

BAB II LANDASAN TEORI

2.1 Penelitian Terkait

Berikut Tabel 2.1 yang merangkum beberapa penelitian terkait sebagai dasar dalam penelitian ini.

Tabel 2.1 Penelitian Terkait

Judul Artikel, Penulis	Nama Jurnal	Objek Penelitian	Metode	Sumber Data	Findings
Sezai Tunca, Bulent Sezen, Violetta Wilk, “ <i>An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing</i> ” [3].	<i>Journal of Big Data</i> / 10 / 82 / 2023	Objeknya penelitiannya adalah konsep "metaverse" yang didapatkan pasca <i>rebranding</i> Facebook menjadi meta dan artikel-artikel yang terkait dengan metaverse yang dipublikasikan di situs web The Guardian dari tahun 2021 hingga 2022.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> (VADER) • Leximancer • MonkeyLearn 	<ul style="list-style-type: none"> • The Guardian Article 	<ul style="list-style-type: none"> • Dari semua artikel yang dipublikasikan dalam periode 2021-2022 tentang metaverse, 61% memiliki sentimen positif, 30% memiliki sentimen negatif, dan 9% bersifat netral. • Tema-tema utama yang muncul dalam analisis Leximancer meliputi "metaverse," "Facebook," "games," dan "platforms." • Penggunaan Monkeylearn menghasilkan klasifikasi peran industri, acara, dan bisnis yang terkait dengan konten metaverse dalam artikel.
Nor Hasliza Md Saad, Alya Syahirah Zainul Abidi, Zulnaidi Yaacod, Muhamad Mu'az Mohd Ali, Zhu Kun, “ <i>Twitter Sentiment Analysis on</i>	<i>International Journal on Recent and Innovation Trends in Computing and</i>	Objek penelitian berfokus pada analisis sentimen terhadap <i>rebranding</i> Facebook menjadi Meta dan topik-topik yang dibicarakan oleh pengguna media sosial terkait Meta.	<ul style="list-style-type: none"> • <i>Sentiment analysis</i> • <i>Lexicon-based polarity detection</i> • RapidMiner 	<ul style="list-style-type: none"> • Twitter Data 	<ul style="list-style-type: none"> • 36% dari tweet memiliki sentimen positif terhadap pengumuman <i>rebranding</i> Meta. • 29% dari tweet memiliki sentimen negatif terhadap pengumuman <i>rebranding</i> Meta.

<i>Meta: A Lexicon-Based Analysis using Rapidminer” [6].</i>	<i>Communication / 11 / 6 / 2023</i>				<ul style="list-style-type: none"> • 35% dari tweet bersifat netral terhadap pengumuman <i>rebranding</i> Meta. • Topik-topik yang sering dibicarakan dalam tweet terkait Meta antara lain NFT (Non-Fungible Token) dan saham, reaksi pengguna terhadap Meta, Metaverse, kemitraan, dan AR (Augmented Reality).
Eskiyaturofikoh and R. R. Suryono, "Analisis sentimen aplikasi X pada Google Play Store menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM)” [14].	Jurnal Ilmiah Penelitian dan Pembelajaran Informatika, vol. 9, no. 3, pp. 1408–1419, 2024	Sentimen pengguna terhadap aplikasi X yang diulas di platform Google Play Store.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Support Vector Machine • Naïve Bayes • SMOTE 	<ul style="list-style-type: none"> • Google Playstore 	<ul style="list-style-type: none"> • Naïve Bayes dan SVM efektif untuk analisis sentimen aplikasi "X". • Penerapan SMOTE meningkatkan akurasi terutama pada model SVM. • Penelitian sebelumnya hanya fokus pada data sesudah rebranding. • Dampak positif atau negatif rebranding belum dijelaskan secara mendalam.
Sulton Nur Hakim, Andika Julianto Putra, Annisa Uswatun Khasanah, “ <i>Sentiment analysis on myindihome user reviews using support vector machine and naïve bayes classifier method</i> ” [15].	<i>International Journal of Industrial Optimization / 2 / 2 / 2021</i>	Ulasan pengguna IndiHome yang ditemukan pada aplikasi myIndiHome di Google Play merupakan objek penelitian. Penelitian ini berfokus pada mengevaluasi sentimen dari ulasan-ulasan ini, khususnya dalam hal pengaruhnya terhadap pemilihan dan penggunaan produk penyedia layanan internet.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Support Vector Machine • Naïve Bayes • FishBone Diagrams 	<ul style="list-style-type: none"> • Google Playstore 	<ul style="list-style-type: none"> • Metode SVM memiliki akurasi total rata-rata sebesar 86.54%. • Metode Naïve Bayes memiliki akurasi total rata-rata sebesar 84.69%.
D. A. Kristiyanti, D. A. Putri, E. Indrayuni, A. Nurhadi, A. H. Umam, ” <i>E-Wallet Sentiment</i>	<i>2020 Journal of Physics: Conference Series,</i>	Objek penelitiannya adalah aplikasi Ovo, dan Dana. Data diambil dari Google PlayStore komentar.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Naïve Bayes • SVM 	<ul style="list-style-type: none"> • Google Play 	<ul style="list-style-type: none"> • Naïve Bayes: akurasi ditemui sebesar 94.90 %. • SVM akurasinya mendapatkan sebesar 91 %.

<i>Analysis Using Naïve Bayes and Support Vector Machine Algorithm” [16].</i>	<i>Conf. Ser. 1641 012079</i>				
Barka Satya, Muhammad Hasan S J, Majid Rahardi, Ferian Fauzi Abdulloh, “ <i>Sentiment Analysis of Review Sestyc Using Support Vector Machine, Naive Bayes, and Logistic Regression Algorithm” [17].</i>	<i>2022 5th International Conference on Information and Communications Technology (ICOIACT)</i>	Objek penelitian dalam penelitian ini adalah ulasan pengguna aplikasi sestyc di Indonesia.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • SVM • Naïve Bayes • Logistic Regression 	<ul style="list-style-type: none"> • Google Play Store 	<ul style="list-style-type: none"> • Mereka mengumpulkan 8.000 ulasan dari Play Store dan melakukan klasifikasi sentimen (positif/negatif) menggunakan Support Vector Machine (SVM), Naïve Bayes, dan Logistic Regression. • Hasil penelitian mereka menunjukkan bahwa SVM memiliki akurasi tertinggi (87,81%), diikuti oleh Naïve Bayes (81,64%) dan Logistic Regression (80,78%).
S. A. Bahtiar, C. Dewa, and A. Luthfi, “ <i>Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling” [19].</i>	<i>Journal of Information Systems and Informatics Vol. 5, No. 3, September 2023</i>	Penelitian ini berfokus pada analisis sentimen ulasan pengguna di marketplace yang terdapat di Google Play Store. Marketplace yang diteliti mencakup platform seperti Shopee dan Lazada, dengan tujuan memahami bagaimana pengguna merespons aplikasi tersebut melalui ulasan yang diberikan.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Naïve Bayes • Logistic Regression 	<ul style="list-style-type: none"> • Google Play Store 	<ul style="list-style-type: none"> • Logistic Regression memberikan akurasi terbaik (84.58%) pada dataset Shopee dengan 2 label. • Lalu untuk Naïve Bayes ditemui akurasi terbaik (84.58%) dengan 2 label. • Secara keseluruhan, klasifikasi dengan 2 label lebih akurat dibandingkan dengan 3 label.
Kisma, A., Widiawati, C., & Suliswaningsih, S, “ <i>Analysis of applications in Playstore based on Rating and Type using Naive Bayes and Logistic Regression” [20].</i>	<i>JATISI (Jurnal Teknik Informatika Dan Sistem Informasi) 10 / 2 / 2023</i>	Aplikasi-aplikasi yang tersedia di Google Play Store, dengan fokus pada klasifikasi jenis aplikasi (gratis atau berbayar) dan rating aplikasi yang paling banyak diunduh oleh pengguna.	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Naïve Bayes • Logistic Regression 	<ul style="list-style-type: none"> • Google Playstore 	<ul style="list-style-type: none"> • Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dan Logistic Regression mampu mengklasifikasikan data aplikasi Google Play Store dengan akurasi tinggi, di mana Naive Bayes memperoleh akurasi sebesar 92,63%

					<p>dan Logistic Regression sebesar 92,60%.</p> <ul style="list-style-type: none"> • Penelitian juga menemukan bahwa aplikasi yang paling banyak diunduh oleh pengguna adalah aplikasi gratis.
<p>Dimas Diandra Audiansyah, Dian Eka Ratnawati, Buce Trias Hanggara, “Analisis Sentimen Aplikasi MyXL menggunakan Metode Support Vector Machine berdasarkan Ulasan Pengguna di Google Play Store” [21].</p>	<p>Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer / 6 / 8 / 2022</p>	<p>Ulasan yang diberikan oleh pengguna aplikasi MyXL pada platform Google Play Store merupakan objek penelitian. Aplikasi MyXL merupakan aplikasi self-service yang disediakan oleh PT XL Axiata Tbk dan digunakan untuk memfasilitasi pengguna dalam melakukan berbagai layanan XL, seperti aktivasi paket internet, pengecekan pulsa, dan sisa kuota, dan sebagainya.</p>	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • SVM 	<ul style="list-style-type: none"> • Google Play Store 	<ul style="list-style-type: none"> • Rerata akurasi ditemui sebesar 88% • Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna aplikasi MyXL di Google Play Store berdasarkan sentimen, dan hasilnya menunjukkan tingkat akurasi yang tinggi dalam mengidentifikasi ulasan positif dan negatif.
<p>J. Setiawan, A. Milenia, and A. Faza, “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” [22].</p>	<p>2023 4th International Conference on Big Data Analytics and Practices (IBDAP)</p>	<p>Bank digital di Indonesia, khususnya Jenius, Jago, dan Line Bank.</p>	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Topic Modeling • Naïve Bayes • LDA 	<ul style="list-style-type: none"> • Web Scraping di Twitter dan Instagram 	<ul style="list-style-type: none"> • Penelitian berhasil melakukan analisis sentimen dan pemodelan topik pada data media sosial (Twitter dan Instagram) terkait tiga bank digital di Indonesia: Bank Jenius, Jago, dan Line Bank. • Data dari Instagram didominasi sentimen negatif, karena platform ini sering digunakan untuk menyampaikan keluhan.
<p>I. Suyuti and D. R. Sari S, "Fine-Grained Sentiment Analysis on PeduliLindungi Application Users with</p>	<p>2022 9th International Conference on Electrical Engineering, Computer Science</p>	<p>Objek Penelitian ini berfokus pada analisis sentimen ulasan pengguna aplikasi PeduliLindungi di Google Play Store, khususnya dalam konteks kebijakan PSBB</p>	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • Naïve Bayes • SMOTE 	<ul style="list-style-type: none"> • Google Play Store 	<ul style="list-style-type: none"> • Naïve Bayes + SMOTE memberikan akurasi tertinggi dibanding metode tanpa SMOTE. • SMOTE terbukti meningkatkan performa dalam mengatasi

<p><i>Multinomial Naive Bayes-SMOTE" [23].</i></p>	<p><i>and Informatics (EECSI)</i></p>	<p>(Pembatasan Sosial Berskala Besar).</p>			<p>ketidakseimbangan data pada analisis sentimen.</p> <ul style="list-style-type: none"> • Hasil penelitian menunjukkan bahwa kombinasi Naïve Bayes dan SMOTE adalah metode terbaik dalam klasifikasi sentimen pengguna PeduliLindungi.
<p>A. N. A. Saputra, R. E. Saputro, and D. I. S. Saputra, “<i>Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments</i>” [24]</p>	<p>Sinkron, vol. 9, no. 2, pp. 687-699, Apr. 2025</p>	<p>Objek penelitian ini berupa komentar pengguna yang terdapat pada platform YouTube, khususnya pada video terkait film Pengabdian Setan 2: Communion.</p>	<ul style="list-style-type: none"> • <i>Sentiment Analysis</i> • SVM • SMOTE-ENN 	<ul style="list-style-type: none"> • Youtube Comment 	<ul style="list-style-type: none"> • Normalisasi terhadap kata-kata slang berhasil meningkatkan pemahaman model terhadap konteks komentar informal. • Teknik SMOTE-ENN efektif dalam mengatasi ketidakseimbangan kelas, sehingga mendukung proses klasifikasi sentimen yang lebih representatif.

UMMN

UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

Berdasarkan Tabel 2.1, penelitian-penelitian yang dipilih tidak hanya berdasarkan kemunculan kata kunci algoritma tertentu, melainkan juga mempertimbangkan relevansi objek penelitian, konteks analisis sentimen, dan metode yang sudah umum digunakan dalam domain klasifikasi teks. Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, serta teknik penanganan data tidak seimbang seperti SMOTE memang merupakan metode yang banyak diadopsi karena efektivitas dan kestabilannya dalam analisis sentimen. Oleh karena itu, kemunculan metode tersebut dalam literatur yang dirangkum mencerminkan tren nyata dan praktik umum di bidang ini, bukan sekadar bias pencarian literatur.

Dengan demikian, pemilihan algoritma dalam penelitian ini tidak hanya berdasarkan popularitas di literatur, melainkan juga karena kesesuaiannya dengan karakteristik data dan tujuan penelitian. Selain itu, penelitian ini melengkapi studi terdahulu dengan fokus yang lebih spesifik pada perubahan sentimen akibat fenomena rebranding Twitter menjadi X, pembagian data berdasarkan periode waktu, serta pengujian teknik penanganan data tidak seimbang yang lebih variatif, termasuk penggunaan metode gabungan seperti SMOTE-ENN. Kebaruan dalam penelitian ini terletak pada kombinasi pendekatan tersebut dalam konteks rebranding platform media sosial, serta analisis terpisah antara periode sebelum dan sesudah rebranding, yang belum banyak dieksplorasi dalam studi terdahulu. Pendekatan ini dilakukan untuk menguji apakah kombinasi teknik tertentu dapat memberikan hasil yang lebih optimal dibandingkan metode konvensional yang sering digunakan dalam penelitian terdahulu.

2.2 Analisis sentimen

Secara kegunaan, analisis sentimen merupakan pengetahuan atau teknik yang digunakan untuk mencari emosi berdasarkan ulasan, baik dalam bentuk teks maupun verbal [25]. Teknik ini sering digunakan untuk mengeksplorasi perasaan yang terkandung dalam ulasan pelanggan terhadap produk tertentu. Dalam menggunakan analisis sentimen, maka diperlukan sebuah dataset yang telah memiliki label. Pendekatan dengan metode pembelajaran *supervised*

merupakan pendekatan menggunakan pembelajaran mesin dan kecerdasan buatan yang memanfaatkan data dengan sifat telah terlabeli. Model tersebut akan mampu menghasilkan output yang akurat dan sesuai dengan label yang diinginkan [4].

Untuk memenuhi pelabelan terkait analisis sentimen, maka diperlukan suatu ulasan sebagai sumber data. Ulasan ini seringkali mengandung opini positif dan negatif tentang perubahan yang terjadi [26]. Pelabelan perlu untuk dilakukan, sebab dalam penelitian analisis sentimen, maka peneliti mencoba untuk mengklasifikasikan terkait apakah teks ulasan yang ditulis pengguna bersifat positif atau negatif [27]. Ulasan tersebut memberikan wawasan tentang dampak yang dihasilkan. Ini membantu dalam konteks analisis sentimen terkait ulasan teks untuk memahami pendapat yang terkait dengan masalah atau objek tertentu, apakah cenderung positif atau negatif [5]. Serta, dalam analisis sentimen merupakan teknik yang digunakan untuk mengetahui opini masyarakat terhadap suatu produk atau layanan berdasarkan ulasan tertulis yang beragam [28].

2.3 Rebranding pada Media Sosial

Rebranding merupakan proses strategis yang digunakan oleh bisnis atau institusi untuk membuat perubahan signifikan pada merek atau identitasnya dengan tetap memperhatikan hasil awal yang direncanakan. Hal ini termasuk melakukan berbagai upaya untuk mengubah reputasi atau citra yang sudah mapan. Inisiatif *rebranding* sering kali melibatkan berbagai komponen, seperti modifikasi logo, identitas, dan nilai perusahaan, yang mengubah esensi dan persepsi merek secara keseluruhan [1]. Hal ini dilakukan untuk membedakan merek dari pesaing, menarik pelanggan baru, dan mempertahankan atau meningkatkan ekuitas merek [29].

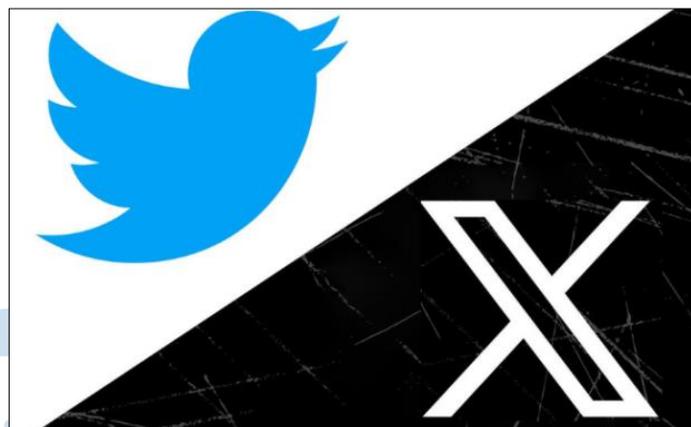
Tidak hanya mencakup modifikasi estetika, tetapi *rebranding* juga merupakan perubahan identitas, reputasi, atau citra merek. Perusahaan sering kali menggunakan taktik ini untuk memperluas pelanggannya, memposisikan diri secara berbeda di pasar, atau merespons perubahan preferensi konsumen.

Seperti ketika media sosial Twitter melakukan *rebranding* namanya menjadi "X", hal itu dilakukan sebagai tanggapan atas pembelian platform media sosial tersebut oleh Elon Musk. Dengan perombakan ini, identitas Twitter akan semakin diperluas dan mencakup menjadi aplikasi serbaguna sehingga tidak lagi hanya sekedar pesan singkat [9].



Gambar 2.1 Akuisisi merupakan rencana untuk membuat X sebagai aplikasi serbaguna
Sumber: [7]

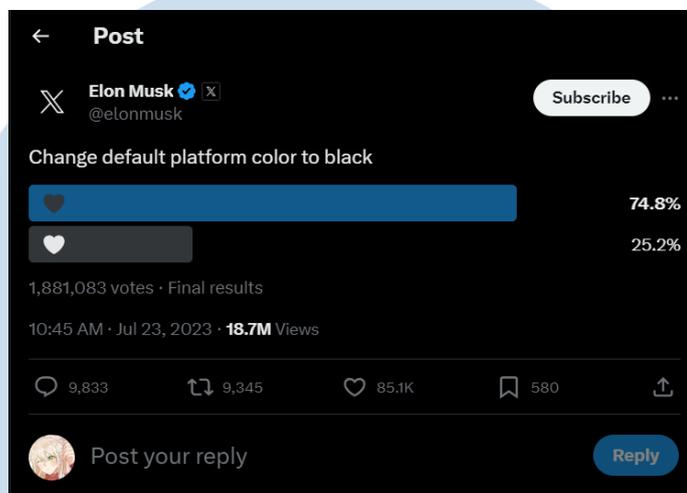
Bukan hanya nama saja yang telah terjadi akan *rebranding* pada Twitter, tetapi logo telah diubah menjadi "X" untuk mencerminkan *rebranding* yang telah terjadi.



Gambar 2.2 Perubahan akan logo dari Burung biru menjadi "X"
Sumber: [30]

Selain itu, Twitter atau "X" juga mengganti palet warna platformnya, dengan warna default yang sekarang adalah hitam berdasarkan hasil polling yang dilakukan oleh Elon Musk [23]. Penggantian palet warna platformnya

didasari akan respon perubahan preferensi yang telah ditawarkan oleh Elon Musk terhadap para pengguna lainnya.



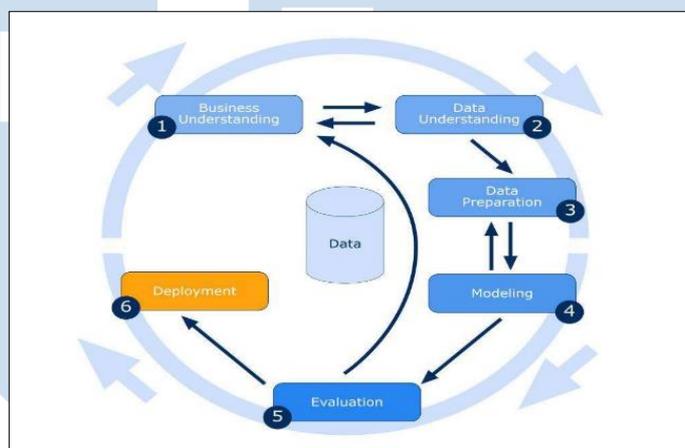
Gambar 2.3 Elon menanyakan preferensi perubahan warna pallet aplikasi terhadap para pengguna Twitter atau "X"
Sumber: [31]

Semua perubahan yang dilakukan oleh Twitter sesudah *rebranding*, seperti perubahan nama menjadi "X", penggantian logo menjadi "X", dan perubahan palet warna platform menjadi hitam sesuai dengan preferensi para pengguna Twitter, adalah bagian dari strategi *rebranding* yang telah dijalankan. Ini mencerminkan bagaimana *rebranding* dapat mencakup perubahan identitas yang mendalam, bertujuan untuk mengubah Twitter menjadi aplikasi serbaguna yang tidak hanya sekedar pesan singkat.

Melalui perubahan yang terjadi ketika *rebranding* dilakukan, Twitter berusaha untuk membedakan dirinya dari pesaing, menarik pelanggan baru, dan mempertahankan atau meningkatkan ekuitas merek yang telah dibangun selama bertahun-tahun. Konsep dasar dari *rebranding* itu sendiri dalam beberapa aspek telah terpenuhi dengan terjadinya *rebranding* pada Twitter. Selain itu, terjadinya *rebranding* tersebut merupakan respons yang cermat terhadap perubahan preferensi pengguna, yang seringkali menjadi salah satu alasan utama di balik keputusan *rebranding*. Hal ini dilakukan untuk memposisikan diri secara berbeda di pasar yang terus berubah, menjadikan Twitter atau "X" dalam era digital yang terus berkembang secara relevan dan komprehensif.

2.4 CRISP - DM

CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan *framework* atau metodologi salah satu model Data Mining [32]. Pendekatan terstruktur yang diberikan metodologi CRISP-DM menawarkan pengerjaan data mining yang baik. Seperti lebih efisien karena pengelolaan yang mudah akibat dari setiap proses memiliki keterhubungan antara satu dengan lainnya [33], dan dapat diandalkan [34].



Gambar 2.4 Enam Tahapan pada *framework* CRISP-DM
Sumber: [35]

Metodologi CRISP-DM terpecah menjadi enam tahapan diantaranya sebagai berikut:

- 1) Tahap *Business understanding*: Pada tahap pertama memiliki fokus utama terkait dengan pemahaman akan tujuan dan kebutuhan. Setelah mengetahuinya, pengetahuan yang sudah di *extract* sebelumnya akan dirubah menjadi permasalahan yang sedang terjadi. Disertakan juga dengan melakukan perencanaan untuk mendapatkan dan menyelesaikan dari tujuan awal [33].
- 2) Tahap *Data understanding*: Tahap kedua, diawali dengan melakukan pengambilan dan pengumpulan data. Setelah itu dilanjutkan dengan melakukan pemahaman terhadap data yang telah diperoleh. Tujuannya untuk mengetahui dan melihat pada berbagai masalah yang ditemukan seperti kualitas data, mencari pengetahuan dari data, serta mengetahui informasi tersembunyi dari data yang dimiliki [33].

- 3) Tahap *Data preparation*: Pada tahap selanjutnya data yang telah diketahui sebelumnya pada *data understanding*, akan dilakukan proses merubah data. Perubahan ini mencakup dari data mentah menjadi data bersih, dan prosesnya dilakukan sampai dengan data tersebut siap digunakan untuk tugas tahap selanjutnya [35].
- 4) Tahap *Modeling*: Setelah pada tahap *data preparation* berhasil diselesaikan dengan mendapatkan data bersih atau data final, maka tahap *modeling* akan dilaksanakan. Tahap modeling, dilakukan untuk mendapatkan hasil performa yang baik dengan cara mengaplikasikan teknik modeling yang dipilih dalam menyelesaikan tujuan awal [35].
- 5) Tahap *Evaluation*: Tahap selanjutnya adalah, *evaluation* yang dilakukan setelah tahap *modeling* berhasil dilakukan. Tujuannya adalah melakukan evaluasi terkait dengan hasil model yang telah digunakan. Serta menindaki lebih lanjut dengan mempertimbangkan apakah model tersebut dapat mencapai tujuan awal atau tidak [35]. Jika mencapai tujuan awal, maka model tersebut telah berhasil mencapainya serta dapat melanjutkan kepada tahap selanjutnya. Sementara itu jika tidak maka akan dilakukan peninjauan lebih lanjut dalam menangani permasalahan model evaluasi tidak mencapai tujuan awal [34].
- 6) Tahap *Deployment*: Tahapan terakhir ini, umumnya bukanlah akhir dari pengerjaan. Namun untuk menggapai apa yang telah didapati dari hasil tujuan awal yang telah terpenuhi dengan model dan evaluasi sebelumnya, maka penerapannya dapat dilakukan. Tujuan akhir yang umum dari tahapan *deployment* adalah penerapan kepada sisi pengguna akan hasil yang telah diperoleh pada tahapan sebelumnya. Pemilihan *deployment* perlu mempertimbangkan dengan baik akan tindakan apa yang perlu dilakukan guna memanfaatkan hasil model dan evaluasi sebelumnya kepada sisi pengguna [35].

Berikut merupakan alasan lain mengapa CRISP – DM dipilih dibandingkan dengan metode lain berada pada tabel di bawah ini:

Tabel 2.2 Perbandingan Metode CRISP-DM, KDD, dan SEMMA

CRISP DM [35]	KDD [36]	SEMMA [37]
Struktur pengerjaan terpecah menjadi 6 tahapan yang bertujuan mengetahui pemahaman bisnis disertai dengan tahapan dalam penerapannya bagaimana.	Struktur pengerjaan terpecah menjadi 5 tahapan yang bertujuan mengetahui pemahaman dari data.	Terfokus pada pengembangan model dengan 5 tahapan berbeda yang saling mendukung satu dengan lainnya.
Tahapan pengerjaan dapat bergerak dengan fleksibel antara mundur dan maju serta melakukan iterasi dengan sesuai kebutuhan.	Tidak memiliki aturan ketat tentang pengerjaan setiap tahapan antara bergerak maju mundur dan mengulangi langkah-langkahnya.	Selama proyek berlangsung, memiliki fleksibilitas yang bebas pada setiap tahapan.
Bagian Implementasi dapat diketahui dengan <i>deployment</i> sebagai model final.	Tidak eksplisit mencakup tahapan implementasi namun lebih fokus pada penerapan pengetahuan.	Fokus pada evaluasi model dan keputusan akhir kegunaan dan keandalannya.
Tahapan dimulai dengan Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Modeling, Evaluasi, Implementasi atau <i>deployment</i> .	Tahapan dimulai dari Pre-KDD, Seleksi, Pre-processing, Transformasi, Data Mining, Interpretasi/Evaluasi.	Awal mula tahapan yaitu Sampel, Jelajah, Modifikasi, Model, Evaluasi.

Mengacu pada Tabel 2.2, CRISP-DM menunjukkan keunggulan sebagai pendekatan yang menyeluruh dan terstruktur dalam proses analisis data, dimulai dari tahap pemahaman bisnis hingga ke tahap implementasi akhir (*deployment*). Framework ini juga dirancang secara fleksibel, memungkinkan perpindahan antar tahap secara iteratif sesuai kebutuhan analisis. Berbeda dengan KDD dan SEMMA yang cenderung menitikberatkan pada aspek teknis serta pengembangan model, CRISP-DM lebih unggul dalam memberikan panduan kerja yang sistematis dan adaptif. Hal ini menjadikannya pilihan utama sebagai kerangka konseptual dalam penelitian ini karena kemampuannya dalam menjembatani aspek teknis dan kebutuhan bisnis secara seimbang.

2.5 Text Mining dan Text PreProcessing

Text mining merupakan Metode mengekstraksi informasi yang diekstraksi secara otomatis dari sejumlah besar teks tidak terstruktur. Melalui pemrosesan, *text mining* dapat mengekstraksi informasi berharga dari teks. Untuk mempermudah pengumpulan informasi berkualitas tinggi, *text mining* juga dapat digunakan untuk menganalisis informasi dengan cepat seperti sentimen suatu pernyataan [38]. Tujuan dari *text mining* adalah untuk mengungkap

pengetahuan yang belum ditemukan sebelumnya, seperti informasi yang belum diketahui oleh siapapun dan belum pernah dicatat. Dalam penggunaannya, *text mining* merupakan teknik yang digunakan dalam menangani klasifikasi [39]. Terdapat tantangan yang diketahui ketika menangani klasifikasi teks. Seperti sifat data yang tidak terstruktur dan sulit untuk menangani, sehingga penggunaan *text mining* diperlukan [40].

Selama proses *text mining* hendak dilakukan, data *text* harus dipersiapkan terlebih dahulu. Proses mempersiapkan data *text* disebut dengan *text preprocessing*. Secara umum, *text preprocessing* merupakan teknik yang mengubah data teks tidak terstruktur menjadi data teks terstruktur. Beberapa langkah prapemrosesan teks yang umum diketahui [16][22][38][41]:

- 1) **Case Folding (Perubahan ukuran huruf)**: Tahap ini umumnya dilakukan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Hal ini membantu memastikan konsistensi dalam data dan mempermudah proses selanjutnya.
- 2) **Remove Number**: Menghapus semua angka dari teks.
- 3) **Remove Punctuation**: Menghapus semua tanda baca (misalnya titik, koma, tanda seru).
- 4) **Remove Emoticons**: Menghapus emotikon dan karakter emoji dari teks.
- 5) **Remove Special Characters**: Menghapus karakter non-ASCII (misalnya simbol "²", karakter khusus lainnya).
- 6) **Remove Whitespace**: Menghapus spasi yang tidak diperlukan dalam analisis sentimen.
- 7) **Tokenize (Tokenisasi)**: Tahap selanjutnya merupakan langkah kunci yang memecah teks menjadi kata-kata atau token. Tahap tokenisasi diperlukan untuk analisis lebih lanjut dan memahami struktur teks.
- 8) **Spelling Normalization (Normalisasi Ejaan)**: Normalisasi ejaan dapat digunakan jika data teks memiliki masalah ejaan yang signifikan. Hal

ini berguna untuk memperbaiki kesalahan ketik yang dapat memengaruhi hasil analisis.

- 9) **Remove Stop Words:** Pada tahap ini, kata-kata yang tidak relevan atau *stopword* dibuang. Dalam pengerjaannya membantu dalam fokus pada kata-kata kunci yang relevan.
- 10) **Stemming:** Proses *stemming* digunakan untuk mengembalikan kata-kata ke bentuk dasarnya. Ini berguna jika ingin menghitung kata-kata dasar dalam teks.

2.6 TF - IDF

TF-IDF atau *Term Frequency - Inverse Document Frequency* merupakan metode untuk menangani pembobotan kata yang berasal dari data teks [22]. Dalam mengolah data teks menjadi matriks kata, TF – IDF perlu dilakukan agar pemodelan yang dilakukan algoritma dapat dilakukan [42]. TF-IDF digunakan untuk menghitung bobot kata dalam dokumen, sehingga membantu mengidentifikasi kata-kata yang paling menonjol atau khas dalam analisis teks [43]. TF-IDF menghitung nilai pentingnya suatu kata berdasarkan dua komponen: frekuensi kemunculannya dalam dokumen tertentu (*TF*), dan seberapa jarang kata tersebut muncul di seluruh dokumen (*IDF*). Semakin sering sebuah kata muncul di dokumen tertentu namun jarang di dokumen lain, maka bobotnya akan semakin tinggi [44]. Proses ekstraksi fitur seperti TF-IDF merupakan bagian dari tahap *modeling*, karena hasil ekstraksi ini akan digunakan sebagai representasi numerik dalam pelatihan model klasifikasi [45].

2.7 Teknik Penanganan Ketidakseimbangan Data (SMOTE & ENN)

Ketidakseimbangan kelas atau *imbalanced* dalam dataset sering kali menjadi tantangan dalam pengembangan model machine learning, karena algoritma cenderung lebih sering mengklasifikasikan sampel ke dalam kelas mayoritas dibandingkan kelas minoritas. Fenomena ini dapat mengakibatkan model yang kurang akurat dalam mengenali pola dari kelas dengan jumlah sampel lebih sedikit. Untuk mengatasi permasalahan ini, pendekatan *Synthetic Minority Oversampling Technique* dan *Edited Nearest Neighbor* (SMOTE-

ENN) diterapkan sebagai metode gabungan yang mengombinasikan teknik oversampling dan undersampling guna menghasilkan distribusi kelas yang lebih seimbang [18].

Metode SMOTE-ENN terdiri dari dua proses utama yang saling melengkapi [46] [47]:

1. **SMOTE (Oversampling)**: Meningkatkan representasi kelas minoritas dengan menghasilkan sampel sintetis baru melalui interpolasi antara instance yang serupa. Teknik ini membantu mengurangi risiko *overfitting* yang sering terjadi pada metode oversampling konvensional.
2. **ENN (Undersampling)**: Menghapus sampel yang dianggap tidak representatif atau redundan dari kelas mayoritas menggunakan pendekatan Edited Nearest Neighbor (ENN). Proses ini bertujuan untuk mengurangi kemungkinan *noise* dalam data dan menghindari dominasi kelas mayoritas, sehingga menciptakan distribusi yang lebih seimbang.

2.8 Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi yang memiliki kemampuan untuk mengerjakan proses data teks dengan jumlah yang banyak. Hal ini disebabkan oleh kemampuannya dalam bekerja menggunakan teori probabilitas yang mempertimbangkan keberadaan suatu kata dalam salah satu target output yang diinginkan [22]. Dalam kinerjanya, algoritma ini memiliki karakteristik sebagai sederhana, efisien, dan populer untuk klasifikasi teks dengan performa baik di berbagai domain [16]. Terdapat beberapa jenis dari Naïve Bayes model diantaranya [48]:

- 1) **Multinomial Naïve Bayes** digunakan untuk merepresentasikan frekuensi dari terms seperti hitungan kata dalam dokumen serta umumnya digunakan dalam analisa teks klasifikasi.
- 2) **Bernoulli Naïve Bayes** umumnya digunakan untuk skenario di mana keberadaan atau ketiadaan suatu kata lebih sesuai ketimbang dengan frekuensinya.

Rumus yang digunakan dalam metode ini dinyatakan sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Rumus 2.1 Rumus Naïve Bayes
Sumber: [34]

Keterangan rumus 2.2 Naïve Bayes:

- 1) $P(H|X)$ merupakan probabilitas suatu hipotesis H yang diberikan data X (posterior probability).
- 2) $P(X|H)$ merupakan probabilitas mendapatkan nilai data X apabila hipotesis H benar (likelihood).
- 3) $P(H)$ merupakan probabilitas awal dari hipotesis H sebelum melihat data X (prior probability).
- 4) $P(X)$ merupakan probabilitas data X yang terjadi secara keseluruhan (evidence).

2.9 Support Vector Machine

Support vector machine merupakan salah satu teknik pembelajaran mesin yang kuat dan sering digunakan dalam bidang tertentu, seperti pengenalan pola, serta analisis sentimen [41]. Metode ini bertujuan untuk mengidentifikasi *hyperplane* optimal yang memisahkan dua kelas dalam ruang input [49]. Support vector machine mencari *hyperplane* terbaik dengan memisahkan data ke dalam dimensi tambahan yang optimal. Ketika metode ini digunakan dalam permasalahan klasifikasi yang melibatkan dua kelas berbeda, support vector machine menciptakan pemisah yang disebut *hyperplane* [21].

Persamaan *hyperplane* dalam support vector machine dapat didefinisikan sebagai berikut:

$$w \cdot x - b = 0$$

Rumus 2.2 Rumus hyperplane
Sumber: [17]

Keterangan rumus 2.3 Rumus hyperplane:

- 1) w merupakan bobot pada parameter
- 2) x merupakan input vektor
- 3) b adalah bias.

Beberapa kernel yang umumnya digunakan pada support vector machine adalah Linear, Polynomial, Radial Basis Function (RBF), dan Sigmoid [15]. Hasil akurasi yang paling baik diantara kernel lainnya akan dijadikan pemilihan penggunaan kernel dalam support vector machine.

2.10 Logistic Regression

Logistic Regression merupakan algoritma yang digunakan untuk klasifikasi maupun analisis prediktif. Algoritma ini bekerja apabila terdapat masalah dengan dua kemungkinan hasil. Fungsi Sigmoid digunakan pada algoritma ini untuk mengubah output menjadi probabilitas antara 0 dan 1. Kurva S merupakan bentuk dari fungsi sigmoid, memastikan bahwa hasil prediksi tetap berada dalam rentang yang dapat diinterpretasikan sebagai probabilitas [50]. Hasilnya bisa berupa dua pilihan apakah sebuah ulasan bersifat positif atau negatif (klasifikasi biner), atau dalam topik klasifikasi berita, model bisa membedakan teks ke dalam beberapa kategori seperti politik, olahraga, atau hiburan (klasifikasi multikelas) [17].

Persamaan logistic regression sebagai berikut [51]:

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1X$$

Rumus 2.3 Persamaan Logit (Fungsi Linear Logit Model)

- 1) w merupakan bobot pada parameter
- 2) p = Probabilitas kejadian $Y = 1$ (misalnya, kemungkinan sebuah teks memiliki sentimen positif).
- 3) $1-p$ = Probabilitas kejadian $Y = 0$ (misalnya, kemungkinan sebuah teks memiliki sentimen negatif).

- 4) $\left(\frac{p}{1-p}\right)$ = Odds atau rasio peluang, yaitu seberapa besar kemungkinan $Y=1$ dibandingkan $Y=0$.
- 5) $\ln()$ = Logaritma natural (basis e) dari rasio peluang. Ini digunakan agar model menjadi linear.
- 6) B_0 = **Intercept (konstanta)**, yaitu nilai awal ketika $X = 0$.
- 7) B_1 = Koefisien regresi, yaitu seberapa besar pengaruh variabel X terhadap peluang kejadian $Y = 1$.
- 8) X = Variabel independen (misalnya, jumlah kata positif dalam teks untuk analisis sentimen).

$$p = \frac{e(B_0 + B_1X)}{1 + e(B_0 + B_1X)}$$

Rumus 2.4 Fungsi Sigmoid (Probabilitas $Y=1$)

- 1) e = Bilangan eksponensial (~ 2.718), digunakan untuk membentuk kurva **S** (sigmoid).
- 2) $e(B_0 + B_1X)$ = Eksponensial dari hasil fungsi linear $B_0 + B_1X$.
- 3) $1 + e(B_0 + B_1X)$ = Normalisasi agar hasil selalu berada dalam rentang 0 hingga 1 (probabilitas).
- 4) p = Probabilitas kejadian $Y = 1$ setelah diterapkan fungsi sigmoid.

Untuk mendukung secara lebih rinci Tabel 2.3 merupakan perbandingan yang dapat diketahui mengenai 3 metode pembelajaran mesin klasifikasi:

Tabel 2.3 Perbandingan 3 Metode Pembelajaran Mesin Klasifikasi

Naïve Bayes	Support Vector Machine	Logistic Regression
Algoritma klasifikasi berbasis teori probabilitas yang bekerja dengan mempertimbangkan keberadaan suatu kata dalam kelas target tertentu [22].	SVM adalah metode pembelajaran yang mencari hyperplane optimal untuk memisahkan kelas dengan margin terbesar, sehingga meningkatkan akurasi klasifikasi dan efisien pada data berdimensi tinggi [52].	Algoritma yang digunakan untuk klasifikasi biner maupun multikelas. Menggunakan fungsi sigmoid untuk mengubah output menjadi probabilitas antara 0 dan 1 [17].

Dapat mengerjakan tugas klasifikasi teks terkait dengan analisis sentimen [15].	Dapat menunjang proses klasifikasi teks dalam tugas analisis sentimen [41].	Memiliki kemampuan dalam tugas klasifikasi teks terkait dengan analisis sentimen [51].
Menghasilkan prediksi berbasis probabilitas yang dapat digunakan untuk mengukur tingkat kepastian suatu teks termasuk dalam kelas tertentu [23].	Menghasilkan skor keputusan (decision score) yang menentukan sejauh mana suatu sampel berada dalam satu kelas dibandingkan kelas lainnya [21].	Dalam klasifikasi teks, algoritma ini dapat digunakan untuk menentukan apakah suatu ulasan bersifat positif atau negatif [50].

Merujuk pada Tabel 2.3, dapat disimpulkan bahwa ketiga algoritma pembelajaran mesin yang dikaji Naïve Bayes, Support Vector Machine (SVM), dan Logistic Regression memiliki karakteristik unggulan masing-masing dalam penerapan analisis sentimen berbasis teks. Naïve Bayes mengandalkan prinsip probabilistik yang ringan secara komputasi namun tetap efektif, terutama dalam menangani volume data yang besar. Sementara itu, SVM unggul dalam hal akurasi melalui pendekatannya yang mencari hyperplane optimal untuk memisahkan kelas secara maksimal.

Serta pada Logistic Regression memiliki perbedaan karena menghasilkan probabilitas yang dapat diinterpretasikan secara kuantitatif, sehingga mendukung kejelasan dalam proses analisis hasil. Ketiga metode ini dianggap relevan untuk diterapkan dalam penelitian ini, mengingat kemampuannya dalam menangani karakteristik data ulasan pengguna dari Google Play Store, sebagaimana telah didukung oleh temuan dari berbagai penelitian sebelumnya.

2.11 Hyperparameter Tuning

Hyperparameter Tuning merupakan proses menyesuaikan nilai hyperparameter dalam model algoritma agar menghasilkan performa yang optimal. Salah satu metode yang umum digunakan untuk melakukan hal ini adalah Grid Search, yang mencoba berbagai kombinasi hyperparameter dan menggunakan *cross-validation* untuk mengevaluasi performanya. Grid Search Cross Validation secara otomatis menguji dan membandingkan berbagai konfigurasi hyperparameter guna menemukan kombinasi terbaik yang menghasilkan model dengan performa prediksi tertinggi [51].

Proses pencarian Grid dilakukan dengan mengeksplorasi berbagai kemungkinan nilai untuk setiap hyperparameter yang telah ditentukan sebelumnya. Setiap kombinasi nilai tersebut akan diuji secara sistematis untuk menentukan konfigurasi terbaik bagi model. *Grid search* merupakan cara terstruktur untuk mencoba berbagai kombinasi pengaturan dalam model agar menemukan yang paling baik. Pengaturan ini disebut *hyperparameter*, dan pemilihan yang tepat dapat membuat model bekerja lebih akurat. Dengan mengoptimalkan pengaturan hyperparameter membantu mengevaluasi setiap kombinasi pengaturan berdasarkan ukuran keberhasilan, seperti akurasi atau skor F1 [53].

2.12 Confusion Matrix

Pengukur performa pada masalah klasifikasi pembelajaran mesin bernama Confusion Matrix. Penggunaan confusion matrix dilakukan untuk mengetahui sebaran kebenaran data prediksi terhadap data aktual [54]. Confusion Matrix memberikan perbandingan antara hasil klasifikasi yang dibuat oleh model dengan hasil klasifikasi yang sebenarnya. Dalam penggunaan confusion Matrix, berisikan akan seperti [55]:

1) TP (*True Positive*)

Nilai prediksi benar diprediksi sebagai *True Positive*.

2) FP (*False Positive*)

Nilai prediksi salah memprediksi positif sebenarnya.

3) FN (*False Negative*)

Nilai positif telah diprediksi sebagai negatif.

4) TN (*True Negative*)

Nilai yang diprediksi dengan benar bahwa sebagai negatif yang sebetulnya.

Terdapat matrik umum untuk memperoleh masalah klasifikasi informasi berharga tentang kinerja metode yang telah di implementasikan. Seperti Precision berfokus pada keakuratan prediksi positif, Recall mengukur kemampuan metode dalam mendeteksi data positif, F1-Score adalah

pengukuran kombinasi dari Precision dan Recall, dan Accuracy mengukur sejauh mana keseluruhan data diklasifikasikan dengan benar [55]. Untuk melakukan perhitungan presisi, recall, F1 – score dan akurasi dilakukan dengan menggunakan persamaan confusion matrix seperti [54]:

$$\frac{TP}{(TP + FP)}$$

Rumus 2.5 Rumus precision

$$\frac{TP}{(TP + FN)}$$

Rumus 2.6 Rumus recall

$$\frac{2 * Precision * Recall}{Precision + Recall}$$

Rumus 2.7 Rumus F1 – Score

$$\frac{(TP + TN)}{(TP + TN + FP + FN)} 100\%$$

Rumus 2.8 Rumus Accuracy

Keterangan rumus 2.6 – 2.9:

- 1) **TP (True Positive)** merupakan nilai yang diprediksi sebagai positif yang benar-benar positif.
- 2) **FP (False Positive)** merupakan nilai yang salah diprediksi sebagai positif, padahal sebenarnya negatif.
- 3) **FN (False Negative)** merupakan nilai yang salah diprediksi sebagai negatif, padahal sebenarnya positif.
- 4) **TN (True Negative)** adalah nilai yang diprediksi sebagai negatif yang benar-benar negatif.
- 5) **Precision** merupakan berapa banyak dari prediksi positif yang benar positif.
- 6) **Recall** merupakan Berapa banyak data positif yang berhasil dideteksi oleh metode.

2.13 Python

Bahasa pemrograman tingkat tinggi yang mudah digunakan, open source, dan menyediakan berbagai macam *library* yang lengkap bernama Python [56]. Python memiliki fitur sistem tipe dinamis dan manajemen memori otomatis, dengan perpustakaan besar atau *library* yang banyak dan komprehensif untuk komputasi ilmiah dan analisis data. Bahasa pemrograman ini kerap kali digunakan dalam menyelesaikan permasalahan *Machine Learning* atau pembelajaran mesin dan *data science*. Alasannya adalah bahasa pemrograman ini memiliki tujuan yang umum dalam skala kecil maupun besar pada lingkup penyelesaian penelitian, pengembangan dan produksi [57].

Tabel 2.4 Perbandingan Bahasa Pemrograman Python dengan R

Python	Bahasa R
Python merupakan bahasa pemrograman yang sering digunakan melakukan tugas analisis data, dan pembelajaran mesin [57].	Bahasa pemrograman R merupakan bahasa yang dibangun untuk mengerjakan analisis statistik, dan visualisasi [57].
Python memiliki dukungan <i>library</i> yang lengkap termasuk <i>scraping data</i> yang berasal dari Google Play Store [58], [59]. <i>Library</i> dengan dukungan untuk pengerjaan klasifikasi teks dan pengerjaan pembelajaran mesin [48][52].	Dukungan <i>library</i> yang dimiliki oleh bahasa pemrograman R termasuk lengkap dalam ranah analisis statistik dan visualisasi [60]. Serta dukungan pengerjaan klasifikasi teks [61].

Merujuk pada Tabel 2.4, Python dan R sama-sama mendukung proses analisis data secara signifikan, namun memiliki keunggulan di bidang yang berbeda. Python unggul dalam efisiensi eksekusi, fleksibilitas untuk berbagai tugas pembelajaran mesin, serta ekosistem pustaka yang luas dan kemampuan integrasi data dari berbagai sumber seperti Google Play Store.

Sementara itu, R tetap menjadi pilihan utama untuk analisis statistik mendalam dan visualisasi eksploratif, berkat fitur statistiknya yang kaya dan berorientasi pada representasi data. Dengan mempertimbangkan fokus penelitian ini pada klasifikasi teks berbasis ulasan pengguna dan kebutuhan integrasi pustaka serta efektivitas pengembangan model, Python dipilih sebagai bahasa pemrograman yang paling sesuai untuk implementasi pembelajaran mesin.

2.14 Jupyter Notebook

Jupyter merupakan *software*, open source, dan layanan web untuk komputasi interaktif pada banyak bahasa pemrograman [62]. Terdapat banyak proyek didalamnya dan salah satunya adalah Jupyter Notebook. Proyek Jupyter Notebook dibangun dengan format seperti buku catatan yang dapat melakukan komputasi yang interaktif, cepat, dapat mengeksplorasi data untuk divisualisasikan dan masih banyak lagi kegunaannya [63]. Jupyter Notebook pada dasarnya memungkinkan pengguna untuk pengeditan skrip dalam browser, eksekusi skrip, dan menampilkan hasil komputasi dengan representasi media pada Notebook.

Berikut merupakan komparasi lebih lengkapnya terkait dengan Jupyter, PyCharm, dan Spyder:

Tabel 2.5 Perbandingan antara Jupyter, PyCharm, dan Spyder
Sumber: [64]

Jupyter	PyCharm	Spyder
Sangat baik untuk data science, output interaktif pada notebook, mendukung banyak bahasa pemrograman lainnya.	Baik untuk data science karena telah terintegrasi dengan berbagai paket library.	Cocok untuk data science, tersedia alat analisis statis.
Gratis bagi siapapun, bersifat <i>open source</i> .	Versi Komunitas gratis dengan limitasi beberapa fitur, versi profesional berbayar tanpa adanya limitasi fitur.	Gratis bagi siapapun dan bersifat <i>open source</i> .
Jupyter cocok untuk pengerjaan dengan lingkup pengerjaan kolaboratif.	PyCharm memiliki dukungan untuk pengerjaan kolaboratif, namun fiturnya terbatas.	Spyder tidak cocok untuk kerja kolaboratif.
Tata letak yang mudah dipahami, serta secara visual menampilkan pada notebook dengan baik.	Tata letak yang sederhana, serta dapat disesuaikan sesuai preferensi.	Antarmuka pengguna Spyder lebih sederhana dan dasar dibandingkan dengan Jupyter dan PyCharm

Berdasarkan Tabel 2.5, dapat diketahui bahwa Jupyter, PyCharm, dan Spyder masing-masing memiliki karakteristik dan keunggulan tersendiri dalam mendukung proses pengembangan berbasis data science. Jupyter unggul dalam mendukung interaktivitas dan eksplorasi data secara visual melalui format notebook yang intuitif. Pada sisi lain, PyCharm memberikan pengalaman pengembangan yang kuat dengan integrasi pustaka yang luas, namun terdapat beberapa fitur lanjutan hanya tersedia dalam versi berbayar.

Sementara itu, Spyder dikenal dengan antarmukanya yang ringan dan efisien serta fitur visualisasi yang cepat. Dengan mempertimbangkan kebutuhan penelitian ini yang melibatkan eksplorasi data, visualisasi, serta pengembangan model pembelajaran mesin secara interaktif, Jupyter dipilih sebagai lingkungan pengembangan utama yang digunakan dalam proyek ini.

2.15 Google Play Store

Google play store merupakan platform distribusi resmi yang dimiliki oleh Android. Para pengguna memiliki keleluasan dalam menikmati mendapatkan aplikasi, games dan juga digital content [65]. Kehadiran google play store, telah terinstall sebelumnya dengan perangkat khususnya adalah android yang mendukung google play store [66]. Para pengguna, juga memiliki kebebasan dalam memberikan ulasan pada semua aplikasi yang tersedia dalam google play store. Untuk memberikan ulasan di Google Play Store, pengguna dapat mengikuti langkah-langkah beriku [67]:

- 1) Pastikan telah mengunduh dan menginstal aplikasi yang ingin diulas. Ulasan hanya dapat diberikan untuk aplikasi yang sudah diunduh dan terpasang. Para pengguna harus memastikan bahwa sudah menggunakan akun Google yang benar.
- 2) Buka aplikasi Google Play Store pada perangkat Android.
- 3) Telusuri atau cari aplikasi yang ingin diulas.
- 4) Temukan dan pilih aplikasi tersebut untuk membuka halaman detailnya.
- 5) Jika ingin meninggalkan ulasan tertulis, pengguna dapat mengetuk "Write a review."
- 6) Ikuti petunjuk yang muncul di layar untuk menulis ulasan dan tambahkan detail yang relevan.
- 7) Setelah menulis ulasan, pengguna dapat mengetuk "Post."

Pada setiap ulasan, para pengguna tidak dapat memberikan ulasan dari akun enterprise seperti akun kerja atau sekolah. Jika salah satu akun di perangkat

pengguna terdaftar dalam program beta untuk sebuah aplikasi, mereka juga tidak dapat memberikan ulasan untuk aplikasi tersebut.

2.16 Streamlit

Streamlit merupakan kerangka kerja open source berbasis Python yang dirancang untuk membantu ilmuwan data serta insinyur AI/ML dalam membangun aplikasi data interaktif secara cepat dan efisien. Dengan hanya beberapa baris skrip, pengguna dapat membuat dan menerapkan aplikasi data yang dinamis tanpa memerlukan pengalaman mendalam dalam pengembangan web. Kemudahan penggunaan serta fleksibilitasnya menjadikan Streamlit sebagai salah satu pilihan utama dalam pengembangan aplikasi berbasis data (Deployment) [68].

