

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
<i>Sentiment Analysis on Customer Satisfaction of Digital Payment in Indonesia: A Comparative Study using KNN and Naïve Bayes</i> [13]	<i>Journal of Physics</i>	Hilman Wisnu, Muhammad Afif, Yova Ruldevyani / 2020	Hasil penelitian menunjukkan bahwa algoritma KNN menghasilkan akurasi yang lebih baik daripada <i>Naïve Bayes</i> . Pada berbagai servis <i>digital payment</i> , algoritma KNN menghasilkan akurasi tertinggi dengan 83.50% pada Go-Pay, 84.00% pada OVO, dan 91.00% pada Link Aja.	Melalui penelitian yang dilakukan menghasilkan algoritma KNN yang akurasinya lebih tinggi dari <i>Naïve Bayes</i> . Selain itu servis Link Aja memiliki sentimen netral terbanyak dengan 0 sentimen negatif, sedangkan OVO memiliki jumlah sentimen negatif terbanyak. Pada Go-Pay pelanggannya lebih memiliki sentimen netral ketimbang negatif.
Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma <i>Naïve Bayes</i> , <i>Neural Network</i> , KNN dan SVM [14]	Inti Nusa Mandiri	Melisa Winda Pertiwi / 2021	Hasil penelitian menunjukkan bahwa dalam melakukan klasifikasi sentimen analisis teks berbahasa Inggris algoritma KNN memiliki akurasi tertinggi dengan 90.76%, algoritma SVM dengan 89.03%, algoritma <i>Naïve Bayes</i> dengan 78.16%, dan	Berdasarkan penelitian sarana dan transportasi mudik pada setiap tahunnya dapat mengalami perubahan, disebabkan opini setiap orang dapat berbeda. Sentimen publik mengenai mudik tahun 2019 menunjukkan lebih banyak opini positif.

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
			algoritma <i>Neural Network</i> dengan 52.73%.	
Metode <i>Support Vector Machine</i> pada <i>Review Game</i> Garena Free Fire Berbasis <i>Website</i> [18]	Jurnal Ilmu Komputer dan Informatika	Siti Saidah, Nurul Hidayati Nusyirwan, Ratu Balkis Khoirunnisa / 2023	Hasil penelitian menunjukkan bahwa model yang dapat digunakan dalam melakukan <i>review game</i> Free Fire dengan menggunakan metode <i>Support Vector Machine</i> dengan akurasi 95%.	Melalui penelitian serta penerapan metode <i>Support Vector Machine</i> pada <i>review game</i> Garena Free Fire berbasis <i>web</i> mampu menjawab opini pengguna terhadap penilaian ulasan dengan kesimpulan positif ataupun negatif. Hasil analisis penerapan berupa matriks, yang kemudian <i>web</i> dengan dukungan <i>Vader Lexicon</i> membantu menentukan pelabelan positif atau negatif pada <i>review</i> .
<i>Sentiment Analysis about Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, and Decision Tree</i> [15]	IJNMT ( <i>International Journal of New Media Technology</i> )	Nico Nathanael Wilim, Raymond Sunardi Oetama / 2021	Hasil penelitian menunjukkan dari penggunaan 3 algoritma, algoritma KNN menghasilkan akurasi tertinggi sebesar 76.94%. Namun tidak terdapat algoritma yang menunjukkan performa terbaik pada seluruh data. Pada tahun 2018 algoritma <i>Naïve Bayes</i> menghasilkan performa terbaik, namun pada tahun 2019	Melalui penelitian menunjukkan bahwa opini publik pada Twitter dapat digunakan untuk melakukan konfirmasi pemenang dari Panasonic Gobel Award. Pada tahun 2018 ILC yang menjadi pemenang, sedangkan Mata Najwa yang menjadi pemenang pada tahun 2019. Hal tersebut

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
			algoritma KNN yang menjadi algoritma dengan performa terbaik.	dikarenakan pada tahun 2018 jumlah sentimen positif lebih banyak ketimbang Mata Najwa, sedangkan pada tahun 2019 terdapat penurunan sentimen positif pada ILC. Serta penggunaan algoritma dengan akurasi terbaik adalah KNN.
<i>Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short-Term Memory</i> [19]	Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)	Dinar Ajeng Kristiyanti, Sri Hardani / 2023	Hasil penelitian menunjukkan penggunaan berbagai algoritma pada analisis sentimen penerimaan publik atas vaksin Covid-19 di Indonesia menghasilkan algoritma SVM menghasilkan akurasi terbaik sebesar 84.89%, dengan presisi 85.00%, 100.00% recall, dan 92.00% F-1. Algoritma Naïve Bayes menghasilkan akurasi 84.65% dan algoritma LSTM menghasilkan akurasi 82.97%.	Berdasarkan penelitian algoritma SVM menjadi algoritma dengan performa terbaik, berdasarkan data penggunaan vaksin Sinovac menjadi vaksin yang menerima respon positif dan negatif pada survei yang dilakukan melalui Twitter. Respon positif didapat berdasarkan jumlah penggunaan banyak dan efek samping kecil. Respon negatif didapat dikarenakan efektifitas yang rendah.
Penerapan Metode K-Nearest Neighbor dan Information Gain pada Klasifikasi Kinerja Siswa [16]	Jurnal Ilmu Pengetahuan dan Teknologi Komputer	Tyas Setiyorini, Rizky Tri Asmono / 2019	Pada penelitian dilakukan perbandingan akurasi metode K-Nearest Neighbor dengan metode K-Nearest Neighbor yang	Penelitian menunjukkan penggunaan seleksi fitur Information Gain dapat mengurangi dimensi vektor, dan

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
			digabungkan dengan <i>Information Gain</i> dalam klasifikasi kinerja siswa. Pada penelitian menunjukkan klasifikasi dengan hanya menggunakan metode <i>K-Nearest Neighbor</i> menghasilkan rata-rata akurasi sebesar 74.068% sedangkan metode <i>K-Nearest Neighbor</i> dengan <i>Information Gain</i> menghasilkan rata-rata sebesar 76.553%.	meningkatkan tingkat akurasi klasifikasi dengan metode KNN. Terdapat perbedaan sebesar 2% yang dihasilkan dengan menggunakan seleksi fitur <i>Information Gain</i> .
<i>A Comparative Study of Feature Selection Techniques in Machine Learning for Predicting Stock Market Trends</i> [17]	<i>Journal of Applied Data Sciences</i>	Adi Suryaputra Paramita, Shalomeira Valencia Winata / 2023	Pada penelitian menunjukkan teknik <i>Recursive Feature Elimination</i> (RFE) menjadi teknik paling unggul dalam melakukan prediksi <i>stock market</i> . Teknik RFE membantu dalam memilih fitur yang paling berpengaruh pada prediksi nilai <i>stock market</i> sehingga memberikan pengetahuan lebih dalam pada tren <i>stock market</i> . Penggunaan teknik RFE dapat mengurangi <i>overfitting</i> , dan meningkatkan performa presisi prediksi.	Pada penelitian, meneliti tentang penggunaan <i>Principal Component Analysis</i> (PCA), <i>Information Gain</i> (GA), dan <i>Recursive Feature Elimination</i> (RFE) dalam memprediksi kondisi <i>stock market</i> . Hasil menunjukkan penggunaan teknik RFE menghasilkan akurasi presisi yang baik dalam melakukan prediksi pasar saham. Penggunaan teknik RFE membantu dalam melakukan pengambilan keputusan dengan

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
				menyeleksi fitur relevan, mengurangi resiko investasi, dan meningkatkan performa strategi.
Analisis Sentimen Menggunakan <i>K-Nearest Neighbor</i> Terhadap <i>New Normal</i> Masa <i>Covid-19</i> Di Indonesia [20]	Techno.Com, Vol 21, No 1	Mhd. Furqan, Sriani, Susan Mayang Sari / 2022	Penelitian menunjukkan penggunaan metode KNN dengan menggunakan pelabelan menggunakan <i>lexicon</i> menghasilkan kelas sentimen positif sebanyak 811 data, sedangkan kelas sentimen negatif 189 data. Penelitian menghasilkan pengujian paling optimal dengan menggunakan data latih 80%, dan data uji 20% dengan hasil akurasi sebesar 94.5%.	Pada penelitian meneliti mengenai analisis sentimen terhadap <i>new normal</i> pada era <i>Covid-19</i> menggunakan KNN pada Twitter. Hasil pelabelan menghasilkan jumlah sentimen positif yang lebih banyak dibandingkan sentimen negatif. Pada penelitian menyimpulkan klasifikasi KNN menggunakan pengujian 80% <i>percentage split</i> menjadi pengujian paling optimal dengan hasil akurasi yang tinggi.
<i>K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation</i> [21]	<i>Journal of Engineering and Applied Technology</i>	Nur Hidayati, Arief Hermawan / 2021	Pada penelitian menghasilkan akurasi tertinggi menggunakan algoritma KNN dengan hasil 85.28%. Hasil akurasi ini didapatkan dengan menggunakan $K=7$ menggunakan perhitungan jarak menggunakan <i>Euclidean</i> dan <i>Manhattan</i> . Hasil yang sama	Penelitian menghasilkan simpulan perhitungan <i>Euclidean</i> ataupun <i>Manhattan</i> tidak berpengaruh pada akurasi prediksi, perbedaan pada perhitungan jarak tidak mempengaruhi klasifikasi data. Hasil terbaik didapatkan pada $K=7$ dengan

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
			tersebut juga didapatkan menggunakan perhitungan <i>Manhattan Distance</i> dengan $K=9$ . Melalui penelitian juga didapatkan peningkatan nilai $K$ tidak selalu menghasilkan akurasi yang lebih baik.	hasil akurasi sebesar 85.28%. Selain itu ditemukan penambahan nilai $K$ tidak selalu memberikan performa yang lebih baik.
Analisis Sentimen Media Sosial Twitter Dengan Algoritma <i>K-Nearest Neighbor</i> Dan Seleksi Fitur <i>Chi-Square</i> (Kasus <i>Omnibus Law</i> Cipta Kerja) [22]	Jurnal Sains Komputer & Informatika (J-SAKTI)	Aditiya Yoga Pratama, Yuyun Umaidah, Apriade Voutama / 2021	Penelitian dilakukan dengan melakukan pengujian pada berbagai model dengan persentase penggunaan fitur seleksi yang berbeda, menghasilkan akurasi yang berbeda. Hasil dengan akurasi terbesar yaitu 81.4% dengan nilai $K=5$ . Evaluasi dilakukan dengan menggunakan pengujian model dengan <i>K Fold Cross Validation</i> dan fitur seleksi <i>Chi Square</i> dengan mengambil data secara acak sebanyak 100 data dan menghasilkan akurasi sebesar 83%. Pelabelan yang dilakukan pada 500 data menghasilkan 78.8% sentimen negatif, 21.2% termasuk pada sentimen positif.	Penelitian menunjukkan penggunaan seleksi fitur <i>Chi Square</i> dengan mengambil fitur sebanyak 25% menghasilkan performa yang lebih baik dibandingkan model lainnya. Selain itu dari seluruh data yang digunakan kata rakyat menjadi kata yang paling dominan. Penelitian juga menunjukkan 78.8% sentimen negatif, dan 21.2% sentimen positif.

Penelitian yang dilakukan memiliki beberapa kebaruan dibandingkan penelitian terlebih dahulu yang terdapat pada tabel 2.1. Pada penelitian sebelumnya yang menggunakan algoritma KNN pada analisis sentimen pembayaran digital menunjukkan bahwa algoritma KNN memiliki akurasi yang terbaik dengan akurasi sebesar 91% [13]. Pada hasil penelitian tersebut masih terdapat akurasi yang rendah pada beberapa aplikasi yang digunakan sebagai objek penelitian. Oleh karena itu penelitian akan dilakukan dengan melakukan optimasi penggunaan KNN sehingga menghasilkan performa yang lebih baik dalam melakukan analisis sentimen.

Pada penelitian sebelumnya terdapat berbagai penelitian analisis sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 pada twitter yang menggunakan berbagai algoritma termasuk KNN. Pada penelitian tersebut melakukan proses *preprocessing* data dan pembobotan teks menggunakan TF-IDF. Penelitian menghasilkan algoritma KNN memiliki akurasi paling tinggi sebesar 90.76% beserta nilai AUC sebesar 0.939. Penelitian dilakukan dengan menganalisis sentimen pada 2 kelas yaitu positif, dan negatif [14].

Terdapat penelitian sebelumnya yang juga melakukan penelitian pada Twitter mengenai analisis sentimen penerimaan masyarakat atas vaksin *Covid-19* di Indonesia yang menggunakan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Long Short-Term Memory*. Penelitian melakukan *labeling* menggunakan *Vader Lexicon* untuk melakukan *labeling* otomatis pada kalimat Bahasa Inggris menjadi kelas positif, dan negatif. Penelitian tersebut menghasilkan penggunaan algoritma SVM yang terbaik diantara algoritma lainnya dengan akurasi sebesar 84.89% [19]. Pada penelitian ini akan dilakukan dengan melakukan analisis sentimen pada 3 kelas yaitu positif, negatif, dan netral dengan menggunakan *Vader Lexicon*. Penelitian akan berfokus pada penggunaan algoritma KNN, melakukan optimasi dan menguji performa algoritma KNN pada data *review* pemain *game* Free Fire dan PUBG Mobile.

Penelitian analisis sentimen pada program televisi *Indonesian Lawyers Club* yang menggunakan algoritma *K-Nearest Neighbor*, *Naïve Bayes*, dan *Decision Tree* menghasilkan algoritma terbaik yang berbeda pada tahun 2018, dan 2019. Pada tahun 2018 algoritma *Naïve Bayes* menjadi algoritma terbaik, sedangkan pada tahun 2019 algoritma KNN menjadi algoritma terbaik dengan akurasi 76.94%. Performa algoritma KNN yang lebih buruk dibandingkan algoritma *Naïve Bayes* pada tahun 2018 disebabkan karena algoritma KNN mengalami kesulitan dalam menentukan *neighbor* terdekat [15]. Pada penelitian akan melakukan optimasi yang dapat membantu algoritma KNN dalam mengatasi kesulitan dalam menentukan *neighbor* terdekat.

Mengatasi kesulitan yang dialami oleh algoritma KNN dapat diselesaikan dengan berbagai solusi, salah satunya adalah penggunaan seleksi fitur yang dilakukan pada penelitian penerapan metode *K-Nearest Neighbor* dan *Information Gain* pada klasifikasi kinerja siswa. Pada penelitian tersebut menghasilkan peningkatan akurasi sebesar 2% ketika menggunakan metode *K-Nearest Neighbor* dengan *Information Gain* [16]. Seleksi fitur dapat menggunakan teknik selain *Information Gain*, pada penelitian sebelumnya yang melakukan studi perbandingan teknik seleksi fitur dalam pembelajaran mesin dalam melakukan prediksi tren pasar saham. Pada penelitian tersebut dilakukan perbandingan teknik *Information Gain* (GA), *Principal Component Analysis* (PCA), *Recursive Feature Elimination* (RFE) dengan hasil teknik seleksi fitur paling unggul adalah teknik RFE dalam melakukan prediksi tren pasar saham [17]. Pada penelitian ini akan melakukan optimasi dengan menggunakan berbagai teknik termasuk fitur seleksi RFE yang digabungkan dengan teknik lainnya pada data yang berbentuk teks.

Pada penelitian lainnya yang melakukan analisis sentimen media sosial twitter menggunakan algoritma KNN dan seleksi fitur *Chi Square* menghasilkan akurasi sebesar 81.4% [22]. Hasil tersebut lebih besar dibandingkan akurasi yang dihasilkan model tanpa menggunakan seleksi fitur, hal tersebut membuktikan optimasi yang dilakukan dengan menambahkan seleksi fitur. Melalui seluruh penelitian terdahulu yang melakukan fitur seleksi

pada pemodelan yang dilakukan menghasilkan hasil yang positif dengan meningkatnya performa model. Melalui berbagai model yang dilakukan seluruh metode yang dilakukan memiliki keuntungan masing-masing. *Chi Square* dan RFE menghasilkan akurasi model diatas 80%, namun fitur seleksi RFE memiliki keuntungan dalam mengurangi dimensi fitur pada *dataset* yang memiliki jumlah fitur yang besar [17]. Penggunaan metode RFE sesuai dengan genre dan *game Battle Royale* yang memiliki keanekaragaman pengalaman dan ulasan yang bermacam.

Penelitian lainnya yang melakukan analisis sentimen menggunakan KNN terhadap era *New Normal Covid-19* pada Indonesia, menunjukkan hasil sentimen positif yang lebih besar dengan 811 data dibandingkan dengan sentimen negatif dengan jumlah 189 data. Hal tersebut menghasilkan kebijakan *new normal* yang dapat diterima dengan baik oleh publik. Penelitian menunjukkan penggunaan *split data* 80% data *training* dan 20% data *testing* menjadi paling optimal [20]. Penelitian menyarankan penggunaan terjemahan bahasa asing yang dapat diterapkan pada penelitian yang akan dilakukan. Algoritma KNN yang menggunakan atau melakukan perhitungan jarak, berbeda dengan algoritma lainnya. Melalui penelitian yang dilakukan dengan melakukan klasifikasi gratifikasi pelajar menggunakan KNN dengan *Euclidean* dan *Manhattan distance* menghasilkan menggunakan metode kalkulasi yang berbeda tidak meningkatkan performa model [21]. Penelitian tersebut juga menunjukkan penambahan nilai K yang tidak selalu memberikan perubahan atau performa yang lebih baik. Melalui karakteristik maupun pemodelan yang dilakukan dengan menggunakan algoritma KNN menunjukkan penggunaan KNN yang cocok dalam melakukan klasifikasi atas berbagai ulasan maupun *review* dari pengguna.

Selain penelitian akan berfokus pada penggunaan algoritma *K-Nearest Neighbor*, pada penelitian ini ingin dilakukan analisis perbedaan dari *game* Free Fire dan juga PUBG Mobile. Penelitian ingin mengungkap pengaruh dari konsep permainan, tampilan, dan hal lainnya yang sebelumnya belum dapat diteliti, penelitian sebelumnya hanya dilakukan pada *game* Free Fire [18].

Penelitian ini dilakukan dengan mengimplementasi dan melakukan publikasi melalui *prototype website*, yang dapat memberikan informasi lebih luas mengenai *game* yang akan diteliti.

## 2.2 Teori tentang Topik Skripsi

### 2.2.1 *Game Online*

*Game online* adalah sebuah jenis permainan yang memanfaatkan jaringan internet sebagai media. *Game online* menggunakan sebuah mekanisme yang dapat menghubungkan beberapa pemain secara bersamaan pada sebuah permainan. *Game online* merupakan sebuah permainan dengan memanfaatkan jaringan, sehingga terdapat interaksi antara satu orang dengan lainnya untuk mencapai suatu tujuan, melakukan misi, dan meraih nilai dalam dunia virtual [23]. *Game online* dapat dapat berkisar dari yang sederhana sampai kompleks, ada yang berbasis teks maupun grafis yang menggabungkan berbagai elemen [24].

Sebelum adanya *game online*, pada umumnya masyarakat bermain *game* yang tidak menggunakan jaringan internet atau yang disebut *game offline*. Seiring berjalannya waktu dan perkembangan teknologi yang ada terjadi transisi berbagai *game* menjadi *game online*. Terdapat peningkatan signifikan pada *game online* secara berkala menandakan perkembangan teknologi dan jaringan internet yang ada [25]. Pada *game online* terdapat berbagai tipe *game online* diantaranya yaitu *game First Person Shooter (FPS)*, *Real-Time Strategy (RTS)*, *Role-Playing Games (RPG)*, *Cross-Platform Online Play*, *Browser Games*, dan *Simulation Games* [25].

### 2.2.2 *Game Mobile*

*Game mobile* adalah sebuah *game* yang dapat diunduh dan dimainkan melalui sebuah *smartphone* atau ponsel pintar. *Game mobile* dikembangkan dan dikhususkan untuk pengguna perangkat seluler. Melalui teknologi maupun *smartphone* yang canggih membuat *game mobile* tidak kalah dengan *game* lainnya [26]. Pada *game mobile* sendiri terdapat berbagai tipe *games* yang umumnya ditemukan, diantaranya:

### 1. *Adventure Games*

Merupakan *game* yang memberikan sebuah cerita interaktif atas sebuah tokoh yang dapat dikendalikan oleh pemain. Membuat pemain dapat membentuk cerita dan melakukan eksplorasi pada *game*.

### 2. *Action Games*

Adalah *game* yang akan memberikan tantangan kepada pemain, sehingga pemain harus siap dan memiliki keterampilan yang baik dalam melawan musuh dan menyelesaikan tantangan.

### 3. *Fighting Games*

Merupakan *game* yang berisi pertarungan, sehingga pemain harus berhadapan dengan orang lain atau AI (*Artificial Intelligence*). Melalui *game* ini pemain dapat mengeluarkan kemampuan berbeda dari karakter yang digunakan oleh pemain.

### 4. *First Person Shooter* dan *Third Person Shooter*

Merupakan *game* tembak menembak dengan beberapa hal atau misi yang perlu dilakukan pemain. *First Person Shooter* memberikan pemain pandangan orang pertama, sedangkan *Third Person Shooter* memberikan pandangan orang ketiga.

### 5. *Role Playing Games*

Adalah *game* dimana pemain dapat menjelajahi dunia *game* dan melakukan pengembangan pada karakter yang digunakannya.

### 6. *Racing Games*

Adalah *game* yang berisikan balapan kendaraan, melalui *game* ini pemain berlomba untuk menjadi yang tercepat dan memenangi balapan.

### 7. *Real Time Strategy*

Merupakan game yang melibatkan unsur strategi, dimana pemain dapat mengelola sumber daya dan mengatur strategi dalam permainan.

### 2.2.3 Game Free Fire

Free Fire merupakan sebuah *game* dengan genre *battle royale* yang dikembangkan oleh 111dots Studio. *Game* ini diterbitkan pada tahun 2017 oleh Garena yang berasal dari Singapura. Semenjak penerbitan yang dilakukan *game* Free Fire semakin berkembang dan diminati oleh berbagai pemain [18].



Gambar 2. 1 Free Fire pada Google Play

*Game* Free Fire memiliki jumlah pemain yang mencapai miliaran di dunia. Pada gambar 2.1 menunjukkan jumlah unduhan *game* yang mencapai lebih dari 1 miliar pada Google Playstore. *Game* Free fire memiliki *rating* atau nilai 4.1 dengan lebih dari 117 juta ulasan.

Free Fire menggunakan basis genre *battle royale* yang memadukan unsur bertahan hidup beserta tembak menembak. Pada *game* ini menggunakan jaringan internet yang kemudian akan menyambungkan beberapa pemain di dunia, yang kemudian akan bermain pada sebuah arena. Kemudian pemain bertujuan untuk menjadi tim atau orang terakhir pada arena tersebut.

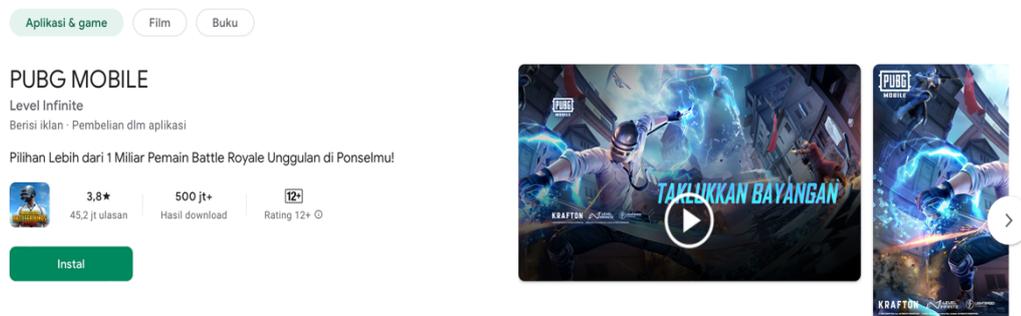
Pada permainan, pemain dapat mengumpulkan persediaan terlebih dahulu. Kemudian pemain dapat bebas mengelilingi arena serta melakukan tembakan jika melihat orang atau tim lain. Pada permainan ini terdapat sebuah area yang harus dicapai oleh pemain setiap beberapa menit, jika pemain berada

pada luar area maka pemain akan terkena serangan dari zona secara berkala. Oleh karena itu pemain harus bergerak ke area yang sudah ditentukan sambil bertahan hidup.

Pada *game* Free Fire memiliki sistem, yang dapat menampilkan peringkat dari pemain berdasarkan poin yang dimiliki. Poin dapat didapatkan pemain ketika pemain berhasil mendapatkan urutan bertahan hidup yang baik ataupun menjadi pemenang permainan. Selain itu pada *game* ini juga terdapat berbagai *event* ataupun program kegiatan yang diadakan oleh Garena. Melalui *event* ini pemain dapat menjalankan misi untuk mendapatkan hadiah yang sudah ditentukan.

#### 2.2.4 *Game* PUBG Mobile

PUBG (Player Unknown's Battle Ground) Mobile adalah sebuah *game* yang memiliki genre sama dengan *game* Free Fire yaitu *battle royale*. *Game* ini dirilis pada 19 Maret 2018 dan dibuat oleh Brendan Greene [27]. PUBG Mobile dikembangkan oleh Tencent Games.



Gambar 2. 2 PUBG Mobile pada Google Playstore

PUBG Mobile memiliki jumlah pemain di dunia yang sangat banyak mencapai jutaan pemain. Pada gambar 2.2 menunjukkan jumlah hasil *download* yang dilakukan untuk *game* PUBG Mobile yaitu lebih dari 500 juta unduhan. PUBG Mobile memiliki *rating* atau penilaian 3.8 dengan lebih dari 45.2 juta ulasan.

Konsep permainan PUBG Mobile adalah sejumlah pemain yang terhubung melalui jaringan internet kemudian dikumpulkan pada sebuah arena yang memaksa pemainnya untuk saling mengeliminasi. Tujuan permainan ini tentunya menjadi orang terakhir yang bertahan hidup pada arena permainan. Pada awal permainan pemain dapat mengumpulkan persediaan, kemudian melanjutkan bertahan hidup sekaligus mengeliminasi lawan. Konsep permainan yang digunakan sama dengan yang digunakan pada Free Fire.

Selain itu pada PUBG Mobile terdapat sistem peringkat yang berdasarkan poin yang dimiliki pemain. Selain itu Tencent Games juga sering mengadakan *event* untuk para pemain, yang biasanya akan memberikan misi bagi para pemain untuk bermain pada pertandingan. Melalui hal tersebut membuktikan konsistensi Tencent Games dalam mengembangkan *game* PUBG Mobile.

#### **2.2.5 Analisis Sentimen**

Analisis sentimen adalah sebuah konsep yang mengacu pada *opinion mining* atau penggunaan data dari komentar, opini, sentimen, maupun perasaan yang dimiliki dalam bentuk tertulis atau teks [28]. Proses analisis dilakukan dengan melakukan pengelompokan teks sentimen. Melalui analisis sentimen dapat menyatakan perasaan emosional sedih, gembira, ataupun marah [28]. Sentimen analisis dilakukan untuk mengetahui dan memahami sentimen dari populasi yang dilakukan analisis. Melalui sentimen analisis mampu membantu sebuah perusahaan untuk lebih memahami target pasar, serta meningkatkan produk yang ditawarkan oleh perusahaan.

Analisis sentimen dapat memberikan pandangan dari pengguna atau konsumen kepada perusahaan, melalui hal ini perusahaan dapat meningkatkan proses bisnis ataupun produk yang dimiliki sehingga lebih dinikmati oleh pengguna. Berdasarkan kritik ataupun opini konsumen, perusahaan dapat mengembangkan produk yang ditawarkan berdasarkan umpan balik dari pengguna atau konsumennya [29]. Selain itu hasil analisis sentimen yang dilakukan dapat membuat sebuah perusahaan lebih unggul dibandingkan

perusahaan lainnya, dikarenakan lebih memahami penggunaannya. Bagi para pengguna dapat memberikan pemahaman ataupun pengalaman dari orang lain terhadap produk ataupun hal yang dianalisis.

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

### 2.3.1 CRISP-DM

CRISP-DM atau singkatan dari *Cross Industry Standard Process - Data Mining* merupakan *open standard* model dari proses pendekatan *data mining* yang biasa digunakan oleh ahli *data mining*. Model ini menggambarkan siklus hidup *data mining* atau *data mining lifecycle*. Model ini sering digunakan dikarenakan berbagai keuntungan yang dapat digunakan untuk menyelesaikan permasalahan pada industri *data mining* [30]. Tahap pada CRISP-DM adalah:

1. *Business understanding*
2. *Data understanding*
3. *Data preparation*
4. *Modeling*
5. *Evaluation*
6. *Deployment*

*Business Understanding*: Pada tahap pertama ini merupakan tahapan penting dari *data mining*. Dikarenakan melalui tahap ini dilakukan pemahaman atas permasalahan yang ingin diselesaikan. Dilakukan pemahaman atas latar belakang, tujuan dan kebutuhan proyek dari perspektif bisnis. Pemahaman dilakukan dengan mengenali berbagai pola pada kasus sehingga dapat dilakukan *data mining*. Kemudian pengetahuan dan pemahaman tersebut diubah dan digunakan dalam mendefinisikan permasalahan proyek tersebut [31].

*Data Understanding*: Tahap kedua pada CRISP-DM yang bertujuan dalam pemahaman data yang menjadi acuan solusi dan tujuan bisnis. Pada tahap ini ditentukan data yang tepat untuk dianalisis, serta mengidentifikasi dan memahami kualitas dan limitasi data. Tahap ini dapat dimulai dengan pengumpulan data dengan berbagai cara yaitu observasi, kuesioner, maupun wawancara. Data yang sudah terkumpul tersebut kemudian akan dijadikan

penunjang pada kasus. Data tersebut kemudian dilakukan identifikasi sehingga dapat mendeskripsi data yang kemudian dapat diolah pada tahap selanjutnya. Tidak hanya mengidentifikasi data, pada tahap ini juga dilakukan eksplorasi data sehingga dapat mengetahui batasan maupun pola dari data. Melewati proses eksplorasi, kemudian data akan diverifikasi untuk membuktikan kualitasnya. Tujuan dari tahap ini adalah mempersiapkan kekurangan yang terjadi, melakukan pendekatan, dan mengidentifikasi sub-kategori [31].

*Data Preparation:* Pada tahap ini mencakup seluruh kegiatan preparasi yang dilakukan dalam menyusun *dataset* yang akan diteliti dan dianalisis. Proses persiapan data ini dapat dilakukan berulang kali tanpa urutan tertentu. Dilakukan perancangan data agar sesuai dengan tujuan dan masalah yang terkait dalam *data mining*. Data yang memiliki informasi maupun keterkaitan dengan objek penelitian ataupun kasus akan dipilih dan dibangun menjadi sebuah *dataset*. Pembangunan data akan dilakukan dengan menggunakan *query* atas data ataupun tabel yang terpilih. Dilakukan pembersihan kekurangan pada data yang sudah ditemukan pada tahap *data understanding* sebelumnya. Seperti data masih memiliki nilai yang hilang ataupun kekurangan lainnya dibetulkan pada tahap ini [31].

*Modeling:* Pada tahap ini akan dilakukan pemilihan dan penerapan model berdasarkan *dataset* yang digunakan. Tahap ini melibatkan penerapan berbagai algoritma *machine learning* serta metode statistik. Algoritma pada *machine learning* sendiri terbagi menjadi 3 yaitu *clustering*, *classification*, dan juga regresi. Algoritma tersebut memiliki karakteristik dan persyaratannya sendiri, dimana jika tidak cocok maka data dapat kembali dipersiapkan. Pemodelan yang dilakukan akan mengaplikasikan algoritma yang dipilih untuk menentukan dan mengidentifikasi pesan maupun pola yang ada pada data. Tahap ini adalah tahap primer dimana proses *data mining* dilakukan, dan diperlukan pemahaman fundamental atas *data mining* tersebut [31].

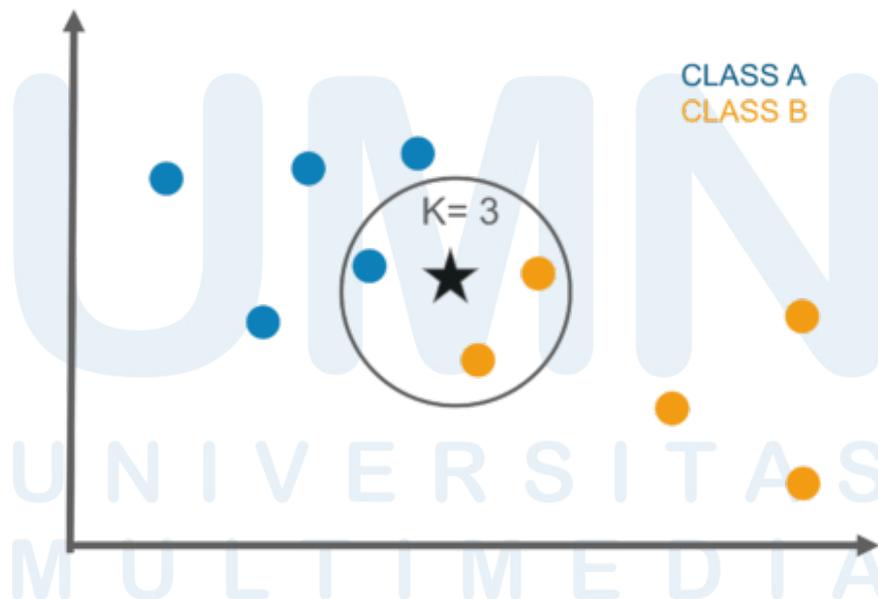
*Evaluation:* Tahap selanjutnya adalah tahap evaluasi dari model-model yang telah dibuat pada tahap selanjutnya. Seluruh model tersebut dilakukan verifikasi serta pengujian sehingga memastikan model sesuai dengan tujuan

dan permasalahan bisnis awal dapat ditangani. Sekaligus pada tahap ini dipilih pemodelan dengan tingkat akurasi, serta penilaian lainnya seperti *misclassification rate* atau dan lain-lain [31].

*Deployment*: Setelah model dievaluasi dan diuji, maka tahap terakhir adalah tahap *deployment* atau penyebaran. Tahap ini diterapkan dengan pembuatan laporan serta memvisualisasikan hasil pemodelan serta *data mining* lainnya. Melalui penyebaran tersebut menghasilkan atau mewujudkan tujuan maupun manfaat dari penelitian [31].

### 2.3.2 *K-Nearest Neighbor*

*K-Nearest Neighbor* (K-NN) adalah algoritma pembelajaran yang diawasi. Teknik *K-Nearest Neighbor* dianggap sebagai salah satu teknik klasifikasi dasar namun sangat efektif. Algoritma ini tidak membuat asumsi umum yang seringkali diperlukan untuk mengklasifikasikan data ketika tidak ada pengetahuan sebelumnya tentang distribusi data. Cara kerja algoritma ini melibatkan pencarian K titik data terdekat dalam *dataset* [32]. Nilai prediksi dari nilai *instance* baru kemudian digunakan untuk mengklasifikasikan lingkungan data yang bersangkutan.



Gambar 2. 3 Contoh Penggunaan Algoritma KNN

Gambar 2.3 menunjukkan contoh penggunaan Algoritma K-NN. Terdapat dua kelas, yaitu kelas A dengan lingkaran biru dan kelas B dengan

lingkaran kuning. Pada tengah gambar, terdapat bintang yang melambangkan sampel uji. Algoritma ini menggunakan konsep jarak untuk menemukan kelas terdekat dengan sampel uji dan memutuskan kelas mana yang sesuai untuk sampel tersebut.

Algoritma K-NN memproses variabel numerik karena bergantung pada *Euclidean Distance*. Oleh karena itu, untuk menghitung jarak antara titik-titik data, rumus *Euclidean Distance* digunakan.

$$\begin{aligned} &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \end{aligned}$$

Rumus 2. 1 Rumus *Euclidean Distance*

Pada rumus 2.1 adalah rumus untuk menghitung *Euclidean Distance*.

## 2.4 Teori tentang *Tools / Software* yang digunakan

### 2.4.1 Python

Python adalah sebuah bahasa pemrograman yang bekerja secara dinamis dan sering digunakan pada pengembangan aplikasi maupun hal lainnya. Python dikenal dengan sintaksis yang fleksibel, dan mudah untuk dipahami, oleh karena itu pemula maupun profesional menjadikan Python sebagai pilihan ideal. Python memiliki berbagai keunggulan, salah satunya adalah sintaks yang digunakan menyerupai dengan bahasa manusia sehingga mudah digunakan. Python memiliki berbagai kegunaan, diantaranya digunakan untuk pengembangan perangkat lunak, pembuatan AI atau kecerdasan buatan, menganalisis data, pengembangan *website*, maupun hal lainnya [33]. Python yang dikembangkan oleh Guido van Rossum, dengan berbagai keunggulannya mampu menangani berbagai tugas pemrograman, dimana bahasa pemrograman Python ini dapat digunakan pada *platform* salah satunya yaitu Google Colaboratory.

### 2.4.2 Google Colab

Google Colab atau kependekan dari Google Colaboratory merupakan sebuah *software* produk dari *Google Research*. *Software* ini memungkinkan mengeksekusi kode Python melalui *browser*. Google Colab dikembangkan dengan menggunakan basis Jupyter notebook dengan bahasa pemrograman

Python [34]. Google Colab memiliki keunggulan dimana menggunakan sistem yang memungkinkan *multi-user*. Fitur *multi-user* ini memungkinkan berbagai pengguna dapat melakukan atau menulis kode secara *online* dan bersamaan. Disediakan *environment open source* untuk berbagai pengguna yang membantu mempelajari mengenai Python. Terdapat berbagai fitur serta *library* pada Google Colab memungkinkan penggunaanya untuk melakukan preparasi data maupun pemodelan data [35].

