitian Terdahulu

No.	Judul Artikel	Penulis	Metode	Objek	Hasil
1.	Analisis Sentimen Twitter Terhadap Pembayaran ShopeePayLater Pada Aplikasi Belanja Online (Shopee) Menggunakan Metode <i>Lexicon Based</i> dan <i>Naïve Bayes Classifier</i> [10]	Indira et al. (2020)	<i>Crawling, Text Pre-processing, Lexicon-Based, Naïve Bayes Classifier, Confusion Matrix</i>	Tweet pengguna Shopee PayLater	Sentimen positif lebih dominan dengan akurasi 82,52%, <i>error rate</i> 17,48%, presisi positif 89,54%, presisi negatif 47,06%, dan <i>recall</i> 89,54%
2.	Analisis Sentimen Pengguna Sistem Pay Later Menggunakan <i>Support Vector Machine</i> Metode Pembobotan <i>Lexicon</i> [11]	Wahidna dan Nerisafitra (2023)	<i>Support Vector Machine (SVM)</i>	Tweet pengguna Shopee PayLater dan Go PayLater	Akurasi model untuk Shopee PayLater: 89,74%, Akurasi model untuk Go PayLater: 90,27%
3.	Analisis Sentimen Pengguna Shopeepaylater Pada Twitter Menggunakan Metode <i>Support Vector Machine (SVM)</i> [15]	Setiawan dan Mulyati (2023)	<i>Support Vector Machine (SVM), Confusion Matrix, Pengujian Blackbox</i>	Tweet pengguna Shopee PayLater	Akurasi 97,1%, Presisi 95,83%, <i>Recall</i> 98,5%, <i>F-1 Score</i> 97%, sistem berbasis web berjalan dengan baik
4.	Penerapan Metode LDA-LSTM Untuk Analisis Sentimen Respon Masyarakat Terhadap Topik Terkait	Wicaksono (2024)	LDA, LSTM	Sentimen Publik terkait Penangkapan Ikan Terukur	Akurasi LSTM 85%

No.	Judul Artikel	Penulis	Metode	Objek	Hasil
	Kebijakan Penangkapan Ikan Terukur [12]				
5.	Analisis Sentimen Tanggapan Masyarakat Indonesia Terhadap Metode Pembayaran Paylater Menggunakan Algoritma <i>Naive Bayes</i> [16]	Nuraeni (2023)	<i>Naive Bayes Classifier, Scraping</i>	Komentar Instagram dari Shopee Indonesia, Gopay Indonesia, Akulaku PayLater, Kredivo, Home Credit	Peringkat paylater berdasarkan persentase sentimen positif: Akulaku PayLater (43,75%), Kredivo (39,39%), SpayLater (34,99%), GopayLater (28,07%), Home Credit (24,63%)
6.	Analisis Sentimen Terhadap Komentar Video Short YouTube Mengenai Paylater Menggunakan Algoritma Bert [17]	Fajris dan Imelda (2025)	<i>Fine Tuning BERT</i>	Komentar video Youtube Short dari akun Raymond Chin (judul: "Kartu Kredit Kalah Populer Sama Paylater")	Akurasi model: 58%, Presisi kelas positif: 48%, Presisi kelas negatif: 88%, <i>Recall</i> kelas positif: 92%, <i>Recall</i> kelas negatif: 35%
7.	Perbandingan Metode Klasifikasi <i>Support Vector Machine</i> (SVM) dan <i>Extreme Gradient Boosting</i> (XGBOOST) Pada Klasifikasi Sentimen Aplikasi Paylater[18]	Zahrah dan Fatimatuz (2023)	<i>Support Vector Machine</i> (SVM) dengan kernel <i>Linear</i> dan RBF, <i>Extreme Gradient Boosting</i> (XGBoost)	Ulasan pengguna Kredivo	SVM <i>Non-linear</i> dengan kernel RBF menghasilkan akurasi 94,45%, <i>F-1 score</i> 96,18%, AUC 92,39%, lebih baik dari SVM <i>Linear</i> dan XGBoost

No.	Judul Artikel	Penulis	Metode	Objek	Hasil
8.	<i>An Integrated Approach for Sentiment Analysis and Topic Modeling of a Digital Bank in Indonesia using Naïve Bayes and Latent Dirichlet Allocation Algorithms on Social Media Data</i> [19]	Setiawan et al. (2024)	<i>Naïve Bayes Algorithm (Sentiment Analysis), Latent Dirichlet Allocation (Topic Modeling)</i>	Data media sosial dari Twitter dan Instagram mengenai Bank Digital Line	<i>Naïve Bayes</i> mencapai <i>F1 score</i> 0.863, sentimen positif lebih banyak di Twitter, sentimen negatif lebih banyak di Instagram. LDA mengidentifikasi 4 topik untuk sentimen positif dan 5 topik untuk sentimen negatif dengan koherensi 0.426279 (positif) dan 0.397232 (negatif).
9.	Analisis Pinjaman Online Pada Sosial Media Twitter Menggunakan Latent Dirichlet Allocation (LDA) [20]	Nautika dan Yustanti (2024)	<i>Latent Dirichlet Allocation (LDA), Knowledge Discovery in Databases (KDD)</i>	<i>Tweet</i> mengenai pinjaman online	Pemodelan LDA terbaik pada skenario satu (iterasi 10, $\alpha$ 0,1, $\beta$ 0,1) dengan <i>coherence score</i> 0,391, menghasilkan 9 topik
10.	<i>Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter</i> [21]	Choirul et al. (2020)	Pemodelan topik menggunakan LDA, and <i>lexicon-based sentiment analysis</i>	Bencana banjir	Sentimen negatif 79%, data tweet dikelompokkan ke dalam 9 topik dan tiap topik direpresentasikan oleh

No.	Judul Artikel	Penulis	Metode	Objek	Hasil
					10 kata dominan.

Penelitian-penelitian sebelumnya telah banyak mengkaji analisis sentimen dan pemodelan topik dalam berbagai konteks, termasuk layanan keuangan digital seperti *paylater*. Untuk memposisikan penelitian ini, tinjauan terhadap beberapa studi relevan telah dilakukan, yang dapat dikelompokkan berdasarkan fokus analisis, metode, dan objek penelitiannya.

Penelitian yang secara khusus menganalisis sentimen terhadap layanan Shopee PayLater menggunakan berbagai metode klasifikasi. Sebagai contoh, penelitian yang dilakukan Indira et al. [10] memanfaatkan kombinasi pendekatan *Lexicon-Based* dan *Naïve Bayes Classifier* untuk mengkaji sentimen pada *tweet* pengguna Shopee PayLater, dengan hasil yang menunjukkan dominasi sentimen positif. Objek yang sama pula diteliti oleh Wahidna [11] serta Setiawan [15] yang juga melakukan investigasi sentimen pengguna Shopee PayLater berdasarkan data *tweet*, namun kedua penelitian ini mengandalkan *Support Vector Machine (SVM)*. Studi oleh Wahidna turut menyertakan perbandingan dengan GoPayLater, sementara Setiawan melaporkan tingkat akurasi yang tinggi untuk model SVM yang mereka kembangkan. Meskipun studi-studi ini sangat relevan karena berfokus pada Shopee PayLater dan beberapa di antaranya telah menggunakan SVM, penelitian yang diusulkan saat ini memiliki cakupan yang lebih luas. Perbedaannya terletak pada upaya untuk tidak hanya menerapkan SVM, tetapi juga melakukan perbandingan kinerja secara komprehensif dengan algoritma *deep learning* seperti *Long Short-Term Memory (LSTM)*. Lebih lanjut, penelitian ini mengintegrasikan hasil analisis sentimen tersebut dengan pemodelan topik menggunakan *Latent Dirichlet Allocation (LDA)* guna memperoleh pemahaman yang lebih holistik, sebuah dimensi analisis yang belum dieksplorasi secara mendalam oleh penelitian-penelitian sebelumnya dalam kelompok ini. Studi-studi tersebut penting karena mengonfirmasi signifikansi Shopee PayLater sebagai subjek kajian analisis sentimen dan menyediakan data performa awal untuk metode SVM yang dapat dijadikan acuan.

Berbagai penelitian yang menganalisis sentimen pada layanan *paylater* yang lebih menerapkan metode *machine learning* alternatif. Penelitian yang dilakukan Nuraeni [16], menggunakan *Naïve Bayes Classifier* untuk menganalisis sentimen terhadap berbagai merek *paylater*, termasuk Shopee PayLater, berdasarkan data komentar Instagram, sehingga menyajikan peta perbandingan sentimen antar penyedia layanan. Di sisi lain, penelitian oleh Fajris mengadopsi pendekatan *deep learning* yang lebih mutakhir dengan melakukan *fine-tuning* pada model BERT untuk analisis sentimen komentar video YouTube Shorts yang membahas topik *paylater* [17]. Sementara itu, penelitian oleh Zahrah melakukan studi komparatif antara SVM dan *Extreme Gradient Boosting (XGBoost)* untuk klasifikasi sentimen pada ulasan pengguna layanan Kredivo [18]. Penelitian ini membedakan diri dari studi-studi tersebut melalui fokusnya yang spesifik pada Shopee PayLater sebagai studi kasus utama, bukan perbandingan antar berbagai merek sebagaimana dilakukan Nuraeni. Meskipun penelitian oleh Fajris juga memanfaatkan *deep learning*, algoritma yang menjadi fokus perbandingan dalam penelitian ini adalah LSTM dan SVM, dengan sumber data yang berbeda yaitu X dan Instagram. Jika dibandingkan dengan penelitian Zahrah [18], penelitian ini tidak hanya terbatas pada perbandingan algoritma klasifikasi sentimen antara LSTM dan SVM tetapi juga pendalaman analisis dengan pemodelan topik LDA yang tidak ada dalam studi tersebut, serta pemilihan objek Shopee PayLater yang sesuai dengan latar belakang merupakan platform *paylater* yang paling banyak digunakan di kalangan Gen Z dan milenial di Indonesia.

## **2.2 Teori tentang Topik Skripsi**

### **2.2.1 Media Sosial**

Media sosial dapat didefinisikan sebagai kumpulan aplikasi berbasis internet yang dibangun di atas fondasi teknologi Web 2.0, yang memungkinkan penciptaan dan pertukaran konten buatan pengguna [22]. Artinya, media sosial menjadi wadah bagi pengguna untuk saling berkomunikasi dan mengekspresikan pendapat secara daring. Sebagai sumber data opini publik, media sosial menawarkan volume informasi yang besar dan *real-time*. Misalnya, penelitian yang dilakukan oleh Dong dan Lian menunjukkan bahwa

dibandingkan dengan survei tradisional, analisis berbasis media sosial dapat memberikan pemahaman yang lebih baik dan komprehensif terhadap persepsi publik tentang topik tertentu secara lebih ilmiah [23]. Dalam konteks ini, opini publik di media sosial sering diekstraksi melalui teknik data mining untuk mengevaluasi sentimen atau tren publik. Jenis-jenis media sosial meliputi jejaring sosial (seperti Facebook), berbagi media (seperti Instagram dan YouTube), forum diskusi (seperti Reddit), serta layanan blog.

Dalam konteks analisis sentimen, media sosial menjadi sumber data utama karena banyaknya opini yang diungkapkan pengguna secara spontan. Analisis sentimen pada media sosial bertujuan untuk memahami emosi atau pendapat publik terhadap suatu isu, produk, atau layanan. Misalnya, penelitian sebelumnya menggunakan analisis sentimen untuk mengukur opini publik terhadap kebijakan tertentu atau performa perusahaan di media sosial seperti X (Twitter) dan Instagram [24]. Dengan metode seperti LSTM atau SVM, sentimen dapat diklasifikasikan menjadi positif atau negatif, memberikan wawasan berharga bagi pengambil keputusan.

### 2.2.2 X

X adalah salah satu aplikasi media sosial yang telah berganti nama sebelumnya dari Twitter sejak Juli 2023[25]. Pada platform X, penggunanya dapat memposting media berupa teks, gambar, video, hingga audio. X lebih dikenal dengan media sosial untuk berbagi teks. Pengguna X dapat melakukan postingan, menyukai postingan (*like*), memposting ulang (*repost*), dan mengutip postingan (*quote post*), juga mengirim pesan ke sesama penggunanya. Pengumpulan data X memungkinkan peneliti menangkap opini publik secara langsung dan real-time. Misalnya, hasil penelitian yang dilakukan Ramlan melaporkan bahwa pada Januari 2023 terdapat sekitar 556 juta pengguna global X, di mana Indonesia menempati peringkat kelima terbanyak pengguna platform ini [26]. Hal ini menunjukkan skala besar data yang tersedia untuk analisis sentimen pengguna Twitter di Indonesia.

Dalam penelitian analisis sentimen, X sering digunakan karena sifatnya yang real-time dan terbuka. Data dari X sangat relevan untuk memahami opini publik terhadap isu-isu terkini. Contohnya, penelitian tentang regulasi TikTok Shop di Indonesia memanfaatkan data dari X untuk mengklasifikasikan sentimen pengguna terhadap kebijakan tersebut menggunakan *algoritma Naïve Bayes* [27]. Data dari X juga digunakan untuk menganalisis popularitas figur politik atau reaksi terhadap acara budaya tertentu [28]. X menjadi platform penting dalam studi analisis sentimen karena volume data yang besar dan beragam

### **2.2.3 Instagram**

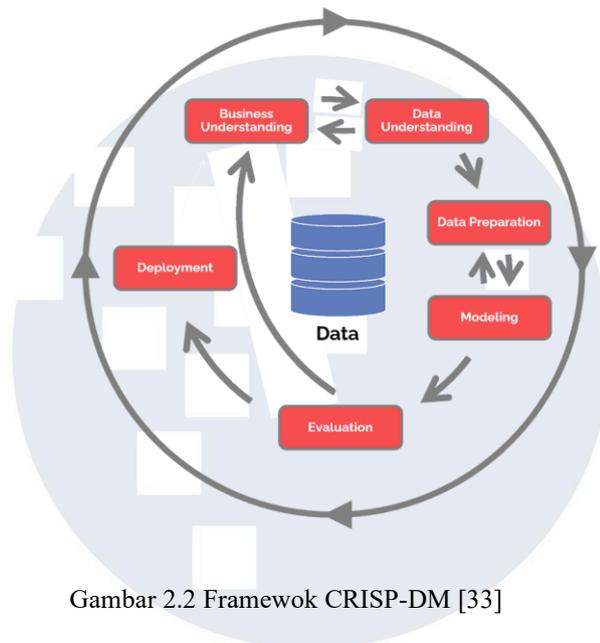
Instagram adalah platform media sosial berbasis visual (foto dan video) yang juga banyak digunakan dalam analisis sosial dan pemasaran. Instagram saat ini menjadi platform media sosial yang paling diidentikkan dengan konten gambar secara daring [26]. Setiap unggahan foto atau video disertai keterangan (*caption*) dan tagar (*hashtag*) yang memberi konteks dan opini pengguna. Dominasi konten visual dan jaringan pengguna yang besar, Instagram sering dimanfaatkan untuk studi pengaruh media sosial terhadap opini publik, analisis tren pasar. Misalnya, dalam penelitian pemasaran, Instagram digunakan untuk mengukur sentimen konsumen terhadap merek melalui analisis komentar dan tagar yang relevan.

Dalam konteks analisis sentimen, Instagram merupakan sumber data yang kaya karena komentar pengguna pada unggahan dapat mencerminkan opini mereka terhadap isu tertentu. Penelitian sebelumnya menunjukkan bahwa analisis sentimen pada komentar Instagram dapat memberikan wawasan tentang pandangan publik terhadap kebijakan atau fenomena tertentu [29]. Misalnya, penelitian tentang kebijakan pendidikan di Indonesia menggunakan komentar Instagram untuk mengidentifikasi sentimen positif dan negatif terhadap kebijakan tersebut [30].

## **2.3 Teori tentang Framework/Algoritma yang digunakan**

### **2.3.1 CRISP-DM**

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah salah satu framework data mining yang memberikan gambaran pada suatu siklus proyek *data mining* [31]. CRISP-DM adalah metodologi yang paling sering digunakan dalam proyek *data mining* [32]. CRISP-DM terdiri dalam 6 tahapan seperti pada Gambar 2.2.



Gambar 2.2 Framework CRISP-DM [33]

Penjelasan dari tiap tahapan tersebut adalah sebagai berikut:

1. *Business Understanding*

*Business understanding* merupakan tahapan di mana dilakukan pemahaman mendalam pada tujuan penelitian yang ditargetkan untuk dicapai. Tahap awal dalam proses ini adalah pemahaman bisnis, di mana dilakukan penentuan masalah yang akan diselesaikan serta penetapan tujuan yang hendak dicapai melalui analisis data.

2. *Data Understanding*

*Data understanding* merupakan tahapan yang melibatkan pengumpulan data yang relevan dengan permasalahan yang ada, eksplorasi data yang ada, serta memahami data lebih dalam dari segi data dan kualitas untuk dilanjutkan ke tahap berikutnya.

3. *Data Preparation*

*Data preparation* merupakan tahapan yang melibatkan pemrosesan dan pembersihan data dengan membersihkan data (*cleaning data*), integrasi data dan mempersiapkan data untuk ke tahap pemodelan (*modeling*).

#### 4. *Modeling*

*Modeling* merupakan tahapan yang melibatkan pengembangan dan pemilihan model yang disesuaikan dengan tujuan penelitian yang sudah ditentukan sebelumnya. Pada tahapan ini, dapat digunakan berbagai algoritma dan metode analisis data untuk membangun model yang dapat melakukan prediksi, klasifikasi, hingga segmentasi.

#### 5. *Evaluation*

*Evaluation* merupakan tahapan yang melibatkan evaluasi dari kinerja model yang sudah dikembangkan di tahap *modeling*. Tahapan ini dilakukan untuk memastikan kinerja model yang dikembangkan sudah sesuai dengan tujuan analisis.

#### 6. *Deployment*

*Deployment* merupakan tahapan yang melibatkan implementasi dari model tersebut ke lingkungan yang dapat berupa aplikasi, sistem, dan lainnya. Pada tahapan ini, perlu diperhatikan agar model dapat berintergrasi dengan sistem yang ada

### 2.3.2 *Long Short-Term Memory (LSTM)*

*Long Short-Term Memory (LSTM)* merupakan salah satu jenis arsitektur dari *Recurrent Neural Network (RNN)* yang dikembangkan secara khusus untuk menangani masalah *vanishing gradient* yang kerap muncul pada RNN tradisional. Algoritma ini menggunakan struktur yang kompleks yang terdiri dari sel memori (*memory cells*) dan berbagai *gates* yang memungkinkan jaringan untuk menyimpan informasi dalam periode waktu jangka panjang [34]. LSTM mengimplementasikan mekanisme gerbang *input*, gerbang *output*, dan gerbang *forget* yang secara selektif mengontrol aliran informasi dan memori, memungkinkan model untuk mempelajari dependensi jangka panjang dalam data sekuensial. Keunggulan utama LSTM adalah kemampuannya untuk mempertahankan informasi penting dan melupakan informasi yang tidak

relevan dari waktu ke waktu. Model LSTM secara efektif dapat mengolah ketergantungan panjang dalam teks dan menawarkan keunggulan akurasi dibandingkan model *machine learning* tradisional pada tugas klasifikasi sentimen [35].

### **2.3.3 Support Vector Machine (SVM)**

*Support Vector Machine* (SVM) merupakan salah satu model machine learning yang sangat populer untuk melakukan klasifikasi dan juga regresi. Tujuan utama SVM adalah menemukan hyperplane terbaik yang memisahkan dua kelas dalam ruang fitur (*feature space*). Dalam konteks klasifikasi, SVM dapat bekerja baik untuk data linier maupun non-linier. Jika data tidak dapat dipisahkan secara linier, SVM menggunakan fungsi kernel untuk mentransformasikan data ke dalam dimensi yang lebih tinggi, di mana pemisahan linier mungkin dapat dicapai. *Support Vector Machine* (SVM) memiliki cakupan penerapan yang luas dan telah dimanfaatkan dalam beragam bidang, seperti pengenalan pola, klasifikasi dokumen teks, identifikasi wajah, serta analisis data *neuroimaging*. Salah satu penelitian yang meneliti analisis sentimen menggunakan SVM pada kasus kenaikan harga BBM menyatakan SVM merupakan salah satu pengklasifikasi teks terkuat karena mampu memisahkan kelas secara *linear* dengan margin yang besar dan menangani vektor fitur berdimensi sangat tinggi [26].

### **2.3.4 Latent Dirichlet Allocation (LDA)**

*Latent Dirichlet Allocation* (LDA) adalah model probabilistik generatif yang digunakan untuk pemodelan topik dalam koleksi dokumen. LDA mengasumsikan setiap dokumen sebagai campuran acak dari sejumlah topik laten, di mana setiap topik diwakili oleh distribusi probabilitas atas kata-kata. Dalam LDA setiap dokumen dituliskan sebagai kombinasi topik-topik tertentu, dan setiap topik digambarkan oleh sebaran kata yang lazim muncul bersama [36]. Dengan pendekatan ini, LDA dapat mengidentifikasi struktur topik yang mendasari korpus teks secara otomatis. Model LDA sering digunakan untuk ekstraksi topik utama dalam analisis teks besar, sehingga membantu peneliti memahami tema umum dari sekumpulan dokumen. Penting untuk membedakan

antara pemodelan topik seperti LDA dengan visualisasi *word cloud*. Sebuah *word cloud* adalah visualisasi sederhana yang hanya menampilkan frekuensi kata, di mana ukuran kata sebanding dengan frekuensi kemunculannya, namun tidak dapat mengungkap hubungan kontekstual antar kata. Sebaliknya, LDA adalah metode statistik yang lebih canggih karena mampu mengidentifikasi kelompok kata yang cenderung muncul bersama dalam dokumen untuk mengungkap topik atau tema laten yang mendasari data. Hal ini memberikan pemahaman yang jauh lebih dalam dan terstruktur mengenai percakapan pengguna.

## **2.4 Teori tentang tools/software yang digunakan**

### **2.4.1 Google Colaboratory**

Google Colab atau singkatan dari Google Colaboratory merupakan layanan yang disediakan oleh Google berbasis *cloud* yang dapat untuk menjalankan kode Python. Beberapa fitur yang disediakan oleh Google Colab yaitu dapat menjalankan Jupyter Notebook didalamnya, sehingga secara tidak langsung setiap blok yang dijalankannya berjalan diatas cloud dan tidak secara tidak langsung tidak menggunakan *local* GPU. Google Colab juga terintegrasi dengan layanan penyimpanan cloud Google Drive, memudahkan penyimpanan dan berbagi notebook secara *online*.

### **2.4.2 Python**

Python dikenal sebagai bahasa pemrograman tingkat lanjut yang sering dijadikan alat utama dalam pengolahan data dan NLP. Kelebihan Python terletak pada sintaks yang mudah dipahami serta *library* yang beragam. Berbagai *library* seperti Pandas, NumPy, scikit-learn, NLTK, spaCy, TensorFlow, dan PyTorch menyediakan fungsi siap pakai untuk pemrosesan data, analisis statistik, *machine learning*, dan pemodelan bahasa alami. Python banyak digunakan sebagai bahasa utama dalam penelitian *data science* dan pengembangan aplikasi analisis sentimen atau pemodelan topik karena fleksibilitas dan dukungan *library* tersebut.

### **2.4.3 Tweet-Harvest**

Tweet-Harvest adalah sebuah alat yang dirancang untuk mengumpulkan dan menganalisis data dari platform media sosial X. Alat ini memungkinkan peneliti dan analis untuk melakukan pengambilan data secara otomatis, sehingga mempermudah proses pengumpulan informasi yang relevan dari tweet, retweet, dan interaksi pengguna lainnya. Tweet-Harvest banyak digunakan dalam penelitian akademik, seperti analisis sentimen dan studi tren sosial. Selain itu, alat ini juga bermanfaat untuk pengumpulan data dalam skala besar tanpa memerlukan akses ke API X resmi, sehingga cocok untuk pengguna yang tidak memiliki akun pengembang (*developer account*). Penggunaan *tools* ini memerlukan otorisasi token dari akun X pengguna untuk mengakses data secara legal dan sesuai kebijakan privasi X [37].

#### 2.4.4 Apify

Apify adalah platform berbasis cloud yang menyediakan layanan pemrograman web untuk pengambilan data (*web scraping*), otomatisasi tugas online, dan integrasi API. Dengan Apify, pengguna dapat membangun dan menjalankan bot yang mampu mengekstrak informasi dari berbagai situs web secara otomatis. Apify sangat berguna bagi peneliti dan perusahaan dalam mengumpulkan data pasar, menganalisis tren industri, serta memantau harga produk kompetitor. Apify mampu menangani berbagai kebutuhan pengambilan data secara otomatis tanpa intervensi manual. Situs resmi Apify menyatakan bahwa Apify adalah platform *full-stack web scraping* untuk developer, dengan *pre-built web scraping* yang telah dibuat sebelumnya sehingga mudah digunakan [38].

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A