

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Sebelum penelitian ini dibuat, terdapat beberapa penelitian serupa yang sudah pernah dilakukan sebelumnya. Beberapa komponen dari berbagai penelitian tersebut kemudian dirajuk agar sebagai dasar dari penelitian ini. Kumpulan penelitian yang menjadi referensi akan dipaparkan dihilat perbedaannya dengan penelitian ini

Tabel 2. 1 Penelitian Terdahulu

UMMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

Penulis	Jurnal	Judul	Tahun	Hasil
Heiko Heidenreich, Geoff Dickson, Nicolas Strohschneid & Markus K urscheidt	<i>Leisure Studies, Vol 1, No 15</i>	<i>Exploring oppositionality and support of Counter-Strike redditors for the World esportsAssociation[4]</i>	2023	Komunitas esports belum merangkul berbagai organisasi pengatur esports domestik dan internasional yang mengaku-ngaku, dan terkenal tidak ramah terhadap organisasi yang dianggap memanfaatkannya. Analisis sentimen ini didasarkan pada tinjauan 5359 komentar (disebar di 29 threads dan lima subreddit) yang dibuat oleh pengguna Counter-Strike di Reddit. Disimpan 861 komentar . Temuan utama penelitian ini adalah bahwa sentimen negatif lebih dominan daripada sentimen positif.
Zhang, Anqi	<i>Electronic Theses and Dissertations . Vol 25, No 2</i>	<i>USING THE RECOMMENDATION OUTCOME OF STEAM GAME REVIEWS AS SENTIMENT PREDICTOR[24]</i>	2022	Dari hasil tersebut, ditemukan bahwa fitur Keywords saja menghasilkan tingkat kesalahan terkecil di semua 3 metrik kesalahan, dengan MAE paling rendah sebesar 2.3504e-16 dari MLR. Hasil ini mungkin disebabkan oleh MLR yang merupakan model regresi paling sederhana tanpa kompleksitas tambahan. Sepertinya semakin kompleks, kinerja model menurun, terutama dengan MLPR. Hal ini mungkin disebabkan oleh MLPR, sebagai desain jaringan saraf, bukan model regresi linear dan mungkin mengandung terlalu banyak fitur yang tidak perlu yang menyebabkan lebih banyak kesalahan.
K. Khadijah, N. Sabillyand F. A. Nugroho	<i>Jurnal Ilmiah KURSOR Vol 12, No 1</i>	<i>Sentiment Analysis Of League Of Legends: Wild Rift Reviews On Google Play Using Naive Bayes Classifier[25]</i>	2023	League of Legends: Wild Rift adalah game mobile dengan lebih dari 48 juta unduhan. Keuntungan diperoleh dari penjualan item dalam game. Kesan pemain pada minggu pertama menentukan pembelian dalam aplikasi. Penelitian ini menggunakan analisis sentimen untuk mempelajari opini pemain berdasarkan ulasan di Google Play Store menggunakan algoritma Naive Bayes Classifier (NBC). Ditemukan bahwa stemming menurunkan akurasi, tetapi transformasi kata informal menjadi formal meningkatkan performa. Ekstraksi fitur BOW bigram memberikan hasil terbaik dengan akurasi 79,3%, presisi 82,10%, recall 83,50%, dan f1-score 82,8%.

Penulis	Jurnal	Judul	Tahun	Hasil
Taylor, Emma.	<i>Master of Science (MS).</i>	<i>An Investigation on the Pricing of Virtual Items & Digital Commodities: Evidence From the Counter-Strike: Global Offensive Market[26]</i>	2022	Penelitian ini mengeksplorasi pasar skin CS:GO dalam konteks pasar virtual lainnya seperti barang video game dan cryptocurrency. Pasar skin CS:GO menarik karena kompleksitasnya: harga bervariasi, informasi detail tersedia, dan pembelian terjadi dalam peledangan peer-to-peer. Penelitian menggunakan model pembelajaran mesin menunjukkan bahwa harga historis lebih baik dalam memprediksi harga saat ini dibandingkan atribut skin. Ini menyarankan bahwa skin CS:GO adalah aset spekulatif mirip cryptocurrency, bukan berdasarkan preferensi individu. Penelitian lanjutan bisa menggunakan analisis sentimen untuk memahami motivasi pembeli.
Suswadi, Moh. Erkamim	<i>ILKOM Jurnal Ilmiah. Vol 15, No 3</i>	<i>Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine[29]</i>	2023	Selama wabah COVID-19, pasar Indonesia, termasuk aplikasi Shopee, sangat terdampak. Evaluasi fitur dan layanan Shopee diperlukan melalui ulasan di Google Play Store. Penelitian menggunakan data Kaggle dari ulasan Shopee dan analisis sentimen dilakukan dengan metode Support Vector Machine (SVM) dan Random Forest. Hasilnya, model Random Forest memiliki akurasi 82,21%, sedangkan model SVM memiliki akurasi lebih tinggi yaitu 84,71%. Kata-kata umum pada sentimen positif seperti “belanja” dan “barang”, sedangkan pada sentimen negatif seperti “ongkir” dan “kirim”, digunakan untuk evaluasi kualitas Shopee.
N.F.N, Bahrawi	<i>JITU Vol 2 No 2</i>	<i>Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based[31]</i>	2019	Penelitian ini menerapkan analisis sentimen menggunakan data dari Twitter dengan algoritma Random Forest. Evaluasi dilakukan terhadap hasil algoritma yang digunakan dalam penelitian ini, dengan tingkat keakuratan sekitar 75%.
N.F.N, Bahrawi	<i>J. Penelitian Pos dan Informasi, vol. 9, no. 1</i>	<i>Online Realtime Sentiment Analysis Tweets by Utilizing Streaming API Features From Twitter[32]</i>	2019	Penelitian ini bertujuan untuk menciptakan sistem yang mampu menganalisis sentimen dari platform media sosial Twitter secara langsung dan otomatis tanpa henti. Dalam percobaan, sistem berhasil mengumpulkan data, menganalisis sentimen, dan menampilkan hasilnya dalam bentuk grafik secara langsung melalui web yang terus

Penulis	Jurnal	Judul	Tahun	Hasil
				diperbarui secara otomatis. Selanjutnya, penelitian akan difokuskan pada meningkatkan keakuratan algoritma yang digunakan dalam analisis sentimen tersebut.
A. Mardjo and C. Choksuchat	<i>IEEE Access, Vol 10</i>	<i>HyVADRF: Hybrid VADER–Random Forest and GWO for Bitcoin Tweet Sentiment Analysis[34]</i>	2022	Model HyVADRF (VADER - random forest) dengan gray wolf optimizer (GWO). VADER digunakan untuk menghitung skor polaritas dan mengklasifikasikan sentimen, sementara hutan acak digunakan sebagai pengklasifikasi. Dari 3,6 juta tweet, model HyVADRF menunjukkan akurasi 75,29%, presisi 70,22%, recall 87,70%, dan F1-score 78%. Ukuran dataset ideal adalah 90% dari total tweet (1.249.060). Model ini memberikan hasil yang stabil dengan standar deviasi rendah.
W. Alinda Wahyuni, S. Saepudin, and F. Sembiring	<i>Mantik, Vol5, No 4</i>	<i>Sentiment Analysis of Online Investment Applications on Google Play Store using Random Forest Algorithm Method[35]</i>	2022	Penelitian ini bertujuan untuk menganalisis sentimen dari review aplikasi investasi online seperti Stockbit, Hsb Investment, dan Bibit. Metode yang digunakan dalam penelitian ini adalah metode Random Forest. Berdasarkan analisis dan pengujian yang dilakukan, kesimpulan yang diperoleh adalah implementasi algoritma Random Forest pada aplikasi Stockbit menghasilkan nilai akurasi sebesar 62.50%, aplikasi Bibit menghasilkan nilai akurasi sebesar 63.39%, dan aplikasi HSB menghasilkan nilai akurasi sebesar 96.25%.
N. Wilim and R. Oetama	<i>IJNMT (International Journal of New Media Technology), vol. 8, no. 1</i>	<i>Sentiment Analysis About Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, And Decision Tree[28]</i>	2019	Penelitian ini membandingkan antara analisis sentimen dari acara TV <i>Indonesia Lawyers Club</i> dan juga Mata Najwa tahun 2018 & 2019 dengan algoritma KNN, <i>Naive Bayes</i> , dan <i>Decision Tree</i> . Hasil penelitian <i>Naive Bayes</i> memiliki akurasi tertinggi dengan 66,94% di ILC dan 70% di mata Najwa untuk data 2018, sedangkan apabila dalam kasus data 2019, algoritma KNN berhasil memperoleh akurasi tertinggi, yaitu 76,94% untuk ILC dan juga 71,94% untuk Mata Najwa
C. Destitus, W. Wella, and	<i>Ultima InfoSys : Jurnal Ilmu</i>	<i>Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di</i>	2020	Identifikasi <i>tweet cyberbullying</i> menggunakan metode <i>Support Vector Machine</i> mencapai tingkat akurasi yang signifikan, mencapai akurasi sebesar 80%, presisi sebesar 75.1%, <i>recall</i> sebesar 96%, dan <i>f-measure</i>

Penulis	Jurnal	Judul	Tahun	Hasil
S. Suryasari	Sistem Informasi, vol. 11, no. 2	<i>Twitter Indonesia</i> [27]		sebesar 85%. Selain itu, dalam pengujian threshold seleksi fitur menggunakan Information Gain, juga mencapai nilai yang tinggi dengan akurasi sebesar 86%, presisi sebesar 81%, <i>recall</i> sebesar 95%, dan <i>f-measure</i> sebesar 87%. Dengan demikian, hasil identifikasi <i>tweet cyberbullying</i> dengan kedua metode menunjukkan kinerja yang optimal.

UMMN

Berdasarkan penelitian yang dicantumkan pada Tabel 2.1, dapat ditemukan bahwa dari setiap artikel yang terdaftar belum ada yang menggunakan ulasan dari *platform Steam*, dimana hal ini dikarenakan sebagian besar penelitian dilakukan dengan data platform *Google Play*[26][29][35] ataupun *Twitter*[27][32][34]. Selain itu, penelitian yang mendalam terkait permainan *Counter Strike 2* dari segi mekanisme permainan juga belum dapat ditemukan, dikarenakan beberapa artikel hanya membahas sisi pasar dan penyelenggaraan kompetisi[4][26]. Terakhir, seluruh penelitian yang dilakukan hanya mengklasifikasikan sentimen berdasarkan komentar saja tanpa melihat keputusan *review* dari pemain. Seluruh hasil tingkat pengukuran dari berbagai algoritma yang dipakai pada penelitian terdahulu juga sebagian besar masih belum ada yang berada diatas 80% ataupun memiliki perbedaan akurasi yang sangat jauh pada saat membandingkan model[28][35].

Dengan demikian, dapat disimpulkan bahwa penelitian ini dapat terautentikasi kebaruannya dikarenakan penelitian ini adalah penelitian satu-satunya yang membahas mengenai analisis sentimen *Counter Strike 2* dari segi mekanisme permainan melalui klasifikasi sentimen terhadap komentar Steam sekaligus juga dari hasil keputusan rekomendasinya dengan menggunakan dan membandingkan empat algoritma klasifikasi yang berbeda-beda, yaitu *SVM*, *Naive Bayes*, *Logistic Regression*, dan *Random Forest*.

2.2 Teori tentang Topik Skripsi

2.1.1 Video Game *First Person Shooter (FPS)*

First-Person Shooter (FPS) adalah *genre* permainan video yang memberikan pengalaman visual dari sudut pandang karakter utama, di mana pemain melihat dan berinteraksi dengan dunia permainan seolah-olah mereka berada dalam tubuh karakter tersebut. Ciri khasnya adalah perspektif orang pertama (dari mata karakter), di mana senjata atau tangan karakter terlihat di layar, memberikan pengalaman yang lebih langsung dan immersif dalam pertempuran dan aksi[7].

Sejarah *FPS* dapat ditelusuri kembali ke era 1970-an dan 1980-an. Salah satu permainan yang dianggap sebagai pelopor genre ini adalah

Maze War, yang dirilis pada tahun 1974[8]. Permainan ini memungkinkan pemain untuk bergerak dalam labirin tiga dimensi dan menembak pemain lain menggunakan senjata.

Namun, pencapaian besar dalam *genre FPS* dimulai pada tahun 1992 dengan rilis *Wolfenstein 3D* yang dikembangkan oleh *Id Software*. Game ini merupakan permainan tembak-menembak pertama yang menggunakan teknologi grafis 3D untuk menciptakan dunia permainan dalam perspektif orang pertama. Kemudian, *id Software* melanjutkan kesuksesannya dengan merilis *Doom* pada tahun 1993, yang menjadi titik puncak popularitas dan pengaruh dalam *genre* ini[9].

Pada pertengahan hingga akhir 1990-an, perkembangan teknologi komputer dan grafis semakin memungkinkan pengembang untuk membuat pengalaman FPS yang lebih kompleks dan realistis. "Quake" (1996) dari *id Software* menjadi salah satu inovator dengan memperkenalkan gameplay multipemain online yang dinamis. Hal ini tidak hanya membuka pintu untuk komunitas permainan yang kuat dan beragam, tetapi juga mengubah lanskap permainan video dengan memperkenalkan dimensi sosial yang baru dan menantang. Dengan koneksi internet yang semakin umum di rumah-rumah, "Quake" menandai awal dari apa yang kemudian menjadi fenomena budaya dan industri yang dikenal sebagai permainan daring (online). Hal ini memungkinkan para pemain untuk berkompetisi secara langsung dengan orang lain dari seluruh dunia, menciptakan komunitas yang kuat dan dinamis serta memperluas pengalaman bermain mereka secara signifikan[10].

Dalam permainan *FPS*, pengguna memiliki kendali penuh atas pergerakan karakternya, termasuk gerakan maju, mundur, berbelok, dan melompat. Pemain juga dapat mengendalikan arah pandang, memungkinkan mereka untuk menyesuaikan penempatan senjata dan mengarahkan tembakan ke berbagai arah. Peta permainan biasanya

dirancang untuk memberikan berbagai situasi pertempuran dan tantangan taktis.

Tujuan umum dalam permainan *FPS* adalah menyelesaikan misi tertentu, seperti menyelamatkan dunia dari ancaman, menyelesaikan skenario tertentu, atau mencapai target spesifik. Seiring dengan itu, pertempuran dan pertarungan melibatkan penggunaan senjata dan peralatan yang bervariasi, termasuk senjata api, granat, dan peralatan khusus. *FPS* juga sering menawarkan mode permainan multipemain online, di mana pemain dapat bergabung dalam pertempuran melawan pemain lain secara *real-time*. Mode multipemain ini memberikan pengalaman sosial dan kompetitif yang intens, seringkali dengan elemen seperti pembangunan karakter, pengembangan kemampuan, dan kompetisi tim[11]. Contoh-permainan *FPS* yang terkenal melibatkan penggunaan senjata modern atau futuristik dan memungkinkan pemain untuk berpartisipasi dalam pertempuran skala besar atau pertarungan satu lawan satu yang intens. Beberapa contoh permainan *FPS* yang sangat populer termasuk *Counter-Strike*, *Call of Duty*, *Battlefield*, dan *Halo*.

FPS telah menjadi salah satu genre paling dominan dan paling dicari dalam industri permainan video, menarik pemain dengan kecepatan aksi dan ketegangan yang intens.

2.1.2 Counter Strike

Counter-Strike, lahir sebagai modifikasi dari permainan *Half-Life* pada tahun 1999, dikembangkan oleh Minh "Gooseman" Le dan Jess "Cliffe" Cliffe, telah menjadi salah satu seri permainan video yang paling ikonik dan sukses di dunia. Dalam *Counter-Strike*, pemain dibagi menjadi dua tim utama, yakni Teroris dan *Counter-Terrorists*, dengan setiap tim memiliki tujuan dan misi yang berbeda-beda tergantung pada skenario atau peta permainan tertentu. Teroris berusaha menyelesaikan misi seperti meletakkan bom, sementara *Counter-Terrorists* berusaha mencegah atau mengatasi tujuan tersebut[12].

Ada beberapa mode permainan utama, termasuk *Bomb Defusal*, di mana Teroris mencoba meletakkan bom dan *Counter-Terrorists* berusaha mencegahnya, serta "*Hostage Rescue*", di mana *Counter-Terrorists* harus menyelamatkan sandera dari Teroris. Kedua mode ini menekankan kerja sama tim, strategi, dan keterampilan tembak-menembak. Pemain memiliki akses ke berbagai senjata dan peralatan yang dapat dibeli menggunakan mata uang dalam permainan. Mulai dari senjata api, granat, hingga peralatan pelindung, pemilihan peralatan menjadi kunci dalam strategi tim dan individu.

Keberhasilan *Counter-Strike* tidak hanya terletak pada pengembangannya oleh Valve, tetapi juga pada kontribusi besar dari komunitas pemain. Modifikasi dan pengembangan peta oleh komunitas menciptakan variasi yang melimpah dalam pengalaman permainan, menjadikan *Counter-Strike* bukan sekadar permainan, tetapi juga sebuah *platform* yang terus berkembang.

Counter-Strike telah menjadi salah satu pionir dalam dunia *e-sports*. Turnamen besar seperti *Major Championships* menarik perhatian jutaan pemirsa di seluruh dunia. Keberhasilan permainan ini dalam dunia *e-sports* tidak hanya menciptakan legenda dalam kompetisi, tetapi juga membentuk dinamika global dalam industri permainan video. *Counter Strike 2* merupakan iterasi terkini dari rangkaian permainan ini dan tetap mempertahankan dominasinya sebagai salah satu *game* paling populer dan berpengaruh di dunia *e-sports* dengan komunitas pemain yang sangat besar dimana setiap pemain memiliki tanggapan dan sentimen yang berbeda-beda yang tersebar di seluruh media sosial[13].

2.1.3 Analisis Sentimen

Sentiment Analysis atau Analisis Sentimen adalah suatu teknik pengolahan bahasa alami yang bertujuan untuk mengekstrak, mengidentifikasi, dan menilai sentimen atau opini dari teks. Tujuan utamanya adalah untuk memahami perasaan atau sikap yang terkandung

dalam teks, baik itu positif, negatif, atau netral. *Sentiment analysis* juga dikenal dengan sebutan *opinion mining*[14].

Terdapat beberapa aspek dan konsep teoritis yang terlibat dalam *sentiment analysis*. Pertama, analisis ini mengandalkan teknik-teknik Pengolahan Bahasa Alami (*Natural Language Processing - NLP*) untuk memahami dan memproses teks manusia. Ini mencakup pemrosesan token, analisis sintaksis, dan pemahaman konteks kata dalam kalimat[28]. Representasi teks menjadi kunci dalam *sentiment analysis*. Dalam analisis sentimen, teks perlu diubah ke dalam representasi yang dapat dimengerti oleh algoritma mesin. Ini sering melibatkan vektor kata atau metode *embedding*, di mana kata-kata direpresentasikan dalam bentuk vektor numerik.

Selanjutnya, analisis sentimen sering dianggap sebagai masalah klasifikasi, di mana teks diklasifikasikan menjadi kategori sentimen tertentu, seperti positif, negatif, atau netral. Model klasifikasi, seperti mesin *vector support (SVM)* atau *deep learning*, dapat digunakan untuk mencapai tujuan ini[15]. Sumber sentimen menjadi fokus penting dalam teori analisis ini. Sentimen dapat muncul dari berbagai sumber, termasuk teks ulasan pengguna, media sosial, artikel berita, atau bahkan transkripsi wawancara.

Aspek subjektivitas juga menjadi pertimbangan, karena beberapa teks mungkin lebih subjektif daripada yang lain. Model analisis sentimen harus dapat memahami dan mengatasi subjektivitas untuk menghasilkan hasil yang lebih akurat. Analisis sentimen juga memerlukan pemahaman terhadap pola bahasa dan konteks. Pemahaman ini membantu dalam mengenali perbedaan antara opini yang sebenarnya dan penggunaan kata-kata dengan makna ganda atau ironi. Terakhir, aplikasi dan pemanfaatan hasil analisis menjadi esensial. *Sentiment analysis* telah diterapkan dalam berbagai bidang, termasuk bisnis (analisis ulasan pelanggan), keuangan (analisis sentimen pasar), dan pengawasan media sosial. Teori ini

melibatkan bagaimana hasil analisis dapat digunakan untuk pengambilan keputusan atau memahami pola perilaku pengguna.

Dalam penelitian ini, pendekatan yang akan dilakukan adalah pendekatan analisis sentimen berbasis *machine learning*. Pendekatan analisis sentimen berbasis *machine learning* menggunakan algoritma pembelajaran mesin untuk mengklasifikasikan teks ke dalam kategori sentimen. Model-model ini dilatih menggunakan data latih yang berisi contoh-contoh teks dengan label sentimen, memungkinkan mereka untuk mempelajari pola-pola yang kompleks dalam teks dan membuat prediksi sentimen yang akurat untuk teks baru. Algoritma yang umum digunakan termasuk *Support Vector Machine (SVM)*, *Naive Bayes*, dan *Random Forest*[14].

Dengan demikian, *sentiment analysis* melibatkan gabungan dari konsep-konsep ini untuk menciptakan model dan teknik yang dapat menginterpretasi opini dan perasaan manusia yang terdapat dalam teks dengan lebih baik. Penerapan teori ini dapat membantu organisasi atau peneliti untuk memahami persepsi dan respon masyarakat terhadap produk, layanan, atau peristiwa tertentu.

2.3 Teori tentang Framework / Algoritma yang digunakan

Model algoritma yang digunakan pada penelitian analisis sentimen ini adalah algoritma klasifikasi seperti *Naive Bayes*, *Support Vector Machine*, *Random Forest*, dan *Logistic Regression*. Pemilihan algoritma klasifikasi didalam penelitian ini tentunya bukan merupakan kebetulan belaka. Hal ini dikarenakan algoritma klasifikasi direkomendasikan untuk analisis sentimen karena tugas ini melibatkan pengelompokan atau klasifikasi teks ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Beberapa alasan mengapa algoritma klasifikasi cocok untuk analisis sentimen adalah sebagai berikut[21]:

1. Analisis sentimen dapat dianggap sebagai tugas klasifikasi di mana teks atau dokumen dikategorikan ke dalam kelas atau label yang

mencerminkan sentimen yang terkandung. Algoritma klasifikasi dirancang khusus untuk menangani tugas semacam ini.

2. Banyak algoritma klasifikasi untuk analisis sentimen beroperasi di bawah paradigma *supervised learning*. Ini berarti model dilatih menggunakan data yang sudah diberi label sentimen sebelumnya, memungkinkan algoritma untuk belajar pola dan hubungan antara fitur-fitur teks dan sentimen yang terkait.
3. Terdapat berbagai metode klasifikasi yang dapat diterapkan untuk analisis sentimen, seperti *Naive Bayes*, *Support Vector Machines*, dan *Logistic Regression*. Keberagaman ini memberikan fleksibilitas dalam memilih algoritma yang paling sesuai dengan karakteristik data dan tujuan analisis.
4. Algoritma klasifikasi dapat dilatih pada dataset yang berisi contoh-contoh teks dengan sentimen yang sudah diketahui. Walaupun demikian/begitu, model dapat belajar untuk mengenali pola atau fitur yang berkaitan dengan berbagai jenis sentimen.
5. Beberapa algoritma klasifikasi dapat menangani konteks dan nuansa dalam teks, memungkinkan mereka untuk memberikan hasil yang lebih akurat dalam situasi di mana sentimen dapat dipengaruhi oleh konteks tertentu.
6. Algoritma klasifikasi memberikan fleksibilitas dalam penggunaan fitur-fitur teks yang relevan. Ini dapat mencakup kata-kata kunci, frasa, atau fitur lain yang dapat membantu model mengidentifikasi sentimen dengan lebih baik.
7. Beberapa algoritma klasifikasi dapat menangani jumlah data yang besar dan dapat diimplementasikan secara efisien. Hal ini memungkinkan untuk menganalisis sentimen dalam skala besar, terutama ketika menghadapi dataset besar dari *review* atau opini.

8. Algoritma klasifikasi menyediakan metrik evaluasi yang baik untuk mengukur kinerja model, seperti akurasi, presisi, recall, dan *F1-score*. Hal ini mempermudah proses evaluasi dan tuning untuk meningkatkan performa model.

Dengan menggunakan algoritma klasifikasi, analisis sentimen dapat dilakukan secara sistematis dan dapat menghasilkan model yang dapat memahami dan mengklasifikasikan sentimen dalam teks dengan tingkat akurasi yang memadai. Setelah analisis selesai dilakukan dan berbagai angka-angka hasil model algoritma didapatkan, maka akan disimpulkan penelitian ini berdasarkan hasil tersebut untuk ditentukan algoritma mana yang memiliki akurasi tertinggi dan juga teks apa yang dinilai menjadi pengaruh yang besar terhadap keputusan seorang pemain dalam menilai *Counter Strike 2*. Bentuk akhir dari penelitian ini adalah sebuah implementasi model klasifikasi berbasis *web* yang menerima inputan teks dari penggunaanya dan berdasarkan teks tersebut akan diprediksi jenis sentimen opini tersebut dan juga akan diprediksi apakah pendapat tersebut merupakan rekomendasi ataupun tidak.

2.3.1 *Naïve Bayes Algorithm*

Algoritma *Naïve Bayes* merupakan salah satu dari berbagai algoritma yang ada di *Machine Learning*. Tujuan dari penggunaan algoritma ini adalah untuk memprediksi masa depan berdasarkan pengalaman dari masa lalu. Selain itu, *Naïve Bayes* juga bisa digunakan untuk menentukan faktor apa yang memiliki pengaruh terbesar atas terjadinya sebuah kejadian. Algoritma ini memiliki sebuah ciri khas, yaitu asumsi yang sangat kuat terhadap independensi dari setiap kejadian/case[16]. Rumus 2.1 dibawah mengilustrasikan persamaan rumus penghitungan dari dari algoritma klasifikasi *Naïve Bayes*.

$$P(A|B) = P(B|A)P(A)/P(B)_{(1)}$$

Rumus 2. 1 Persamaan algoritma *Naïve Bayes*

$P(A|B)$: Probabilitas A akan terjadi apabila B sudah terjadi

$P(B|A)$: Probabilitas B akan terjadi apabila A sudah terjadi

$P(A)$: Probabilitas kejadian A terjadi

$P(B)$: Probabilitas kejadian B terjadi

Layaknya setiap algoritma, tentunya pasti *Naïve Bayes* memiliki keunggulan dan kerugiannya di bidang tertentu. Berikut adalah kelebihan dari algoritma *Naïve Bayes*:

- Bisa mencakup berbagai jenis data, baik numerical ataupun kategorikal
- Tidak membutuhkan jumlah sampel yang banyak, sehingga cocok untuk digunakan pada penelitian dengan dataset yang kecil
- Bisa digunakan tidak hanya untuk faktor yang bersifat biner, tetapi juga faktor yang bersifat multiclass

Berikut adalah kekurangan atau kelemahan dari penggunaan algoritma *Naïve Bayes*[17]:

- Asumsi independensi *variable* dapat membuat akurasi menjadi turun, karena dengan menganggap bahwa semua variabel independen, maka secara tidak langsung juga akan mengabaikan korelasi yang mungkin terjadi antara variabel
- Keakuratan algoritma tidak bisa diukur hanya dengan satu probabilitas saja, sehingga harus menggunakan banyak faktor penentu lainnya untuk menghasilkan hasil yang optimal

2.3.2 *Support Vector Machine*

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* berfokus pada konsep pemisahan linier, mencari garis atau bidang pemisah optimal antara dua kelas dalam ruang fitur. Jika pemisahan linier tidak

mungkin, *SVM* menggunakan *kernel trick* untuk mentransformasi data ke dimensi yang lebih tinggi, di mana pemisahan linier mungkin tercapai. Rumus 2.2 mengilustrasikan rumus penghitungan dari berbagai jenis Support Vector Machine[18].

$$dH(\Phi(X_0)) = \frac{|W^T(\Phi(X_0)) + b|}{\|W\|_2} \quad (2)$$

Rumus 2. 2 Rumus Algoritma *Support Vector Machine*

Support Vector, yaitu titik-titik data yang paling dekat dengan garis atau bidang pemisah, dan hyperplane, representasi pemisahan antara kelas, merupakan elemen kunci dalam *SVM*. Tujuan utama *SVM* adalah memaksimalkan margin, yaitu jarak antara vektor dukungan dan *hyperplane*.

SVM menciptakan fungsi keputusan untuk memetakan input ke output kelas berdasarkan lokasi relatif data terhadap hyperplane. Fungsi keputusan ini bergantung pada konsep regularisasi dengan parameter *C*, yang mengontrol *trade-off* antara memaksimalkan margin dan mengizinkan pelanggaran margin.

Konsep *kernel functions* memungkinkan *SVM* menangani data yang tidak dapat dipisahkan secara linier. Beberapa kernel umum melibatkan fungsi linear, polinomial, dan *radial basis function (RBF)*, dengan pemilihan *kernel* tergantung pada karakteristik data.

Proses optimasi *SVM* melibatkan pencarian solusi untuk dual problem, yang mencakup penentuan bobot untuk setiap titik data. Kernel trick memungkinkan *SVM* untuk bekerja dengan data yang memiliki struktur non- linear dengan memetakan data ke dalam dimensi yang lebih tinggi.

SVM juga dapat diadaptasi untuk tugas klasifikasi multikelas menggunakan metode seperti *one-vs-one* atau *one-vs-all*. Keseluruhan, dasar- dasar teoritis ini menjadikan *SVM* algoritma yang kuat dan efektif

dalam berbagai tugas pembelajaran mesin, terutama dalam konteks klasifikasi dan regresi.

Seperti juga halnya dengan algoritma lainnya, *Support Vector Machine* juga pastinya memiliki keunggulan dan keterbatasannya sendiri. Berikut Merupakan kelebihan dari algoritma *Support Vector Machine*[18]:

1. Efektif pada dimensi tinggi dan cocok untuk tugas analisis data kompleks.
2. Kemampuan menangani data non-linier dengan menggunakan kernel functions.
3. Fokus pada konsep margin untuk meningkatkan generalisasi model.
4. Efektif pada dataset terbatas, bahkan dengan jumlah fitur yang besar.
5. -Fleksibilitas dalam parameter tuning untuk optimalisasi model.
6. -Mampu menangani klasifikasi multikelas dengan berbagai metode.

Berikut merupakan kekurangan dari algoritma ini adalah sebagai berikut:

1. Kinerja menurun pada dataset besar dan membutuhkan sumber daya komputasi yang signifikan.
2. Sensitif terhadap pemilihan kernel yang dapat mempengaruhi kinerja.
3. Kurang interpretatif, sulit diinterpretasikan secara intuitif.
4. Memerlukan pemahaman matematis yang mendalam.
5. Rentan terhadap noise dalam data yang dapat mempengaruhi hasil.
6. Proses pelatihan SVM dapat memakan waktu lama pada dataset besar.

2.3.3 Random Forest

Random Forest adalah salah satu algoritma dalam *Machine Learning*. Konsep dasar dari ensemble learning adalah menggabungkan prediksi dari beberapa model untuk meningkatkan kinerja dan ketepatan

hasil. *Random Forest* menggunakan pohon keputusan (*Decision Trees*) sebagai model dasar[19]. Pohon keputusan adalah struktur hierarkis yang memetakan fitur-fitur dari data input ke dalam keputusan atau prediksi. Dalam konteks *Random Forest*, bukan hanya satu pohon keputusan yang digunakan, tetapi sejumlah pohon keputusan yang dibangun secara bersamaan. Ini menciptakan suatu "hutan" dari pohon keputusan.

Salah satu fitur utama dari *Random Forest* adalah penggunaan teknik *bootstrapped sampling*. Ini berarti setiap pohon keputusan dilatih pada dataset yang dihasilkan dari sampling acak dengan pengembalian dari dataset asli. Dengan kata lain, setiap pohon melihat sebagian data yang berbeda-beda. Selain itu, *Random Forest* juga mengadopsi konsep random feature selection. Ketika membangun setiap pohon, algoritma secara acak memilih subset dari semua fitur yang tersedia. Hal ini bertujuan untuk menghindari ketergantungan yang tinggi pada satu fitur tertentu dan memastikan variasi antar pohon.

Proses penggabungan hasil dari setiap pohon juga memiliki ciri khasnya. Dalam klasifikasi, *Random Forest* menggunakan voting mayoritas untuk menentukan label kelas, sementara dalam regresi, hasil dari setiap pohon diambil rata-rata untuk mendapatkan prediksi akhir. Terdapat pula konsep *Out-of-Bag (OOB) error*, yang memungkinkan evaluasi performa model tanpa memerlukan dataset validasi terpisah. Beberapa pohon keputusan tidak menggunakan setiap sampel dalam proses pelatihan (karena sampling bootstrap), dan OOB error dihitung dari prediksi model pada sampel-sampel yang tidak digunakan.

Dalam penggunaannya, *Random Forest* memiliki beberapa hyperparameter yang dapat disesuaikan, seperti jumlah pohon, kedalaman pohon, dan jumlah fitur yang dipilih secara acak. Berdasarkan hal tersebut, maka dapat membantu mengoptimalkan kinerja model. Kelebihan dari *Random Forest* meliputi kestabilan, kinerja yang baik pada data berukuran besar, dan kemampuan untuk menangani banyak jenis data. Namun,

kekurangan utamanya termasuk kompleksitas model dan sulitnya interpretasi karena melibatkan banyak pohon[19]. Rumus 2.3 mengilustrasikan persamaan matematika *Random Forest*

$$\hat{y} = \text{modus}(y^1, y^2, \dots, y^N) \quad (3)$$

Rumus 2.3 Persamaan *Random Forest*

N = jumlah pohon keputusan

\hat{y} = rata-rata prediksi semua pohon

Secara aplikatif, *Random Forest* telah berhasil digunakan dalam berbagai bidang, termasuk klasifikasi dan regresi. Algoritma ini juga telah diterapkan dalam tugas-tugas seperti deteksi anomali dan seleksi fitur. Dengan menggabungkan kekuatan pohon keputusan individu, *Random Forest* memberikan model yang kuat dan sering memberikan hasil yang baik pada berbagai tugas *Machine Learning*.

2.3.4 Logistic Regression

Logistic Regression adalah metode dalam statistik dan *Machine Learning* yang digunakan untuk tugas klasifikasi biner. Konsep utamanya melibatkan fungsi logistik atau sigmoid function, yang mengubah nilai kontinu menjadi rentang antara 0 dan 1, membuatnya cocok untuk masalah klasifikasi biner[19].

Model *Logistic Regression* beroperasi dengan rasio probabilitas, yang menyatakan perbandingan antara probabilitas kejadian sukses dan kejadian gagal dimana pelatihan model melibatkan optimasi parameter untuk meminimalkan deviasi antara probabilitas prediksi dan hasil observasional menggunakan metode seperti maksimum *likelihood*. Fungsi cost atau loss function mencerminkan sejauh mana model meleset dari hasil yang sebenarnya, dan dalam beberapa kasus, dapat melibatkan teknik regularisasi. Rumus 2.4 menggambarkan persamaan matematika *Logistic Regression*.

$$P(y=1 | \mathbf{x}) = \frac{1}{1 + e^{-w \cdot \mathbf{x} - b}} \quad (4)$$

Rumus 2.4 Persamaan *Logistic Regression*

$P(y=1 | \mathbf{x})$: probabilitas bahwa sampel x termasuk dalam kelas positif

x : Vektor Input

w : vektor parameter (bobot) yang harus dipelajari selama pelatihan model.

b : bias

e : konstanta Euler

Model *Logistic Regression* mengandalkan asumsi asimptotik besar, yang menyiratkan bahwa semakin banyak data yang digunakan untuk pelatihan, semakin mendekati distribusi normal. Evaluasi kinerja model melibatkan pengukuran seperti akurasi, presisi, recall, dan area di bawah kurva ROC (AUC-ROC) tergantung pada tujuan klasifikasi dan kebutuhan bisnis. *Logistic Regression* sering diterapkan dalam berbagai bidang, seperti ilmu sosial, kesehatan, dan ekonomi. Meskipun sederhana, model ini dapat memberikan hasil yang baik dalam konteks klasifikasi biner ketika asumsi- asumsi yang mendasarinya terpenuhi.

2.3.5 CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang banyak digunakan dalam proyek data mining dan analisis data, yang menyediakan kerangka kerja terstruktur dan iteratif[42]. Proses ini dimulai dengan pemahaman bisnis, yang melibatkan identifikasi tujuan bisnis dan kebutuhan proyek serta penentuan kriteria kesuksesan melalui diskusi dengan pemangku kepentingan. Tahap ini penting karena memastikan bahwa proyek data mining sesuai dengan tujuan strategis organisasi dan memberikan nilai bisnis yang jelas. Selanjutnya, tahap pemahaman data melibatkan pengumpulan data awal dan eksplorasi data

untuk mendapatkan wawasan serta menemukan pola awal dan anomali. Proses ini membantu mengidentifikasi potensi masalah dengan kualitas data dan menentukan strategi untuk mengatasinya.

Pada tahap persiapan data, data mentah dibersihkan dan disiapkan untuk analisis lebih lanjut. Ini termasuk menghapus duplikat data, menangani nilai yang hilang, dan mengubah variabel kategoris menjadi numerik. Tahap ini sangat penting karena kualitas data yang buruk dapat mempengaruhi hasil akhir dari analisis data. Modeling, sebagai tahap berikutnya, melibatkan pembangunan model analisis data menggunakan teknik seperti regresi, klasifikasi, atau *clustering*[43]. Model ini dirancang untuk memprediksi atau mengelompokkan data dengan baik dan harus dikalibrasi secara tepat agar dapat memberikan hasil yang akurat dan andal.

Evaluasi dilakukan untuk memastikan bahwa model memenuhi kebutuhan bisnis dan kriteria kesuksesan yang telah ditentukan pada tahap awal. Proses ini melibatkan pengujian model menggunakan data validasi dan membandingkan kinerjanya dengan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Evaluasi yang menyeluruh memastikan bahwa model tidak hanya akurat tetapi juga relevan dan dapat digunakan dalam konteks bisnis yang nyata[42].

Tahap terakhir, *Deployment*, melibatkan implementasi model dalam lingkungan operasional dan monitoring kinerjanya secara berkelanjutan. Ini memastikan bahwa model tetap berfungsi dengan baik dan dapat beradaptasi dengan perubahan dalam data atau lingkungan bisnis. Implementasi yang efektif dari model juga mencakup pelatihan pengguna akhir dan pengembangan prosedur operasional standar untuk penggunaan model sehari-hari[42].

Framework CRISP-DM memastikan bahwa setiap fase proyek data mining dilakukan dengan seksama dan hasil yang diperoleh relevan dengan tujuan bisnis[43]. Hal ini sangat penting dalam konteks penelitian ilmiah, di mana reproducibility dan validitas adalah kunci. Dengan

mengikuti tahapan CRISP-DM, peneliti dapat memastikan bahwa setiap langkah dari pemahaman bisnis hingga implementasi model dilakukan secara sistematis dan dapat diulang, yang meningkatkan kredibilitas hasil penelitian. Framework ini tidak hanya membantu dalam mengelola proyek data mining dengan lebih efektif tetapi juga memastikan bahwa hasil akhirnya memberikan nilai nyata bagi organisasi.

2.3.6 Evaluasi Kinerja Model Klasifikasi

Evaluasi kinerja model klasifikasi merupakan tahap penting dalam proses *machine learning* yang bertujuan untuk mengukur seberapa baik model tersebut mampu melakukan prediksi terhadap kelas atau label dari data baru yang tidak pernah dilihat sebelumnya. Proses ini melibatkan sejumlah metrik evaluasi yang memberikan gambaran lebih mendalam tentang kemampuan model dalam melakukan tugas klasifikasi. Beberapa metrik utama yang digunakan dalam evaluasi ini adalah *recall score*, *precision*, *accuracy*, dan *F1 score*[44].

Recall score, juga dikenal sebagai *sensitivity* atau *true positive rate*, mengukur proporsi instance positif yang berhasil diidentifikasi oleh model. *Precision*, di sisi lain, mengukur proporsi instance yang secara tepat diidentifikasi sebagai positif dari semua instance yang diidentifikasi sebagai positif oleh model, dan sering kali disebut sebagai *positive predictive value*. *Accuracy*, sebagai metrik yang paling umum digunakan, adalah proporsi dari semua instance yang diidentifikasi dengan benar oleh model, baik positif maupun negatif, dari total instance yang ada. *F1 score*, sebagai *harmonic mean* dari *precision* dan *recall*, memberikan keseimbangan antara kedua metrik tersebut. Hal ini sangat bermanfaat ketika kelas target tidak seimbang, sehingga membantu dalam menghindari kesalahan interpretasi kinerja model[44].

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positif\ (FP)} \quad (5)$$

$$\text{Recall} = \text{True Positives (TP)} / (\text{True Positives (TP)} + \text{False Negatives (FN)}) \quad (6)$$

$$\text{F1 Score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}) \quad (7)$$

$$\text{Accuracy} = (\text{True Positives (TP)} + \text{True Negatives (TN)}) / \text{Total Sample} \quad (8)$$

Rumus 2.5 Persamaan Evaluasi Tingkat Keberhasilan Model

1. *True Positives (TP)*: Jumlah sampel yang benar-benar positif dan diprediksi positif oleh model.
2. *True Negatives (TN)*: Jumlah sampel yang benar-benar negatif dan diprediksi negatif oleh model.
3. *False Positives (FP)*: Jumlah sampel yang sebenarnya negatif tetapi diprediksi positif oleh model.
4. *False Negatives (FN)*: Jumlah sampel yang sebenarnya positif tetapi diprediksi negatif oleh model.

Evaluasi kinerja model klasifikasi bukan hanya sekedar mengukur seberapa baik model dapat memprediksi kelas dari data baru, tetapi juga penting dalam konteks pengambilan keputusan, seperti dalam bidang kedokteran, keuangan, dan teknologi, di mana hasil yang akurat dan andal sangat diharapkan untuk mendukung keputusan yang tepat. Dengan memahami metrik evaluasi ini secara mendalam, pengguna model dapat membuat keputusan yang lebih cerdas dan tepat dalam berbagai skenario aplikasi[45].

2.3.7 *TextBlob*

TextBlob adalah *library Python* untuk memproses data teks. *Library* ini menyediakan API sederhana untuk menangani tugas-tugas *Natural Language Processing (NLP)* umum seperti penandaan bagian ucapan, ekstraksi frasa kata benda, analisis sentimen, klasifikasi, terjemahan, dan banyak lagi. Teknik menetapkan salah satu dari bagian-bagian ucapan ke kata tertentu dikenal dengan POS atau penandaan *part-*

of-speech, dapat dibuat daftar *tag part-of-speech* melalui properti *tag*[46]. Ini juga dikenal sebagai penandaan "*point-of-sale*" dalam bentuk panjang. Secara sederhana, penandaan POS adalah tugas memberi label pada setiap kata dalam frase dengan bagian ucapan yang sesuai. Kata benda, kata kerja, kata keterangan, kata sifat, kata ganti, kata penghubung, dan subkategori adalah bagian dari ucapan yang dikenal. Penandaan POS berbasis aturan dan penandaan berbasis transformasi adalah jenis penandaan POS yang paling umum[46].

Penandaan POS berbasis aturan adalah salah satu metode penandaan yang paling kuno. Untuk menemukan *tag-tag* yang mungkin untuk setiap kata, penandaan berbasis aturan melihat kamus. Penandaan berbasis aturan menggunakan aturan yang ditulis tangan untuk memilih penandaan yang tepat jika kata tersebut memiliki lebih dari satu kemungkinan penandaan. Karakteristik linguistik sebuah kata, serta kata-kata sebelum dan sesudahnya, dapat dianalisis dalam penandaan berbasis aturan untuk membedakannya[46][47].

Penandaan berbasis transformasi juga dikenal sebagai penandaan Brill. Ini adalah contoh *Transformation Based Learning* (TBL), yang merupakan sistem berbasis aturan untuk memberi label POS secara otomatis pada teks yang diberikan. TBL mengubah satu status ke status lainnya menggunakan aturan transformasi, sehingga memungkinkan untuk memiliki pengetahuan linguistik secara mudah dimengerti[47].

2.3.8 ANOVA (*Analysis of Variance*)

Analysis of Variance (ANOVA) adalah teknik statistik yang digunakan untuk membandingkan rata-rata dari tiga atau lebih kelompok untuk menentukan apakah ada perbedaan yang signifikan di antara mereka. ANOVA telah menjadi alat yang sangat penting dalam penelitian eksperimental. Pada dasarnya, ANOVA mengevaluasi dua jenis variabilitas dalam data: variabilitas antara kelompok dan variabilitas dalam kelompok. Dengan menganalisis variabilitas ini, ANOVA dapat

mengidentifikasi apakah perbedaan yang diamati antara kelompok-kelompok lebih besar daripada yang diharapkan dari variasi acak[50].

ANOVA membagi total variasi dalam data (Total Sum of Squares atau SS_{total}) menjadi dua komponen utama: variabilitas antara kelompok (Between-Group Sum of Squares atau $SS_{between}$) dan variabilitas dalam kelompok (Within-Group Sum of Squares atau SS_{within}). $SS_{between}$ mengukur seberapa jauh rata-rata kelompok berbeda dari rata-rata keseluruhan, sedangkan SS_{within} mengukur variasi pengamatan dalam kelompok yang sama[50].

Untuk menghitung ANOVA, pertama-tama dihitung rata-rata dari setiap kelompok dan rata-rata keseluruhan. Kemudian, dihitung *sum of squares (SS)* untuk total variasi, antara kelompok, dan dalam kelompok. *Mean Square (MS)* diperoleh dengan membagi SS dengan derajat kebebasan yang sesuai. *F-statistic* kemudian dihitung dengan membagi $MS_{between}$ dengan MS_{within} . Nilai *F-statistic* ini dibandingkan dengan nilai kritis dari distribusi F untuk menentukan apakah perbedaan antar kelompok signifikan secara statistik[50].

$$F\text{-Statistic} = MS_{between} / MS_{within} \quad (9)$$

Rumus 2.5 Persamaan Skor F ANOVA

1. $MS_{between}$: Hasil Kuadrat Rata-rata Jarak tiap model.
2. MS_{within} : Hasil Kuadrat Rata-rata Jarak antara model.

ANOVA didasarkan pada beberapa asumsi, yaitu observasi dalam setiap kelompok harus independen, data dalam setiap kelompok harus mengikuti distribusi normal, dan varians antar kelompok harus homogen. Ketika asumsi-asumsi ini terpenuhi, ANOVA memberikan hasil yang dapat diandalkan. Jika ANOVA menunjukkan bahwa ada perbedaan yang signifikan antara kelompok, tes post-hoc seperti Tukey's HSD dapat digunakan untuk mengidentifikasi pasangan kelompok mana yang berbeda signifikan[51].

ANOVA memiliki beberapa variasi, termasuk *One-Way* ANOVA, yang membandingkan satu faktor independen di beberapa kelompok, dan *Two-Way* ANOVA, yang mengevaluasi dua faktor independen. *Repeated Measures* ANOVA digunakan ketika pengamatan dilakukan pada subjek yang sama di bawah kondisi yang berbeda. Meskipun ANOVA adalah alat yang kuat, hasilnya dapat terpengaruh jika asumsi dasar tidak terpenuhi, sehingga penting untuk memverifikasi asumsi-asumsi ini sebelum melakukan analisis[51].

Dalam kesimpulannya, ANOVA adalah metode statistik yang efektif untuk membandingkan rata-rata antara beberapa kelompok dan menentukan apakah perbedaan yang diamati signifikan secara statistik. Dengan membagi total variasi menjadi variabilitas antara kelompok dan dalam kelompok, ANOVA menyediakan kerangka kerja yang kuat untuk analisis data eksperimental dan pengambilan keputusan berbasis data[51].

2.4 Teori tentang Tools / Software yang digunakan

2.4.1 Jupyter Notebook

Jupyter Notebook adalah aplikasi open sources yang memungkinkan pembuatan dan berbagi dokumen interaktif yang menggabungkan kode, teks naratif, visualisasi, dan formula matematika. Nama "Jupyter" berasal dari gabungan tiga bahasa pemrograman utama yang didukung oleh platform ini: *Julia*, *Python*, dan *R*[20].

Jupyter Notebook dibagi menjadi sel-sel yang masing-masing dapat berisi teks atau kode. Sel kode digunakan untuk menulis dan mengeksekusi kode *Python*, *Julia*, *R*, dan beberapa bahasa pemrograman lainnya. Sementara sel Markdown digunakan untuk menyajikan teks naratif dengan format markdown, memungkinkan penulisan yang lebih kaya dan terstruktur.

Setiap *notebook* terkait dengan kernel tertentu, yang merupakan entitas yang mengeksekusi kode dalam sel-sel *notebook*. Kernel dapat

dihentikan dan dijalankan ulang tanpa kehilangan data. Hasil eksekusi kode, seperti output cetak atau grafik, ditampilkan langsung di bawah sel kode yang sesuai.

Jupyter Notebook juga menyertakan *toolbar* dan menu yang memberikan akses ke berbagai fungsi, seperti menyimpan notebook, menambahkan sel, dan mengganti kernel. Proyek ini dimulai dengan IPython pada tahun 2001 oleh Fernando Pérez, dan pada tahun 2014, berkembang menjadi proyek Jupyter yang lebih besar, mendukung bahasa pemrograman lainnya.

Seiring berjalannya waktu, *Jupyter* terus berkembang. Pada tahun 2018, *JupyterLab* dirilis, memberikan antarmuka pengguna baru dengan tata letak yang lebih fleksibel dan fitur-fitur baru. Saat ini, *Jupyter Notebook* digunakan secara luas dalam pendidikan, penelitian, dan industri untuk analisis data interaktif, visualisasi, dan pengembangan perangkat lunak. Perkembangan dan popularitasnya mencerminkan kebutuhan yang terus berkembang untuk alat yang mendukung interaktivitas dan pemahaman yang mendalam terhadap data dalam komputasi ilmiah dan analisis[20].

2.4.2 Streamlit

Streamlit adalah sebuah *framework open-source* yang digunakan untuk membuat aplikasi *web* dengan mudah menggunakan *Python*. *Streamlit* dirancang khusus untuk pengembangan aplikasi data dan *Machine Learning* yang interaktif. Dengan *Streamlit*, pengguna dapat membuat antarmuka *web* dengan cepat hanya dengan menulis kode *Python* sederhana tanpa perlu memiliki pengetahuan mendalam tentang HTML, CSS, atau JavaScript. *Framework* ini menyediakan berbagai komponen seperti tombol, input, dan plot yang dapat diintegrasikan dengan mudah, sehingga memungkinkan pengembang untuk fokus pada pemrosesan data dan logika aplikasi[38].

Streamlit juga menawarkan fitur-fitur yang mempermudah proses pengembangan aplikasi, seperti "*hot reloading*" yang memungkinkan pengguna untuk melihat perubahan secara langsung saat mengedit kode, serta kemampuan untuk menyajikan data dari berbagai sumber seperti *pandas DataFrame* atau *TensorFlow*. Dengan kelebihan ini, *Streamlit* menjadi pilihan populer bagi para *data scientist* dan pengembang *Machine Learning* yang ingin membuat prototipe atau aplikasi interaktif dengan cepat dan mudah, tanpa harus memikirkan kompleksitas pengembangan web[38].



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA