BAB II

TINJAUAN PUSTAKA

II.1 *Game Online* PUBG

Game atau permainan online sesuatu yang terdiri dari peraturan, budaya, dan play atau permaianan. Dalam sebuah permaianan, pemain terlibat dan mengikuti aturan sesuai yang telah ditetapkan dalam sebuah permainan atau game tersebut yang permaianan tersebut merupakan buatan atau rekayasa. Dalam sebuah permaianan terdapat aturan yang tidak boleh dilanggar dalam bermain. Dibuatnya game ini bertujuan untuk mengasah intelektual manusia atau disebut (Intelektual playability). Game adalah arena untuk membuat pemain merasa puas dalam menjalankan aksi-aksinya. Ada target-target yang dimainkan secara maksimal (Safitri, 2020).

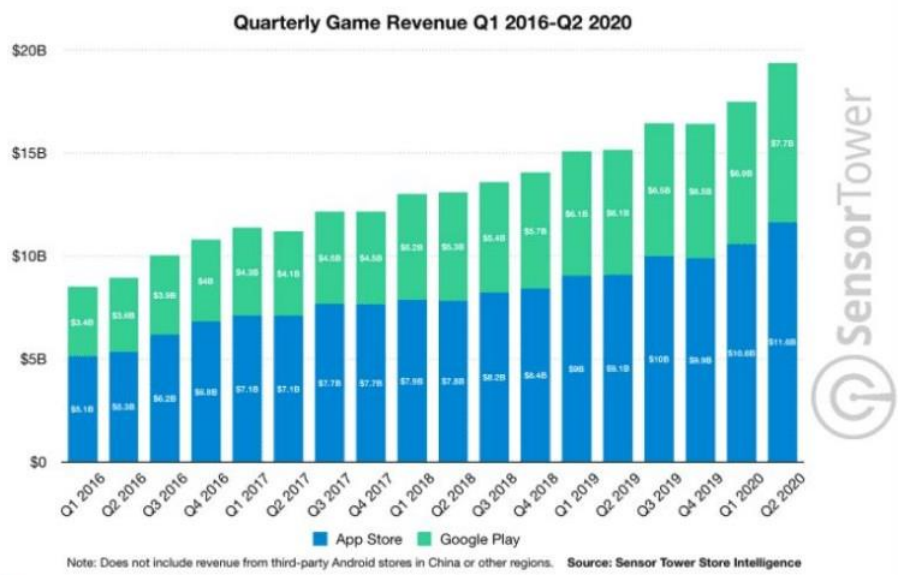
PUBG (PlayerUnknown's BattleGround) adalah sebuah game multiplayer kompetitif yang menjadikan "Battle-Royale" sebagai genre utama. Battle-Royal sendiri merupakan sebuah filem klasik jepang populer pada tahun 2000 silam, yang diadaptasikan dari sebuiah novel keluaran tahun 1999.

PUBG memuat pertempuran 100 orang secara bersamaan di sebuah area yang besar, yang semuanya datang tanpa perbekalan apapun. Setiap dari mereka harus memperkuat dan mempersenjatai diri mereka dengan apapun yang mereka temukan di arena yang ada, dari sekedar panci penggorengan untuk senjata melee, boody armor untuk menahan sedikit laju peluru, hingga senjata api kaliber berat. Bisa juga menemukan kendaraan air atau darat untuk ekstra mobilitas. Ini perlu bertahan

hingga akhir, dan menjadi satu-satunya player yang selamat ditengah pertempuran. Tidak banyak kesempatan bersembunyi hanya satu tempat saja, karena daerah dimana player bisa bergerak dan terus diperkecil seiring dengan waktu berjalan di sebuah arena besar yang terus memaksa pemain untuk berhadapan langsung dengan satu sama lain. Selain bermain solo, juga bisa bermain dalam format Duo (2 orang) dan juga squad (4 orang), baik secara acak atau dengan mengundang teman dari Friend List. Bermain dalam format kooperatif seperti ini lebih sulit karena lawan akan bergerak dalam format strategi tertentu, maka perlu memikirkan lebih banyak strategi untuk kemenangan (Fauzi, 2019).

PUBG yang sudah diunduh sekitar 400 juta pada tahun 2019 dan pengguna harian sampai 50 juta merupakan game battle royal yang dimainkan secara online dan secara bersama-sama sehingga membuat pemain kecanduan game ini (Ilham Rofiqi & Hindarto, 2021). Menurut situs resmi https://www.businessofapps.com/ PUBG Mobile adalah game terlaris pada tahun 2020, menghasilkan pendapatan $2.6 miliar, Telah diunduh lebih dari 730 juta kali.

Berdasarkan data yang bersumberr dari https://sensortower.com/ dilaporkan bahwa pendatana game online PUBG untuk kuartal kedua 2020 meningkat menjadi 10.3% dibandingkan dengan kuartal pertama dan PUBG pun kembali menjadi game terlaris peringkat ke 2 diseluruh dunia.



Gambar II.1 Hasil Survei Sensor Tower

II.2 Supervised Learning

Supervised learning merupakan salah satu metode untuk mengklasifikasikan masing-masing objek dalam data ke beberapa kelas. Pada supervised learning setiap objek pada suatu data memiliki fitur, yaitu ciri-ciri yang ada pada masing-masing objek. Setiap objek dalam suatu data memiliki jumlah fitur yang sama. Fitur digunakan sebagai input untuk menentukan kelas pada objek. Dalam supervised learning, kelas dari masing-masing objek sudah diketahui. Oleh karena itu, permasalahan yang dihadapi dalam supervised learning adalah bagaimana memetakan objek ke dalam kelas yang tepat menggunakan fitur-fitur yang dimiliki oleh setiap objek.

Klasifikasi pada supervised learning dilakukan dengan melakukan training (pelatihan) untuk membentuk model. Classifier (algoritma klasifikasi) akan membentuk model yang beradaptasi sesuai dengan fitur-fitur yang ada pada data.

Model yang dihasilkan dapat berupa tree, rule, atau suatu fungsi yang dapat memprediksikan suatu kelas berdasarkan fitur-fitur yang dimiliki oleh data tersebut. Langkah selanjutnya adalah melakukan validasi terhadap model yang dihasilkan.(Dougherty, 2012)

II.3 Text Mining

Text mining adalah proses ekstraksi pola berupa informasi dan pengetahuan yang sebelumnya tidak diketahui pada sejumlah besar sumber data yang berupa teks (Mahmoud et al., 2015). Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Proses yang umum dilakukan oleh penambangan teks di antaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks, dan lain-lain (DHARMAWAN, n.d.).

Media sosial yang merupakan sumber potensial data terstruktur dianggap sebagai sumber informasi intelijen pasar dan pelanggan yang berharga. Banyak perusahaan yang menggunakan *text mining* untuk menganalisis atau memprediksi kebutuhan pelanggan dan menilai persepsi merek mereka. Analisis teks dapat mengatasi isu-isu dengan menganalisis volume besar dari data terstruktur,

mengeluarkan pendapat, emosi dan sentimen dan hubungan mereka dengan merek dan produk. (Mahmoud et al., 2015).

II.4 Algoritma Naive Bayes

Naïve Bayes Classifier (NBC) adalah sebuah pengklasifikasi sederhana berdasarkan penerapan teorema Bayes (dari statistic Bayesian) dengan asumsi independen (naif) yang kuat. Sebuah istilah yang lebih deskriptif untuk model probabilitas yang digaris bawahi adalah "model fitur independen". Dalam terminologi sederhana, sebuah NBC mengasumsikan bahwa kehadiran (atau ketiadaan) fitur tertentu dari suatu kelas tidak berhubungan dengan kehadiran (atau ketiadaan) fitur lainnya. NBC hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (ratarata dan varian dari variabel) yang diperlukan untuk klasifikasi. Karena variabel diasumsikan independen, hanya varian dari variabel-variabel untuk setiap kelas yang perlu ditentukan dan bukan keseluruhan covariance matrix.(Nooraeni et al., 2020)

Berikut Tahapan dari proses algoritma Naïve Bayes adalah sebagai berikut:

1. Menghitung jumlah *class*/label.
2. Menghitung jumlah kasus yang sama dengan *class* yang sama.
3. Mengalikan semua variabel *class*.
4. Membandingkan hasil semua variabel.

Algoritma Naïve Bayes *classification* memiliki persamaan yang dapat dijadikan acuan untuk menghitung nilai probabilitas dalam pengambilan suatu keputusan, adapun persamaannya dapat dilihat pada persamaan II.1 berikut.

$$P(X|H) = \frac{P(H|X) \cdot P(X)}{P(H)} \quad \text{(II.1)}$$

di mana:

X	:	Data dengan *class* yang belum diketahui
H	:	Hipotesis data X merupakan suatu *class* spesifik
$P(H\|X)$	:	Probabilitas hipotesis H berdasarkan kondisi X
$P(H)$	:	Probabilitas hipotesis H
$P(X\|H)$	:	Probabilitas X berdasarkan kondisi pada hipotesis H
$P(X)$	:	Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. (Lubis et al., 2021)

II.5 *Support Vector Machine* (SVM)

Support Vector Machines (SVM) adalah metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Metode ini dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. *SVM* termasuk klasifikasi yang memberi aturan untuk mengubah dokumen teks menjadi vector sebelum diklasifikasi. Biasanya teks dokumen diubah menjadi vector multidimensional tf-idf. (Buntoro, 2017)

Konsep SVM mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. *Pattern* yang merupakan anggota dari dua buah kelas: +1 dan -1 dan berbagi alternative garis pemisah (*discrimination boundaries*).

Margin adalah jarak antara hyperplane tersebut dengan pattern terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Usaha untuk mencari lokasi hyperplane ini merupakan inti dari proses pembelajaran pada SVM. Secara intuitif, model SVM merupakan representasi dari data sebagai titik dalam ruang, dipetakan sehingga kategori contoh terpisah dibagi oleh celah jelas yang selebar mungkin. Data baru kemudian dipetakan ke dalam ruang yang sama dan diperkirakan termasuk kategori berdasarkan sisi mana dari celah data tersebut berada. (Adijaya, 2020)

SVM bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperlane* terbaik yang memisahkan dua buah kelas pada input *space*. Prinsip dasar SVM adalah klasifikasikan linear. Pada SVM pemisahan yang baik dicapai oleh *hyperplane* yang memiliki jarak terbesar ke titik data *training* terdekat dari setiap kelas (margin fungsional disebut), karena pada umumnya semakin besar margin semakin rendah *error* generalisasi dari pemilah. Misalkan terdapat m *data training* $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ adalah sampel data dan $y_1 \in \{1, -1\}$ adalah target atau kelas dari sampel data. Misalkan juga bahwa data untuk kedua kelas terpisah maka ingin dicari fungsi pemisah (*hyperplane*) (Wu & Kumar, 2009)

$$f(x) = xw + b \qquad \text{(II.2)}$$

di mana w adalah parameter bobot dan b adalah parameter bias.

Pattern $\vec{x_i}$ yang termasuk kelas +1 (sampel positif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan

$$x_i w + b > 0 \text{ untuk } y_i = 1 \qquad \text{(II.3)}$$

Pattern $\vec{x}_s$ yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan

$$x_i w + b < 0 \text{ untuk } y_i = -1 \quad \text{(II.4)}$$

Masalah mencari parameter w dan b yang optimal agar diperoleh *hyperplane* yang optimal merupakan *quadratic programming*.

$$\min_{\&,b} \frac{1}{2} w^T w + C \sum_{i*1}^{m} \mu_i \quad \text{(II.5)}$$

Masalh ini dapat dipecahkan dengan berbagi teknik komputasi, diantaranya dengan *Lagrange Multiplier*.

$$\sum_{i*1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j*1}^{l} \alpha_i \alpha_j y_i y_j \vec{x}_s \vec{x}_y \quad \text{(II.6)}$$

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif $(\alpha_i \geq 0)$. Dari hasil perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *Support Vector Machine*.

SVM pada prinsipnya adalah klasifikasi linear. Tetapi SVM memiliki keunggulan dalam klasifikasi untuk problem non-linier. Pada prosesnya data diproyeksikan ke ruang vektor baru, sering disebut feature space, berdimensi lebih tinggi sedemikian hingga data itu dapat terpisah secara linier. Selanjutnya, di ruang baru, SVM mencari hyperplane optimal yaitu bekerja sebagai klasifier linear.

Dapat diasumsikan bahwa kedua belah kelas dapat terpisah secara sempurna oleh *hyperplane* (*linear separable*). Akan tetapi, pada umumnya dua belah kelas pada *input space* tidak dapat terpisah secara sempurna (*non linear separable*).

Untuk mengatasi masalah ini, SVM dirumuskan ulang dengan memperkenalkan metode *Margin Soft*. Cortes et al. (1995) menyarankan ide margin maksimal dimodifikasi yang memungkinkan untuk contoh *mislabeled*. Jika terdapat hyperplane yang dapat memecah contoh "ya" dan "tidak", metode *Margin Soft* akan memilih hyperplane yang membagi contoh-contoh sebersih mungkin. Metode ini memperkenalkan variabel slack ξi ($\xi i \geq 0$), yang mengukur tingkat kesalahan klasifikasi.

$$y_i(< \vec{w}.\vec{x_i} > +b) \geq 1 - \xi i \quad (i = 1,2, \dots, l) \tag{II.7}$$

Sedangkan objective function (2.5) yang dioptimasikan menjadi.

$$\text{minimze}\frac{1}{2}\|\vec{w}\|^2 + C \sum_{i=1}^{l} \xi_i \tag{II.8}$$

C merupakan parameter yang mengkontrol *trade-off* antara margin dan error klasifikasi ξ. Semakin besar nilai C, berarti nilai akhir terhadap kesalahan menjadi semakin besar, sehingga proses training menjadi lebih ketat. Parameter C ditentukan dengan mencoba beberapa nilai dan dievaluasi efeknya terhadap akurasi yang dicapai oleh SVM. (Norwawi, 2020)

II.6 Text Preprocessing

Data pre-processing adalah teknik *data mining* yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Langkah data pre-processing diperlukan untuk menyelesaikan beberapa jenis masalah termasuk *noisy data*, data redundansi, nilai data yang hilang, dll.

Adapun langkah-langkah *data pre-proccesing* adalah tokenisasi, *case folding, stemming, filtering,* dan *labelling*.

1. Tokenisasi

Tokenisasi merupakan proses pemisahan suatu rangkaian karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca.

2. Case Folding

Case folding merupakan proses mengubah semua huruf dalam suatu dokumen atau kalimat menjadi huruf kecil. Case folding digunakan untuk mempermudah pencarian. Tidak semua data konsisten dalam penggunaan huruf kapital.

3. Stemming

Kata-kata yang sudah diubah menjadi huruf kecil perlu dilakukan pengecekan. *Stemming* digunakan untuk menyeragamkan kata sehingga mengurangi daftar kata yang ada pada data latih.

4. Filtering

Filtering / eliminasi *stopwords* memiliki banyak keuntungan, yaitu akan mengurangi *space* pada tabel *term index* hingga 40% atau lebih. Proses *stopword removal* merupakan proses penghapusan *term* yang tidak memiliki arti atau tidak relevan.

5. Labelling

Labelling berasal dari kata label. Label berarti karakter atau himpunan karakter yang digunakan untuk mengidentifikasi suatu variabel atau bagian dari data atau berkas. *Labelling* / proses pemberian label ada 2 (dua), yaitu pemberian label

kepada *token* dengan kata penguat (*exaggeration*) dan kata negasi (*negation*). (Firdaus, 2017).

II.7 Pembobotan *Term Frequency- Document Frequency* (TF-IDF)

Pembobotan kata adalah proses pemberian bobot untuk setiap kata yang terdapat dalam sebuah dokumen. Dalam pencarian informasi peringkat berdasarkan frekuensi kata, salah satu metode yang paling populer adalah metode TFIDF (*Term Frequency-Inversed Document Frequency*).(Hidayat, 2015)

Dalam metode TF-IDF, *Term Frequency* lebih berfokus pada istilah yang sering muncul dalam suatu dokumen sedangkan *Inverse Document Frequency* lebih berfokus pada pemberian bobot rendah untuk istilah yang muncul dalam banyak dokumen. Persamaan TF-IDF adalah sebagai berikut:

$$IDF(w) = \log \frac{N}{DF(w)} \qquad \text{(II.9)}$$

$$W\,dt = TF \cdot IDF \qquad \text{(II.10)}$$

$$TF - IDF(w,d) = \frac{TF - IDF(w,d)}{\sqrt{\sum_{n=1}^{\&} TF - IDF(w,d)^2}} \qquad \text{(II.11)}$$

Keterangan:

$IDF(w)$	:	bobot kata dalam seluruh dokumen
w	:	sebuah kata
$TF(w,d)$	:	Jumlah kemunculan kata w dalam dokumen d
$IDF(w,d)$	:	invers DF dari kata w
N	:	jumlah seluruh dokumen

$DF(w)$: jumlah dokumen yang memuat kata w

II.8 Pengujian

Pada pengujian Naïve Bayes dan SVM akan dilakukan pengujian akurasi yaitu menggunakan *Confussion matrix*. Confusion matrix adalah suatu metode yang umumnya digunakan untuk melakukan perhitungan tingkat akurasi pada data mining. Confusion matrix memuat informasi tentang klasifikasi yang diprediksi dengan benar oleh sebuah sistem klasifikasi. (Cavalin & Oliveira, 2018)

Dalam *confusion matrix* terdapat jenis klasifikasi *binary* yang hanya memiliki dua output kelas yaitu:

Tabel II.1 Klasifikasi Binary

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (*True Positive*)	TN (*True Negative*)
Negatif	FP (*False Positive*)	FN (*False Negative*)

Confusion matrix melakukan perhitungan yang menghasilkan 4 *output*, yaitu *recall, precision, accuracy*, dan *error rate*. *Recall* merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Rumus *confusion matrix* adalah sebagai berikut:

$$recall = \frac{TP}{FN + TP} \times 100\% \quad \text{(II.12)}$$

$$precission = \frac{TP}{FP + TP} \times 100\% \quad \text{(II.13)}$$

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% \quad \text{(II.14)}$$

Keterangan:

TP : *true positive*, jumlah data positif dan terklasifikasi dengan benar

TN : *true negative*, jumlah data negatif namun terklasifikasi dengan benar

FN : *false negative*, jumlah data negatif yang terklasifikasi salah

FP : *false positive*, jumlah data positif namun terklasifikasi salah

II.9 Kelebihan *Naive Bayes dan Support Vector Machine*

Adapun yang menjadi Kelebihan *Support Vector Machine* yaitu

- Support Vector Machine adalah jenis algoritma sistem pembelajaran yang digunakan untuk membuat klasifikasi lebih akurat. SVM merupakan model yang berasal dari teori pembelajaran statistika yang akan memberikan hasil yang lebih baik dibandingkan metode lainnya. SVM telah banyak diterapkan pada digit tulisan tangan, pengenalan karakter dan teks. (Vimala & Sharmili, 2018)
- Teknik SVM berakar pada teori pembelajaran statistik dan telah menunjukkan hasil empiris menjanjikan dalam berbagai aplikasi praktis dari pengenalan digit tulisan tangan sampai kategorisasi teks. SVM juga bekerja sangat baik pada data dengan banyak dimensi dan menghindari kesulitan dari permasalahan dimensionalitas.

- SVM digunakan untuk menangani ketidakseimbangan kelas Ketidakseimbangan kelas adalah masalah dalam pembelajaran mesin ketika jumlah total positif dan negatif tidak sama dan pengklasifikasi tidak akan berkinerja baik.(Darwis et al., 2020)

  Adapun Kelebihan Naïve Bayes yaitu

- Naïve Bayes classifier menunjukkan akurasi dan kecepatan yang tinggi bila diterapkan pada database yang besar.
- Tidak membutuhkan jumlah data latih yang besar untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.
- Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang mesin pembelajaran karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana. (Handayani & Pribadi, 2015)

II.10 Penelitian Terkait

Dalam Penelitiannya (Octaviani et al., 2014) Juga menerapkan metode klasifikasi Support Vector Machine (SVM) pada data akreditasi SD di Magelang. Klasifikasi Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang menggunakan metode Support Vector Machine (SVM) dengan data training yang diujikan sebanyak 337 data memiliki akurasi klasifikasi sebesar 100% menggunakan fungsi kernel Gaussian Radial Basic Function (RBF). Sedangkan menggunakan fungsi kernel Polynomial akurasi klasifikasi adalah sebesar 98.810 %.

Dalam Penelitiannya (Darmawan et al., 2018) memprediksi kepuasan pengunjung taman Tabebuya dengan metode algoritma SVM (Support Vector

Machine) dimana eksperimen pada model divalidasi dan dievaluasi dengan Confusion Matrix dan AUC (Area Under Curve) dengan ROC (Receiver Operating Characteristic). Hasil dari penelitian melakukan evaluasi dan validasi dapat disimpulkan bahwa algortima SVM memiliki akurasi dan performa secara rata-rata yaitu sebesar 86,00% dan nilai AUC sebesar 0.947.

Penelitian yang dilakukan oleh (Darwis et al., 2020) yaitu memanfaatkan twitterscrapper dalam mengambil data yang akan digunakan pada penelitian. Yaitu dengan menggunakan dataset dari komentar tweet KPK RI yang menggunakan teks bahasa Indonesia sebagai topik penelitian. Opini akan diolah melalui metode-metode yang disesuaikan pada proses pengambilan teks atau text mining. Metode klasifikasi yang digunakan adalah Support Vector Machine (SVM) dan ekstraksi fitur menggunakan TF-IDF. Dari 2000 data hasil twitter crawling, penelitian ini menghasilkan 1890 data dan 3846 term/kata dari hasil preprocessing lalu dihitung nilai dari kemunculan kata untuk labeling yang menghasilkan sentimen positif, negatif dan netral. Berdasarkan hasil pengujian yang dihasilkan, penerapan metode SVM menghasilkan nilai Akurasi sebesar 82% dan menghasilkan sentimen dengan label negatif lebih besar dengan jumlah 77%, label positif 8% dan label netral 25%.

Penelitian Oleh (Handayani & Pribadi, 2015) yaitu dengan mengadakan sistem klasifikasi teks pelaporan dan pengaduan melalui layanan 110 untuk membantu tugas yang dikerjakan oleh Badan Kepolisisan Negara dalam menyediakan pelaporan dan pengaduan dengan call center 110. Yaitu dengan menggunakan Metode Naive Bayes Classifiers yang adalah salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan

dokumen latih dan dokumen uji. Keduanya dibandingkan melalui beberapa tahap persamaan, yang akhirnya diperoleh hasil probabilitas tertinggi yang ditetapkan sebagai kategori dokumen baru.Hasil penelitian yang dilakukan oleh peneliti yaitu pengklasifikasian teks otomatis pelaporan dan pengaduan mayarakat dengan menggunakan metode Naive bayes Classifiers menghasilkan rata-rata akurasi yang tinggi,yaitu recall 93%, precission 90 %,dan f-measure 92%.

Penelitian oleh (Rahutomo et al., 2018) meneliti tentang Implementasi Twitter Sentimen Analisis Untuk Review Film Menggunakan Algoritma Support Vector Machine. Proses klasifikasi memudahkan melihat opini positif, negatif, atau netral dengan proses klasifikasi menggunakan algoritma Support Vector Machine. Evaluasi akurasi menghasilkan tingkat akurasi yang stabil hingga 76% dari seluruh proses training dan tingkat akurasi ini lebih baik dari algoritma Naïve Bayes Classifier yang hanya stabil sampai dengan 75%.

Penelitian oleh (HR et al., n.d.), mempelajari tentang Analisis sentimen pada twitter terhadap kinerja Komisi Pemberantasan Korupsi (KPK) di Indonesia dengan prosedur Naïve Bayes. Dengan tujuan untuk megakomodasi permasalahan terhadap kinerja KPK berdasarkan data twitter. Hasil pada penelitian ini menunjuka bahwa sebanyak 50,96% akun twitter memiliki sentimen positif dan 49,03% memiliki sentimen negatif. Sedangkan hasil analisis pada data uji menunjukan tingkat akurasi 65,51%, tingkat presisi 51,35%, dan recall 90,47%.

Penelitian oleh (Sitepu et al., 2021) Mengidentifikasi komentar perundungan pada jejaring sosial dengan memanfaatkan 26egati data mining, system diharapkan mampu mengklasifikasikan informasi yang beredar di

masyarakat. Penelititan ini menggunakan 27egati klasifikasi Support Vector Machine karena algoritma tersebut termasuk baik dalam melakukan proses klasifikasi. Data yang digunakan adalah 1000 dataset komentar. Data dikelompokkan dahulu secara manual ke label “bully” dan “tidak bully” kemudian data dibagi menjadi data latih dan data uji. Untuk menguji kemampuan 27egati, data yang sudah diklasifikasikan kemundian dianalisis menggunakan confusion matrix. Hasil penelitian menunjukkan Algoritma SVM mampu melakukan klasifikasi dengan tingkat keakuratan 87,75%, presisi 89% dan Recal 91%. Algoritma SVM mampu merumuskan data latih dengan akurasi sebesar 98,3%.

Tabel II.2 Penelitian Terkait

No	Judul	Penulis	Metode	Hasil	Kekurangan
1	Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (Sd) Di Kabupaten Magelang	Pusphita Anna Octaviani , Yuciana Wilandari , Dwi Ispriyanti	Dengan menggunakan algoritma SVM dan dievaluasi dengan Akurasi RBF dan menggunakan fungsi kernel Polynomial	Hasil klasifikasi SVM dengan akurasi klasifikasi sebesar 100% menggunakan fungsi kernel Gaussian Radial Basic Function (RBF) dan Dengan fungsi kernel Polynomial akurasi klasifikasi adalah sebesar 98.810 %.	Hasil klasifikasi dari 6 SD diklasifikasikan berbeda dengan kelas asli.
2	Implementasi Data Mining Menggunakan Model Svm Untuk Prediksi Kepuasan Pengunjung Taman Tabebuya	Agus Darmawan , Nunu Kustian , Wanti Rahayu	Metode Algoritma Svm (Support Vector Machine) Dimana Eksperimen Pada Model Divalidasi Dan Dievaluasi Dengan Confusion Matrix Dan Auc (Area Under Curve) Dengan Roc (Receiver Operating Characteristic)	Klasifikasi SVM 86,00% dan nilai AUC sebesar 0.947.	Aplikasi dirancang dengan tidak dengan menggunakan software analisis sehingga data dapat dimanipulasi dengan mudah.

No	Judul	Penulis	Metode	Hasil	Kekurangan
3	Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia	Dedi Darwis , Eka Shintya Pratiwi , A. Ferico Octavian syah Pasaribu	Support Vector Machine (SVM) dan ekstraksi fitur menggunakan TF-IDF	Akurasi sebesar 82% dan menghasilkan 29egative29 dengan label 29egative lebih besar dengan jumlah 77%, label positif 8% dan label netral 25%.	Penelitian tidak menggunakan perbandingan dengan metode lain sehingga hasil prediksi tidak dapat diuji akurasinya
4	Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110	Fitri Handayani dan Feddy Setio Pribadi	Metode Naive Bayes Classifiers	Naive bayes Classifiers menghasilkan rata-rata akurasi yang tinggi, yaitu recall 93%, precission 90 %,dan f-measure 92%.	Penelitian ini hanya menggunakan data sebanyak 113 dokumen dan dibagi menjadi 70% data sebagai data latih dan 30% data sebagai data uji
5	Implementasi *Twitter Sentiment Analysis* Untuk *Review* Film Menggunakan Algoritma *Support Vector Machine*	Faisal Rahutomo , Pramana Yoga Saputra , Miftahul Agtamas Fidyawan	Proses klasifikasi menggunakan algoritma Support Vector Machine	Evaluasi akurasi menghasilkan tingkat akurasi yang stabil hingga 75% dari seluruh proses training	Aplikasi dirancang dengan tidak dengan menggunakan software analisis sehingga data dapat dimanipulasi dengan mudah.

No	Judul	Penulis	Metode	Hasil	Kekurangan
6	Analisis Sentimen Pada Twitter Terhadap Kinerja Komisi Pemberantasan Korupsi (KPK) Di Indonesia Dengan Metode Naive Bayes	Hendri Uut Fahrullah	Klasifikasi Naïve Bayes	Hasil analisis pada data uji menunjukan tingkat akurasi 65,51%, tingkat presisi 51,35%, dan recall 90,47%.	Proses stemming dapat tidak menggunakan algoritma lain atau mengembangkan stemming yang sudah ada sehingga beberapa kata yang belum bisa diklasifikasi selanjutnya bisa diklasifikasi.
7	Analisis Kinerja Support Vector Machine dalam Mengidentifikasi Komentar Perundungan pada Jejaring Sosial	Ade Clinton Sitepu, Wanayumini, Zakarias Situmorang	Klasifikasi Support Vector Machine	Algoritma SVM mampu melakukan klasifikasi dengan tingkat keakuratan 87,75%, presisi 89% dan Recal 91%. Dan mampu merumuskan data latih dengan akurasi sebesar 98,3%.	Kalsifikasi hanya menggunakan 2 katerogi yaitu kategori bully dan kategori tidak bully

II.11 Ringkasan Bab II

PlayerUnknown's Battle Ground (PUBG) merupakan game *online multiplayer* dengan genre *battle royale*, yang pada tahun 2020 menjadi game terlaris peringkat ke-2 di seluruh dunia. *Review* pengguna permainan ini, khususnya pada

platform mobile, sangat beragam, sehingga penulis melakukan penelitian prediksi *rating game* PUBG menggunakan kombinasi *supervised learning* dan *text mining*. Dalam penelitian prediksi, *supervised learning* membutuhkan target dan fitur dalam proses pelatihan dan pengujian data. Dalam *text mining*, pola dan informasi yang terdapat di dalam sebuah teks dapat di ekstrak untuk kemudian digunakan sebagai bahan analisis atau memprediksi suatu masalah. Tiga algoritma yang cukup populer digunakan dalam masalah prediksi adalah K-Nearest Neighbors, Naïve Bayes, dan Support Vector Machine. Ketiga algoritma ini membutuhkan target dalam proses prediksi, yang diperoleh dari hasil *text processing* isi komentar review *game online* PUBG. Tokenisasi, Case Folding, Stemming, Filtering, dan Labeling merupakan langkah-langkah umum yang dilakukan *text processing* dalam menghasilkan target bagi algoritma prediksi. Tokenisasi berfungsi untuk memisahkan kata-kata di dalam teks, Case Folding berfungsi untuk mengubah seluruh kata yang dipisahkan menjadi bentuk huruf kecil, Stemming berfungsi untuk menyeragamkan kata, Filtering berfungsi untuk memfilter kata menggunakan daftar *stopword*, sedangkanLabeling berfungsi untuk memberikan identitas akhir dalam bentuk kategori atau variabel terhadap inti dari isi komentar yang diproses.