

**ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK
ANALISIS SENTIMEN BERBASIS ASPEK ULASAN *GAME*
*ONLINE MOBILE LEGENDS: BANG-BANG***

MAR ATUL AJI TYAS UTAMI



**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2022**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Algoritma *Support Vector Machine* untuk Analisis Sentimen Berbasis Aspek Ulasan *Game Online Mobile Legends: Bang-Bang*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2022

Mar Atul Aji Tyas Utami
G14180059

ABSTRAK

MAR ATUL AJI TYAS UTAMI. Algoritma *Support Vector Machine* untuk Analisis Sentimen Berbasis Aspek Ulasan *Game Online* Mobile Legends: Bang-Bang. Supervised by PIKA SILVIANTI and MOHAMMAD MASJKUR.

Kehadiran era teknologi digital dipermudah dengan adanya koneksi internet yang mudah diakses serta menyediakan banyak fitur dan hiburan salah satunya *game online*. Mobile Legends: Bang-Bang merupakan *game online* berjenis MOBA yang populer sejak diluncurkan pada tahun 2016. Kepopuleran tersebut tidak terlepas dari ulasan pengguna yang memberikan informasi dan sentimen berbeda. Penelitian ini akan mengidentifikasi sentimen ulasan pengguna aplikasi berdasarkan aspek *Gameplay*, *Performa*, Visualisasi, dan *Player*. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM). Aplikasi game online Mobile Legends: Bang-Bang cenderung memiliki sentimen negatif berdasarkan aspek *Gameplay*, *Performa*, dan *Player*. Akan tetapi, dari aspek Visualisasi ulasan cenderung memiliki sentimen positif. Hasil evaluasi model dilakukan berdasarkan nilai akurasi, *f1-score*, dan AUC didapatkan bahwa aspek *Gameplay*, *Performa*, dan *Player* memberikan tingkat klasifikasi yang lebih baik dibandingkan aspek Visualisasi.

Kata Kunci: aspect based sentimen analysis, game online, mobile legends: bang-bang, svm

ABSTRACT

MAR ATUL AJI TYAS UTAMI. Support Vector Machine Algorithm for Sentimen Analysis Based on Online Game Reviews Mobile Legends: Bang-Bang. Supervised by PIKA SILVIANTI and MOHAMMAD MASJKUR.

The presence of the digital technology era is facilitated by an internet connection that is easily accessible and provides many features and entertainment, one of which is online games. Mobile Legends: Bang-Bang is a MOBA-type online game that has been popular since its launch in 2016. Currently, Mobile Legends: Bang-Bang is still the top free game on the Google Play Store. This popularity is inseparable from user reviews that provide different information and sentiments. This research will identify the sentiment of application user reviews based on aspects of *Gameplay*, *Performance*, Visualization, and *Player*. The classification method used in this study is the Support Vector Machine (SVM). The online game application Mobile Legends: Bang-Bang tends to have negative sentiments from aspects of *Gameplay*, *Performance*, and *Player*. However, from the Visualization aspect, they tend to have positive sentiments. The results of the evaluation of the model based on the value of accuracy, *f1-score*, and AUC, it was found that the *Gameplay*, *Performance*, and *Player* aspects gave better classification results than the Visualization aspect.

Keywords: aspect based sentiment analysis, game online, mobile legends: bang-bang, svm

© Hak Cipta milik IPB, tahun 2022¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

**ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK
ANALISIS SENTIMEN BERBASIS ASPEK ULASAN *GAME*
*ONLINE MOBILE LEGENDS: BANG-BANG***

MAR ATUL AJI TYAS UTAMI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2022**

Tim Penguji pada Ujian Skripsi:

Judul Skripsi : Algoritma *Support Vector Machine* untuk Analisis Sentimen
Berbasis Aspek Ulasan *Game Online* Mobile Legends: Bang-Bang
Nama : Mar Atul Aji Tyas Utami
NIM : G14180059

Disetujui oleh

Pembimbing 1:

Pika Silvianti, M.Si

Pembimbing 2:

Ir. Mohammad Masjkur, M.S

Diketahui oleh

Ketua Departemen Statistika:

Dr. Anang Kurnia, M.Si

NIP 197308241997021001

Tanggal Ujian:

Tanggal Lulus:

PRAKATA

Puji dan syukur kehadiran Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Desember 2021 ini ialah klasifikasi sentimen dengan judul Algoritma *Support Vector Machine* untuk Analisis Sentimen Berbasis Aspek Ulasan *Game Online Mobile Legends: Bang-Bang*.

Terima kasih penulis ucapkan kepada pihak-pihak yang telah memberi dukungan dan arahan selama proses penulisan skripsi ini berlangsung, diantaranya:

1. Keluarga yang selalu memberikan dukungan berupa doa dan semangat kepada penulis.
2. Ibu Pika Silvianti, M.Si dan Bapak Ir. Mohammad Masjkur, M.S selaku dosen pembimbing skripsi Departemen Statistika IPB yang telah membimbing dan memberikan arahan.
3. Dosen penguji
4. Seluruh dosen, staf dan TU Departemen Statistika IPB yang membantu dan melancarkan penulisan.
5. Mahasiswa/i Departemen Statistika IPB angkatan 55.
6. Seluruh pihak terkait yang telah membantu penelitian ini dan tidak dapat dituliskan satu persatu.

Penulis sadar karya ilmiah ini masih jauh dari sempurna, penulis mohon maaf atas kekurangan dan kesalahan yang terdapat pada karya ilmiah ini. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan..

Bogor, Agustus 2022

Mar Atul Aji Tyas Utami

DAFTAR ISI

DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Analisis Sentimen Berbasis Aspek	3
2.2 Pembobotan TF-IDF	3
2.3 <i>Syntethic Minority Oversampling Technique</i> (SMOTE)	4
2.4 <i>Support Vector Machine</i> (SVM)	4
2.5 Evaluasi Model	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis Data	9
IV HASIL DAN PEMBAHASAN	15
4.1 Eksplorasi Data	15
4.2 <i>Term Frequency – Inverse Document Frequency</i> (TF-IDF)	19
4.3 <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	19
4.4 Pemodelan Klasifikasi	21
4.5 Evaluasi Model	22
V SIMPULAN DAN SARAN	24
5.1 Simpulan	24
5.2 Saran	24
DAFTAR PUSTAKA	25
RIWAYAT HIDUP	27

DAFTAR TABEL

1	Ilustrasi <i>confusion matrix</i>	7
2	Nilai rentang AUC dan tingkatan klasifikasinya	8
3	Rincian peubah hasil penarikan data	9
4	Contoh hasil penarikan data	10
5	Contoh ulasan yang sudah diberi label	10
6	Contoh ulasan setelah cleaning data	11
7	Contoh kata setelah normalisasi	11
8	Contoh ulasan setelah <i>tokenizing</i>	12
9	Contoh ulasan setelah <i>stopwords removal</i>	12
10	Contoh ulasan setelah <i>stemming</i>	12
11	Contoh hasil pembobotan TF-IDF	13
12	Peubah data pemodelan SVM	14
13	Banyaknya data hasil pelabelan	16
14	Hasil pembagian data <i>training</i> dan data <i>testing</i> per aspek	20
15	Presentase data sebelum dan sesudah dilakukan SMOTE	20
16	<i>Confusion matrix</i> aspek <i>Gameplay</i>	21
17	<i>Confusion matrix</i> aspek <i>Performa</i>	21
18	<i>Confusion matrix</i> aspek Visualisasi	22
19	<i>Confusion matrix</i> aspek <i>Player</i>	22
20	Hasil performansi klasifikasi SVM	22
21	Nilai AUC masing-masing aspek	23

DAFTAR GAMBAR

1	Ilustrasi <i>hyperplane</i> yang mungkin dalam set data	5
2	Ilustrasi <i>hyperplane</i> optimal dan margin sesuai	6
3	Ilustrasi fungsi <i>kernel</i> pada data non linear	7
4	Diagram alir penelitian	9
5	Diagram batang frekuensi ulasan per bulan	15
6	Histogram polaritas sentimen keseluruhan ulasan	15
7	Banyaknya masing-masing sentimen pada setiap aspek	16
8	Awan kata untuk data aspek <i>Gameplay</i> sentimen positif (a) dan sentimen negatif (b)	17
9	Awan kata untuk data aspek <i>Performa</i> sentimen positif (a) dan sentimen negatif (b)	18
10	Awan kata untuk data aspek Visualisasi sentimen positif (a) dan sentimen negatif (b)	18
11	Awan kata untuk data aspek <i>Player</i> sentimen positif (a) dan sentimen negatif (b)	19

I PENDAHULUAN

1.1 Latar Belakang

Ilmu pengetahuan dan teknologi di era globalisasi berkembang sangat pesat dan semakin canggih. Kehadiran era teknologi digital dipermudah dengan adanya koneksi internet yang mudah diakses serta menyediakan banyak fitur dan hiburan yang dapat dinikmati dalam kehidupan sehari-hari, salah satunya *game online*. *Game online* merupakan permainan yang dapat diakses secara *online* oleh banyak pemain menggunakan jaringan internet (Lebho *et al.* 2020). *Game online* memiliki berbagai jenis, salah satu jenisnya adalah MOBA atau *Multiplayer Online Battle Arena*. Secara umum, peraturan *game* MOBA adalah pemain akan memainkan satu karakter pahlawan dan tergabung dalam sebuah tim. Terdapat dua tim yang saling bermusuhan dalam satu kali permainan. Masing-masing tim bertugas untuk menghancurkan lawan sekaligus mempertahankan tim sendiri untuk memenangkan permainan (Simorangkir *et al.* 2018). Bermain *game online* saat ini menjadi hiburan dan hobi baru yang sangat diminati ketika memiliki waktu luang. Banyaknya pilihan *game* yang tersedia serta dapat diunduh secara gratis membuat pengguna *game online* terutama MOBA semakin meningkat.

Mobile Legends: Bang-Bang merupakan *game online* berjenis MOBA yang dibuat dan dikembangkan oleh Moonton dari Shanghai, China yang diluncurkan pada tahun 2016. Google dalam aplikasi *store*-nya yaitu Google Play Store merupakan salah satu *platform* penyedia aplikasi Mobile Legends: Bang-Bang yang dapat diunduh secara gratis. Menurut Fendy Tan, pengembang game Mobile Legends: Bang-Bang di Indonesia, di tahun 2017 telah diunduh sebanyak 35 juta kali dengan delapan juta pengguna aktif harian. Tahun 2018, sebanyak 43 juta pengguna aktif bulanan berada di Asia Tenggara dengan lima puluh persen jumlah tersebut merupakan pengguna Indonesia (Yogatama *et al.* 2019). Menurut data dari website Esports Chart, Indonesia menduduki *top-1 country* dengan Mobile Legends: Bang-Bang sebagai *game* yang populer dimainkan. Saat ini, Mobile Legends: Bang-Bang juga masih menjadi *top free games* di dalam Google Play Store.

Kepopuleran aplikasi Mobile Legends: Bang-Bang dapat dipengaruhi ulasan yang diberikan oleh penggunanya. Sebagian besar pengguna akan membaca ulasan terlebih dahulu sebelum memutuskan untuk mengunduh aplikasi. Semakin banyak ulasan pembaca akan sulit untuk menemukan informasi yang relevan terhadap preferensi mereka. Informasi yang terdapat dalam ulasan perlu diekstrak terlebih dahulu untuk mengetahui opini atau sentimen pengguna terhadap aplikasi. Studi komputasi mengenai pendapat, sikap, dan emosi seseorang terhadap suatu entitas tersebut dapat dilakukan dengan analisis sentimen. Akan tetapi, informasi dalam sebuah ulasan dapat menyampaikan sentimen berbeda terhadap aspek yang berbeda juga sehingga diperlukan pendekatan sentimen berbasis aspek (Liu 2012).

Klasifikasi sentimen dilakukan untuk melihat polaritas dari ulasan aplikasi *game online* Mobile Legends: Bang-Bang mengandung opini positif atau opini negatif. Hasil klasifikasi sentimen tersebut dapat digunakan oleh pengguna untuk mengetahui kelebihan dan kekurangan dari aplikasi serta dapat dijadikan evaluasi keputusan peningkatan mutu oleh pengembang game. Metode yang digunakan

untuk menunjang klasifikasi dalam penelitian ini adalah *Support Vector Machine* (SVM). SVM merupakan algoritma yang telah mendapatkan pengakuan luas dengan akurasi yang baik dalam klasifikasi. Menurut penelitian yang telah dilakukan Iskandar dan Nataliani (2021) mengenai perbandingan *Naïve Bayes*, SVM, dan KNN untuk analisis sentimen gadget berbasis aspek, terbukti bahwa model klasifikasi SVM menunjukkan hasil terbaik dengan rata-rata akurasi sebesar 96,43%. Penelitian lainnya yang telah dilakukan oleh Ilmawan dan Mude (2020) terkait perbandingan metode klasifikasi SVM dan *Naïve Bayes* untuk analisis sentimen menghasilkan bahwa SVM Classifier memiliki nilai akurasi yang lebih tinggi dibandingkan *Naïve Bayes*, yaitu sebesar 81,46%.

1.2 Tujuan

Tujuan dari penelitian ini yaitu:

- a. Mengidentifikasi sentimen ulasan pengguna aplikasi *game online* Mobile Legends: Bang-Bang di Google Play Store berdasarkan aspek *Gameplay*, *Performa*, *Visualisasi*, dan *Player*.
- b. Menerapkan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi sentimen berbasis aspek pada aplikasi *game online* Mobile Legends: Bang-Bang.

II TINJAUAN PUSTAKA

2.1 Analisis Sentimen Berbasis Aspek

Analisis sentimen secara umum dikategorikan kedalam tiga tingkatan, yaitu tingkat kalimat, tingkat dokumen, dan tingkat aspek. Analisis tingkat aspek atau biasa disebut analisis sentimen berbasis aspek digunakan untuk mengetahui sasaran dari suatu opini positif atau negatif. Analisis sentimen berbasis aspek mencakup sekaligus entitas dan aspek dari suatu opini. Secara umum, analisis sentimen berbasis aspek memiliki dua tugas utama, yaitu ekstraksi aspek dan klasifikasi sentimen aspek. Ekstraksi aspek bertujuan mengetahui aspek dan entitas yang sedang dievaluasi sedangkan klasifikasi sentimen aspek untuk mengetahui arah dari sentimen suatu aspek tersebut yaitu positif atau negatif (Liu 2012).

Penelitian ini menggunakan empat kategori aspek untuk mengidentifikasi sentimen ulasan. Aspek tersebut, yaitu aspek *Gameplay*, *Performa*, *Visualisasi*, dan *Player*. Aspek *Gameplay* menggambarkan proses interaksi pemain dengan karakter game pada saat permainan berlangsung seperti aturan, alur, tujuan, dan pengalaman dalam memainkan game. Aspek *Performa* menggambarkan kinerja sistem yang telah dibuat untuk dapat memainkan sebuah game. Aspek *Visualisasi* menggambarkan segala sesuatu yang berhubungan dengan bentuk visual atau tampilan grafik sebuah game seperti animasi, karakter, dan kualitas video gambar. Aspek *Player* menggambarkan pemain yang terlibat dan berinteraksi satu sama lain di dalam game.

2.2 Pembobotan TF-IDF

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan proses transformasi data teks ke dalam data numerik untuk dilakukan pembobotan pada setiap kata atau fiturnya. TF-IDF juga merupakan sebuah ukuran statistik untuk mengevaluasi seberapa pentingnya sebuah kata di dalam dokumen (Septian *et al.* 2019). Adapun rumus dari pembobotan TF-IDF sebagai berikut:

$$TFIDF_{ab} = TF_{ab} \times IDF_a = \frac{freq_{ab}}{\max freq_{ab}} \times \left(1 + \log \frac{N}{df_a}\right) \quad (1)$$

Keterangan:

$TFIDF_{ab}$: bobot dari kata a pada dokumen ke b

TF_{ab} : *term frequency*

IDF_a : *inverse document frequency*

$freq_{ab}$: banyaknya kemunculan kata a dalam dokumen ke b

$\max freq_{ab}$: total kata a pada dokumen ke b

N : total dokumen

df_a : banyaknya dokumen yang mengandung kata a

Pembobotan kata dengan TF-IDF ini akan menghasilkan bobot kata yang semakin besar apabila sering muncul dalam suatu dokumen, sebaliknya bobot kata akan semakin kecil apabila muncul dalam banyak dokumen.

2.3 Syntethic Minority Oversampling Technique (SMOTE)

Syntethic Minority Oversampling Technique (SMOTE) merupakan salah satu metode untuk mengatasi kelas data tidak seimbang. Data dikatakan tidak seimbang apabila suatu kelas data memiliki jumlah amatan yang jauh berbeda dibandingkan dengan kelas lainnya. Kelas dengan jumlah amatan yang sedikit disebut kelas minoritas sedangkan kelas lainnya disebut kelas mayoritas. Penggunaan data tidak seimbang untuk pembuatan model memiliki pengaruh yang sangat besar pada model yang dihasilkan. Data akan cenderung diliputi kelas mayoritas dan mengacuhkan kelas minoritas apabila pengolahan algoritma tidak menghiraukan ketidakseimbangan data (Chawla *et al.* 2002).

SMOTE dalam mengatasi ketidakseimbangan kelas data dengan cara membangkitkan data buatan (sintetik) pada kelas data minoritas. Pembangkitan data dilakukan secara *oversampling* dengan memanfaatkan konsep *k*-tetangga terdekat yang dipilih secara acak. Perhitungan jarak tetangga terdekat untuk data bertipe numerik dihitung menggunakan jarak Euclidean. Data gabungan bertipe numerik dan kategorik untuk perhitungan jarak menggunakan jarak Euclidean dan menggunakan nilai median dari simpangan baku peubah numerik sebagai selisih nilai peubah kategorik (Umma *et al.* 2021). Jarak Euclidean dapat dicari menggunakan persamaan sebagai berikut:

$$D(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (2)$$

Keterangan:

$D(x, y)$: jarak antara amatan x dengan y
 x : amatan x
 y : amatan y
 k : merepresentasikan amatan ke-1 hingga n
 n : banyaknya amatan

Pembangkitan data buatan pada data bertipe numerik berbeda dengan data bertipe kategorik. Data kategorik dibangkitkan dengan menentukan kategori yang paling sering muncul (modus) diantara amatan dengan *k*-tetangga terdekatnya. Sementara itu, pembangkitan data bertipe numerik dilakukan berdasarkan persamaan berikut:

$$X_{baru} = x + (x^* - x) \times rand[0,1] \quad (3)$$

Keterangan:

X_{baru} : data buatan hasil replikasi
 x : data yang akan direplikasi
 x^* : data yang memiliki jarak tetangga terdekat yang terpilih
 $rand[0,1]$: bilangan acak antara 0 sampai 1

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) termasuk salah satu metode regresi atau pengklasifikasi data berdasarkan data yang ada sebelumnya dan terlebih dahulu pemodelannya disupervisi. Tiga komponen utama metode SVM yaitu *support*

vectors, *hyperplane*, dan margin. *Support vectors* merupakan dua data yang memiliki jarak terdekat dari kelas yang berbeda. *Hyperplane* merupakan garis pembatas antar *support vectors* dan margin merupakan jarak antar *support vectors* dengan *hyperplane* (Hutami 2021). Penentuan klasifikasi SVM menggunakan *decision boundary* (batas keputusan) berdasarkan data latih sehingga dapat dibentuk model linear atau *hyperplane* yang paling optimal untuk mengklasifikasi data. Sederhananya, konsep SVM dapat dijelaskan sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas yang berbeda pada ruang inputnya (Christianini dan Shawe-Taylor 2000).

Data yang tersedia dinotasikan $\vec{x}_i \in R^d$, label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$ dengan l adalah banyaknya data. Kelas -1 dan +1 merupakan kelas negatif dan positif yang diasumsikan dapat terpisah secara sempurna oleh *hyperplane* berdimensi d yang didefinisikan sebagai berikut:

$$\vec{w} \cdot \vec{x}^T + b = 0 \quad (4)$$

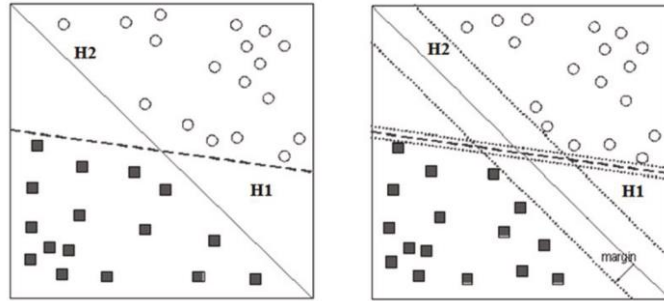
dengan \vec{w} adalah vektor bobot, \vec{x}^T adalah vektor input, dan b adalah bias (merekpresentasikan posisi bidang relatif terhadap pusat koordinat). Sesuai dengan fungsi keputusan sebagai berikut:

$$f(x) = \text{sgn}((\vec{w} \cdot \vec{x}) + b) \quad (5)$$

Ilustrasi konsep *hyperplane* yang mungkin untuk set data dapat dilihat pada Gambar 1. Data didefinisikan dengan persamaan sebagai berikut:

$$\vec{w} \cdot \vec{x}^T + b \geq 0 \text{ jika } d_i = +1 \quad (6)$$

$$\vec{w} \cdot \vec{x}^T + b \leq 0 \text{ jika } d_i = -1 \quad (7)$$



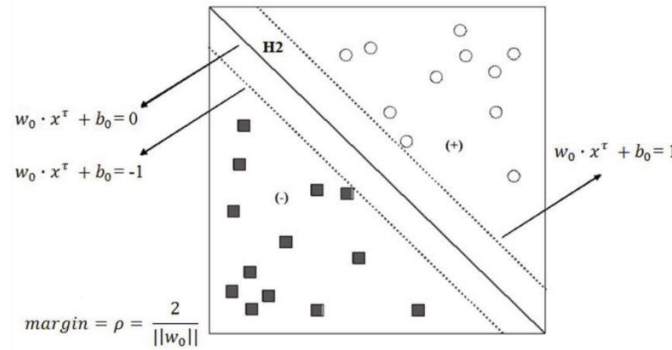
Gambar 1 Ilustrasi *hyperplane* yang mungkin dalam set data

Nilai margin dapat diperoleh dengan memaksimalkan jarak antara *hyperplane* dan titik terdekatnya. *Hyperplane* yang optimal dapat dibangun dengan menyelesaikan persamaan menggunakan *Quadratic Programming* (QP) yang telah dimodifikasi dengan teknik *soft margin* sebagai berikut:

$$\min \tau(\vec{w}, \xi) = \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (8)$$

dengan slack variabel $\xi_i (\xi_i > 0)$ dan C adalah parameter Cost untuk mengontrol tradeoff antara margin dan error klasifikasi ξ . Semakin kecil nilai C maka margin akan semakin lebar sebaliknya semakin besar nilai C maka margin akan semakin

sempit. Ilustrasi penentuan *hyperplane* yang optimal dan margin yang sesuai dapat dilihat pada Gambar 2.



Gambar 2 Ilustrasi *hyperplane* optimal dan margin sesuai

Permasalahan optimasi yang terdapat dalam persamaan dapat dipecahkan menggunakan teknik *Lagrange Multiplier* dengan persamaan seagai berikut:

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (x_i \cdot w + b) - 1) \quad (9)$$

Persamaan *Lagrange Multiplier* tersebut kemudian dimodifikasi sehingga menjadi dualitas *Lagrange Multiplier*, yaitu dengan meminimumkan L terhadap w dan b sehingga didapatkan persamaan sebagai berikut:

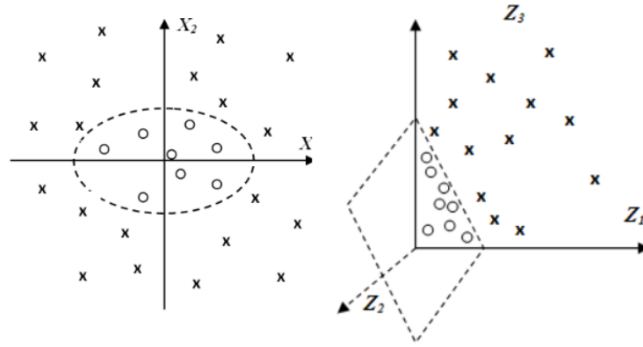
$$L(w, b, a) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (10)$$

$$\text{syarat } \alpha_i \geq 0 \quad (i = 1, 2, 3, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0$$

dengan α_i dan α_j adalah koefisien lagrange data ke- i dan ke- j , y_i dan y_j adalah kelas data ke- i dan ke- j . Nilai α_i yang diperoleh digunakan untuk mencari w . *Support vectors* memiliki nilai $\alpha_i > 0$ sedangkan titik amatan lainnya memiliki $\alpha_i = 0$. Oleh karena itu, fungsi keputusan hanya dipengaruhi oleh *support vectors*. Fungsi keputusan dapat dilihat pada persamaan (5).

2.4.1 Kernel Trick

SVM pada dasarnya hanya bekerja pada data yang dapat dipisahkan secara linear. Mengatasi permasalahan data yang bersifat non-linier, SVM dimodifikasi dengan memasukkan fungsi *kernel* atau biasa disebut dengan *kernel trick*. *Kernel* merupakan suatu fungsi yang memetakan fitur data dari dimensi rendah ke fitur baru dengan dimensi yang relatif lebih tinggi seperti ilustrasi yang dapat dilihat pada Gambar 3 (Prasetyo 2014).



Gambar 3 Ilustrasi fungsi *kernel* pada data non linear

Ada tiga fungsi *kernel* yang biasa digunakan, yaitu *linear kernel*, *polynomial kernel*, dan *radial basis function kernel*. Pemilihan fungsi *kernel* yang tepat dan sesuai menjadi hal yang penting untuk diperhatikan. Kasus data teks seperti klasifikasi umumnya dapat dipisahkan secara linear sehingga fungsi *kernel* yang tepat digunakan adalah *linear kernel* (Joachims 1997). Penelitian ini menggunakan fungsi *linear kernel* yang didefinisikan sebagai berikut:

$$k(\vec{x}_i, \vec{x}_j) = \vec{x}_i^T \vec{x}_j \quad (9)$$

dengan $i, j = 1, 2, 3, \dots, l$; \vec{x}_i dan \vec{x}_j adalah data ke- i dan data ke- j .

2.5 Evaluasi Model

Evaluasi model dilakukan untuk menilai kualitas klasifikasi. Agar dapat mengukur tingkat akurasi dari prediksi klasifikasi yang dihasilkan, sistem klasifikasi tersebut harus dinilai *Performanya*. *Confusion matrix* merupakan salah satu metode yang biasa digunakan dalam perhitungan akurasi pada konsep data mining. Ilustrasi dari *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1 Ilustrasi *confusion matrix*

Kelas aktual	Kelas prediksi	
	Positif	Negatif
Positif	<i>True Positif</i> (TP)	<i>False Negatif</i> (FN)
Negatif	<i>False Positif</i> (FP)	<i>True Negatif</i> (TN)

Penelitian ini menggunakan hasil dari *confusion matrix* untuk mengukur nilai akurasi dan *F1-score*. Perhitungan akurasi dan *F1-score* dapat dilihat pada persamaan berikut:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (10)$$

$$F1 - \text{score} = \frac{TP}{TP + \frac{1}{2}(TN + FP)} \quad (11)$$

Keterangan:

TP : banyaknya kelas positif yang diprediksi benar
 TN : banyaknya kelas negatif yang diprediksi benar
 FP : banyaknya kelas positif yang diprediksi salah
 FN : banyaknya kelas negatif yang diprediksi salah

Performa model juga dapat dilihat menggunakan nilai AUC (*area under curve*). AUC merupakan luas di bawah kurva yang dinamakan kurva *Receiver Operating Characteristic* (ROC). AUC merupakan bentuk skalar dari ROC yang nilainya antara 0 sampai 1. Semakin mendekati satu maka akurasi model atau klasifikasi semakin tinggi. Menurut Gorunescu (2011) nilai rentang AUC dan tingkatan klasifikasinya dikategorikan seperti pada Tabel 2.

Tabel 2 Nilai rentang AUC dan tingkatan klasifikasinya

Nilai AUC	Tingkat klasifikasi
0,90 – 1,00	Sangat Baik
0,80 – 0,90	Baik
0,70 – 0,80	Cukup
0,60 – 0,70	Kurang
0,50 – 0,60	Gagal

III METODE

3.1 Data

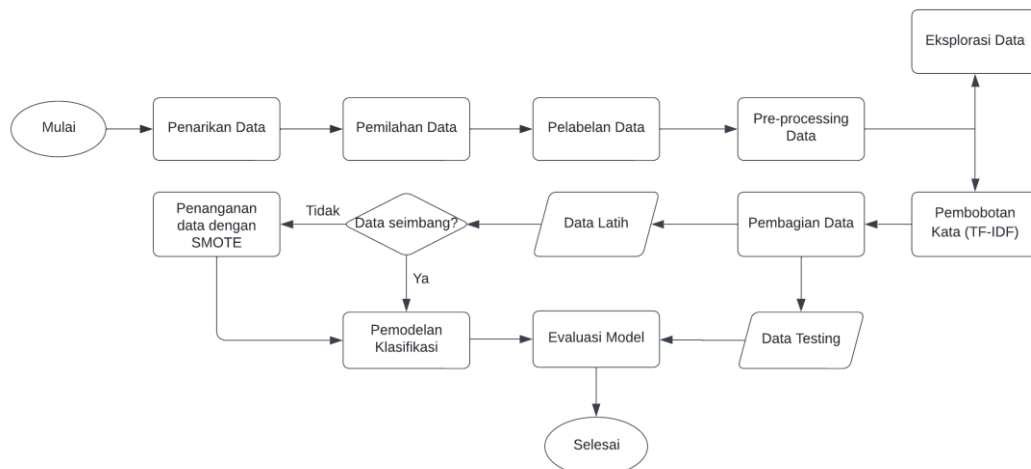
Data yang digunakan dalam penelitian ini adalah data ulasan berbahasa Indonesia pada aplikasi game online Mobile Legends: Bang-Bang. Data merupakan hasil *scrapping* pada website Google Play Store menggunakan bantuan Google Colab. Data ulasan hasil *scrapping* sebanyak 5000 ulasan dengan 2295 ulasan yang digunakan untuk analisis, yaitu pada periode Januari sampai Maret 2022. Alasan dipilihnya periode tersebut adalah data yang digunakan merupakan data terbaru di tahun 2022 sehingga masih selaras dengan pembaruan terakhir aplikasi Mobile Legends: Bang-Bang pada Desember 2021. Peubah yang diambil pada proses penarikan data dijelaskan pada Tabel 3.

Tabel 3 Rincian peubah hasil penarikan data

No	Nama peubah	Keterangan
1	<i>Content</i>	Isi ulasan
2	<i>Username</i>	Nama pengguna pemberi ulasan
3	<i>Score</i>	Rating yang diberikan pengguna
4	<i>At</i>	Waktu pembuatan ulasan

3.2 Prosedur Analisis Data

Prosedur analisis data pada penelitian ini dapat dilihat pada Gambar 4.



Gambar 4 Diagram alir penelitian

Penjelasan prosedur analisis data adalah sebagai berikut:

a. Melakukan penarikan data

Penarikan data dilakukan secara *scrapping* menggunakan bahasa pemrograman Python pada Google Colab dengan mengambil ulasan aplikasi game online Mobile Legends: Bang-Bang di Google Play Store. Data *scrapping* sebanyak 5000 ulasan yang diambil berdasarkan ulasan paling relevan terhadap aplikasi. Contoh data hasil penarikan dapat dilihat pada Tabel 4.

Tabel 4 Contoh hasil penarikan data

<i>Username</i>	<i>Score</i>	<i>At</i>	<i>Content</i>
GUNAWAN FEBRIAN	2	1/1/2022 15:39	Perbaiki di masalah server, servernya sering error dan jaringan sering lag padahal uda pake data sama wifi
Muhammad Danish Fairuzzabadi	3	3/31/202 2 23:25	game nya memuaskan tetapi banyak <i>Player</i> rasis dan menggunakan bahasa yang tidak pantas

b. Melakukan pemilahan data

Data hasil *scrapping* masih terdapat ulasan dengan periode tahun 2018 hingga tahun 2021. Penelitian ini hanya fokus untuk ulasan pada periode Januari sampai Maret 2022 sehingga ulasan selain itu akan dihapus. Selain itu, terdapat ulasan yang bersifat *redundant* dan memberikan informasi yang persis sama sehingga ulasan tersebut hanya diambil salah satu saja.

c. Melakukan pelabelan data

Setiap ulasan yang terdapat dalam data dapat menyampaikan sentimen berbeda terhadap aspek yang berbeda sehingga dilakukan proses pelabelan data. Pelabelan data dilakukan menggunakan data hasil proses pemilahan dengan peubah yang diambil hanya *username* dan *content*. Pelabelan data dilakukan dengan mengkategorikan sentimen setiap ulasan berdasarkan masing-masing aspek. Terdapat empat kategori aspek yang digunakan, yaitu aspek *Gameplay*, *Performa*, Visualisasi, dan *Player*. Setiap ulasan akan diberi label secara manual dengan tiga kriteria pelabelan, yaitu label 1, label -1, dan label 0.

Label 1 dengan sentimen positif menunjukkan ulasan yang mendukung atau mengapresiasi aspek di dalam game. Label -1 dengan sentimen negatif menunjukkan ulasan yang mengandung ejekan atau pernyataan kontra terhadap aspek game. Label 0 apabila ulasan di dalam data tidak mencakup pembahasan mengenai kategori aspek. Ulasan yang tidak mencakup aspek (label 0) tersebut akan dihapus sehingga tersisa ulasan dengan sentimen positif dan negatif yang siap untuk dilakukan analisis. Contoh ulasan yang sudah diberi label terdapat pada Tabel 5.

Tabel 5 Contoh ulasan yang sudah diberi label

<i>Username</i>	<i>Content</i>	<i>Aspek^a</i>			
		A	B	C	D
GUNAWAN FEBRIAN	Perbaiki di masalah server, servernya sering error dan jaringan sering lag padahal uda pake data sama wifi	0	-1	0	0
Muhammad Danish Fairuzzabadi	game nya memuaskan tetapi banyak <i>Player</i> rasis dan menggunakan bahasa yang tidak pantas	1	1	1	-1
Ilham Aja	Gimana aplikasinya kok gak keluar di hp saya bisanya main di playstore	0	0	0	0

^aA : *Gameplay*, B : *Performa*, C : Visualisasi, D : *Player*

d. Melakukan *pre-processing*

Pre-processing data dilakukan menggunakan bahasa pemrograman Python pada Google Colab. Tahapan *pre-processing* yang akan dilakukan pada data ulasan meliputi:

1) Cleaning data

Tahap cleaning data merupakan proses pembersihan teks, menghilangkan elemen-elemen yang tidak diperlukan. Proses ini akan menghapus seluruh karakter lain selain huruf, seperti tanda baca, angka, simbol, tag, html, dan emoji. Penyeragaman bentuk huruf juga dilakukan untuk mengubah kata-kata ulasan menjadi bentuk huruf kecil (*lowercase*). Contoh ulasan yang telah dilakukan tahap cleaning data dapat dilihat di Tabel 6.

Tabel 6 Contoh ulasan setelah cleaning data

Sebelum cleaning data	Setelah cleaning data
Game ini sangat seru!!! Gak nyesel sih kalau main ini karena semua Hero disini uih jago!!!! banget kalo hero kesukaan sy adalah Beatrix, maksih ml sudah membuat game ini♥️❏♥️❏	game ini sangat seru gak nyesel sih kalau main ini karena semua hero disini uih jago banget kalo hero kesukaan sy adalah beatrix maksih ml sudah membuat game ini

2) Normalisasi kata

Tahap normalisasi kata dilakukan untuk menyeragamkan format kata yang memiliki makna sama tetapi dengan penulisan yang berbeda. Proses ini dilakukan dikarenakan masih terdapat kesalahan penulisan dan penggunaan kata yang tidak baku pada ulasan. Normalisasi kata diterapkan menggunakan kamus yang berasal dari penelitian Salsabila *et al.* (2018) yang berjudul “*Colloquial Indonesian Lexicon*”. Kamus tersebut telah dimodifikasi dan ditambahkan secara manual agar sesuai dengan topik penelitian penulis. Total kata tidak baku yang terdapat di dalam kamus sebanyak 5098 kata. Contoh kata sebelum dan setelah dilakukan normalisasi dapat dilihat pada Tabel 7.

Tabel 7 Contoh kata setelah normalisasi

Sebelum normalisasi	Setelah normalisasi
game ini sangat seru gak nyesel sih kalau main ini karena semua hero disini uh jago banget kalo hero kesukaan sy adalah beatrix maksih ml sudah membuat game ini	game ini sangat seru tidak menyesal sih kalau main ini karena semua hero disini uh jago banget kalau hero kesukaan saya adalah beatrix terima kasih mobile legends sudah membuat game ini

3) *Tokenizing*

Tahap *tokenizing* dilakukan proses pemisahan struktur kalimat menjadi potongan kata atau token. Contoh ulasan setelah dilakukan *tokenizing* dapat dilihat pada Tabel 8.

Tabel 8 Contoh ulasan setelah *tokenizing*

Sebelum <i>tokenizing</i>	Setelah <i>tokenizing</i>
game ini sangat seru tidak menyesal sih kalau main ini karena semua hero disini uh jago banget kalau hero kesukaan saya adalah beatrix terima kasih mobile legends sudah membuat game ini	['game', 'ini', 'sangat', 'seru', 'tidak', 'menyesal', 'sih', 'kalau', 'main', 'ini', 'karena', 'semua', 'hero', 'disini', 'uh', 'jago', 'banget', 'kalau', 'hero', 'kesukaan', 'saya', 'adalah', 'beatrix', 'terima', 'kasih', 'mobile', 'legends', 'sudah', 'membuat', 'game', 'ini']

4) *Stopwords removal*

Tahap *stopwords removal* bertujuan untuk menghilangkan kata-kata yang tidak berpengaruh ke dalam proses klasifikasi. Kata-kata tersebut seperti kata penghubung, kata ganti orang, kata seruan, dan kata lainnya yang terdapat dalam daftar *stopwords*. Penelitian ini menggunakan daftar kata dalam Bahasa Indonesia yang didefinisikan sebagai stopwords berasal dari gabungan modul Sastrawi dan Natural Language Toolkit (NLTK) pada bahasa pemrograman Python yang telah dimodifikasi oleh penulis. Contoh ulasan setelah dilakukan *stopwords removal* dapat dilihat pada Tabel 9.

Tabel 9 Contoh ulasan setelah *stopwords removal*

Sebelum <i>stopwords removal</i>	Setelah <i>stopwords removal</i>
['game', 'ini', 'sangat', 'seru', 'tidak', 'menyesal', 'sih', 'kalau', 'main', 'ini', 'karena', 'semua', 'hero', 'disini', 'uh', 'jago', 'banget', 'kalau', 'hero', 'kesukaan', 'saya', 'adalah', 'beatrix', 'terima', 'kasih', 'mobile', 'legends', 'sudah', 'membuat', 'game', 'ini']	['game', 'seru', 'menyesal', 'main', 'hero', 'jago', 'banget', 'hero', 'kesukaan', 'beatrix', 'terima', 'kasih', 'mobile', 'legends', 'game']

5) *Stemming*

Tahap *stemming* bertujuan untuk menguraikan bentuk suatu kata pada teks sehingga didapatkan kata dasarnya. Penelitian ini menggunakan modul Sastrawi sebagai acuan bentuk kata dasarnya. Contoh ulasan setelah dilakukan *stemming* dapat dilihat pada Tabel 10.

Tabel 10 Contoh ulasan setelah *stemming*

Sebelum <i>stemming</i>	Setelah <i>stemming</i>
['game', 'seru', 'menyesal', 'main', 'hero', 'jago', 'banget', 'hero', 'kesukaan', 'beatrix', 'terima', 'kasih', 'mobile', 'legends', 'game']	game seru sesal main hero jago banget hero suka beatrix terima kasih mobile legends game

- e. Melakukan eksplorasi data
- 1) Eksplorasi data keseluruhan
Melihat deskripsi data serta dilakukan polaritas sentimen pada data ulasan secara keseluruhan. Nilai polaritas ditentukan berdasarkan setiap ulasan menggunakan *library TextBlob*.
 - 2) Eksplorasi data masing-masing aspek
Mengelompokkan data hasil pelabelan sesuai dengan masing-masing aspek serta mengkategorikannya sesuai dengan sentimen positif dan negatif.
 - 3) Pembentukan awan kata
Pembentukan awan kata berdasarkan data masing-masing aspek yang sudah melalui tahap *pre-processing*. Awan kata dilakukan pada kategori sentimen positif dan negatif untuk melihat kata yang berkaitan dengan sentimen tersebut. Ukuran kata yang semakin besar menunjukkan semakin besar juga frekuensi munculnya kata di dalam data.
- f. Melakukan pembobotan TF-IDF
Data hasil *pre-processing* dilakukan pembobotan pada masing-masing dataset aspek dengan menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan TF-IDF ini akan dihasilkan vektor atau matriks berdimensi yang menunjukkan baris dokumen dan jumlah kata unik di dalam data. Setiap kata unik yang terdapat di dalam dokumen tersebut memiliki bobot kata sesuai dengan pembobotan TF-IDF. Ilustrasi hasil pembobotan TF-IDF terdapat pada Tabel 11.

Tabel 11 Contoh hasil pembobotan TF-IDF

Dokumen	Kata				
	hero	main	...	game	seru
D1	0	0	...	0.163	0.125
D2	0.125	0	...	0	0
D3	0	0.167	...	0.167	0

- g. Melakukan pembagian data dan penanganan data tidak seimbang
Data masing-masing aspek dibagi menjadi data *training* dan data *testing* sebelum dilakukan proses klasifikasi. Pembagian data menggunakan perbandingan jumlah data *training* dan data *testing* sebesar 80:20. Pemeriksaan ketidakseimbangan pada data dilakukan berdasarkan data *training*. Data yang tidak seimbang akan dilakukan penanganan menggunakan metode SMOTE.
- h. Melakukan pemodelan klasifikasi sentimen menggunakan SVM
Pemodelan klasifikasi dilakukan menggunakan metode SVM. Fungsi yang digunakan adalah fungsi *linear kernel*. Model klasifikasi dibangun menggunakan *package Support Vector Classification* (SVC) dari *library Scikit-Learn* berdasarkan data *training* pada masing-masing aspek. Rincian peubah data yang siap dilakukan pemodelan dapat dilihat pada Tabel 12.

Tabel 12 Peubah data pemodelan SVM

Peubah	Bentuk Data	Keterangan
X	$\vec{x}_i = \{x_1, x_2 \dots x_l\} \in R^d$	<i>Document term matrix</i> dari pembobotan TF-IDF
Y	$y_i \in \{-1, +1\}$	Label sentimen negatif (-1) dan positif (1)

dengan $i = 1, 2, \dots, l$; l adalah banyaknya data, d adalah dimensi data

i. Melakukan evaluasi model

Proses evaluasi dilakukan untuk menghitung *performa* sistem klasifikasi yang telah dibangun. Data *testing* akan digunakan untuk mengevaluasi model klasifikasi. Pengujian model dilihat dari nilai akurasi dan *f1-score* berdasarkan perhitungan dari *confusion matrix* serta melihat dari nilai AUC.

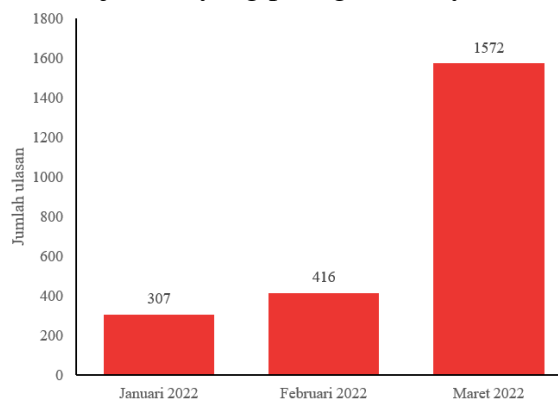
j. Menarik kesimpulan

IV HASIL DAN PEMBAHASAN

4.1 Eksplorasi Data

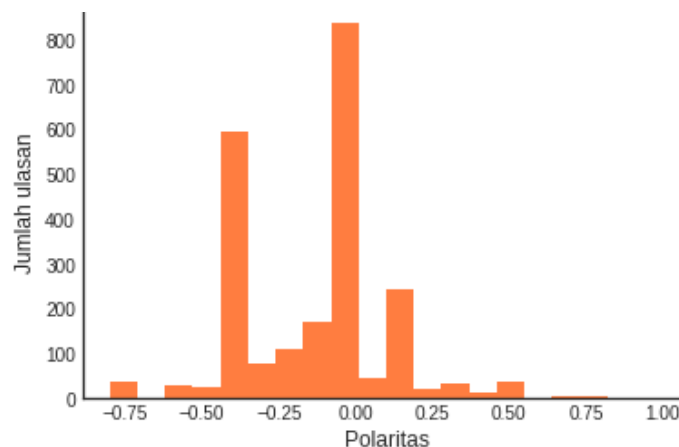
4.1.1 Eksplorasi Keseluruhan Data

Data setelah dilakukan proses pemilahan didapatkan sebanyak 2295 ulasan yang siap dilakukan tahap pelabelan data. Diagram batang digunakan untuk menunjukkan banyaknya ulasan yang digunakan pada periode Januari sampai Maret 2022. Gambar 5 menunjukkan bahwa ulasan yang paling banyak diambil pada saat penarikan data yaitu pada bulan Maret 2022 sebanyak 1572 ulasan. Jumlah ulasan pada bulan Februari 2022 sebanyak 416 ulasan, sedangkan pada bulan Januari memiliki jumlah yang paling sedikit yaitu sebanyak 307 ulasan.



Gambar 5 Diagram batang frekuensi ulasan per bulan

Ulasan yang disampaikan pengguna dapat dilihat seberapa positif atau negatif ulasan tersebut terhadap aplikasi. Nilai polaritas dapat digunakan untuk melihat sentimen ulasan tersebut di dalam data. Hasil nilai polaritas dapat memberikan informasi terkait kelas yang dapat dimasuki oleh setiap ulasan, yaitu $1 \geq \text{nilai polaritas} > 0$ dikategorikan ke dalam sentimen positif, nilai polaritas = 0 dikategorikan netral, dan $0 > \text{nilai polaritas} \geq -1$ dikategorikan ke dalam sentiment negatif (Prasetyo dan Subagyo 2021). Polaritas sentimen pada ulasan keseluruhan dapat dilihat melalui histogram pada Gambar 6.



Gambar 6 Histogram polaritas sentimen keseluruhann ulasan

Hasil histogram polaritas sentimen pada Gambar 6 menunjukkan bahwa mayoritas ulasan memiliki rentang $0 > \text{nilai polaritas} \geq -0.50$. Hal tersebut menandakan bahwa data didominasi oleh ulasan negatif. Jumlah ulasan terbanyak yang menunjukkan sentimen negatif yaitu lebih dari 800 ulasan sedangkan jumlah ulasan terbanyak yang menunjukkan sentimen positif hanya sebanyak 250 ulasan. Ulasan dengan sentimen netral tidak ditemukan di dalam data karena tidak adanya ulasan yang memiliki nilai polaritas 0 di dalam histogram.

4.1.2 Eksplorasi Data masing-masing Aspek

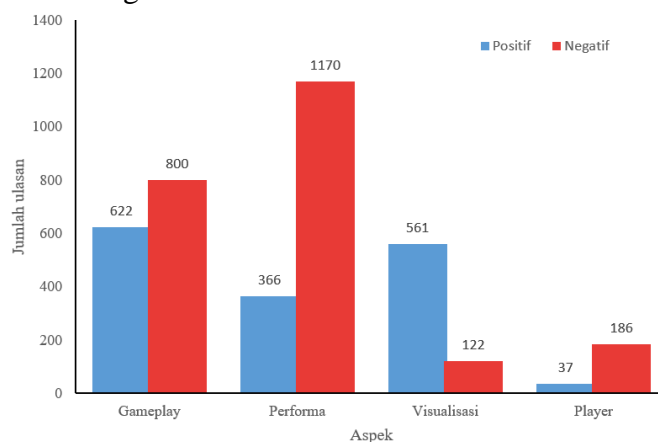
Proses pelabelan setiap ulasan menghasilkan empat dataset berdasarkan masing-masing aspek. Banyaknya data hasil pelabelan dapat dilihat pada Tabel 11.

Tabel 13 Banyaknya data hasil pelabelan

Aspek	Label		
	1	0	-1
<i>Gameplay</i>	622	873	800
<i>Performa</i>	366	759	1170
Visualisasi	561	1612	122
<i>Player</i>	37	2072	186

Label 1 menunjukkan bahwa ulasan memiliki sentimen positif sedangkan label -1 menunjukkan sentimen negatif. Ulasan dengan label 0 akan dihapus karena ulasan tersebut tidak mencakup pembahasan pada setiap aspeknya. Total ulasan yang digunakan untuk analisis pada masing-masing aspek, yaitu sebanyak 1422 ulasan untuk aspek *Gameplay*, 1536 ulasan untuk aspek *Performa*, 683 ulasan untuk aspek Visualisasi, dan 223 ulasan untuk aspek *Player*. Total masing-masing aspek tersebut menunjukkan bahwa data didominasi oleh pengguna yang memberikan ulasan mengenai aspek *Gameplay* dan *Performa* pada aplikasi game Mobile Legends: Bang-Bang.

Sentimen ulasan masing-masing aspek berdasarkan data hasil pelabelan dapat dilihat pada Gambar 7. Sentimen terdiri dari dua kategori, yaitu sentimen positif dan sentimen negatif.



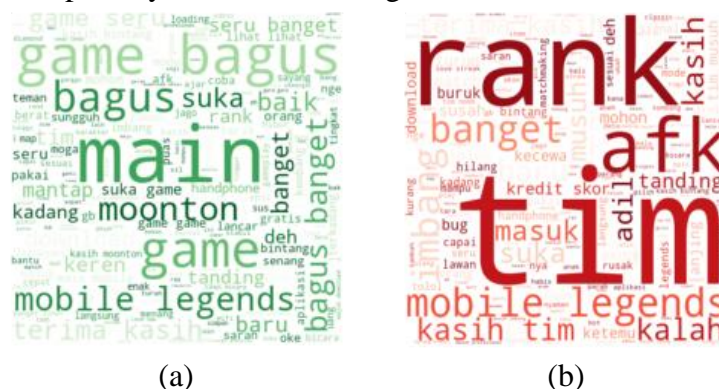
Gambar 7 Banyaknya masing-masing sentimen pada setiap aspek

Aspek *Gameplay*, *Performa*, dan *Player* memiliki mayoritas ulasan dengan sentimen negatif. Banyaknya masing-masing ulasan tersebut berturut-turut sebanyak 800 ulasan (56.3%), 1169 ulasan (76.1%), dan 186 ulasan (83.4%). Hasil sentimen ketiga aspek tersebut menunjukkan bahwa pengguna aplikasi game online Mobile Legends: Bang-Bang cenderung menyatakan kontra atau ejekan terhadap aplikasi dari aspek *Gameplay*, *Performa*, dan *Player*. Ketiga aspek tersebut di dalam aplikasi memiliki urgensi untuk segera diperbaiki dan ditingkatkan oleh pihak pengembang game Montoon. Akan tetapi, pada aspek Visualisasi menunjukkan bahwa mayoritas ulasan memiliki sentimen positif yaitu sebanyak 561 ulasan (82.1%). Hal tersebut menunjukkan bahwa pengguna cenderung menyukai atau mengapresiasi game dari aspek Visualisasinya.

4.1.3 Pembentukan Awan Kata

Pembentukan awan kata dilakukan agar dapat memberikan gambaran yang lebih jelas mengenai kata-kata yang sering muncul di dalam data. Semakin besar kata yang terlihat menunjukkan bahwa kata tersebut semakin sering muncul atau semakin tinggi frekuensi kata tersebut muncul di dalam data. Pembentukan awan kata dilakukan pada masing-masing sentimen aspek.

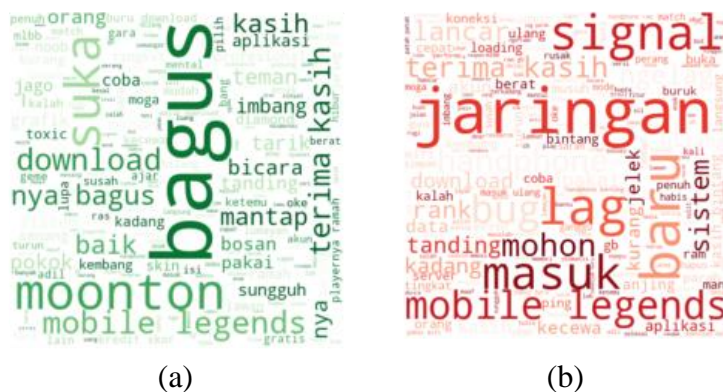
Awan kata yang dihasilkan dari data aspek *Gameplay* dapat dilihat pada Gambar 8. Hasil awan kata pada sentimen positif (a) menunjukkan bahwa dominasi kata yang sering muncul adalah kata “main”, “game”, dan “bagus”. Kata-kata tersebut di dalam aspek *Gameplay* dapat memberikan gambaran bahwa pengguna suka bermain game Mobile Legends: Bang-Bang. Awan kata pada sentimen negatif (b) menunjukkan kata “rank”, “tim”, dan “afk” memiliki frekuensi kemunculan terbanyak di dalam ulasan. Kata “afk” menunjukkan bahwa gamer meninggalkan meja bermain ketika sedang online, hal tersebut umumnya mengganggu jalannya permainan. Kata “rank” umumnya menunjukkan mode tingkatan pangkat atau level yang bisa dimainkan berdasarkan jumlah jam terbang. Kata “tim” dapat diartikan sebagian besar pengguna merasa tidak nyaman terhadap timnya ketika bermain game.



Gambar 8 Awan kata untuk data aspek *Gameplay* sentimen positif (a) dan sentimen negatif (b)

Awan kata yang dihasilkan dari data aspek *Performa* dapat dilihat pada Gambar 9. Hasil awan kata pada sentimen positif (a) menunjukkan bahwa dominasi kata yang sering muncul adalah kata “bagus”, “suka”, dan “moonton”.

Kata-kata tersebut dapat memberikan gambaran bahwa pengguna menyukai dan menyatakan kinerja dari moonton sebagai pengembang game sudah bagus. Awan kata pada sentimen negatif (b) menunjukkan kata “signal”, “jaringan”, dan “lag” memiliki frekuensi kemunculan terbanyak di dalam ulasan. Kata “signal” dan “jaringan” tersebut dapat memberikan gambaran bahwa pengguna terkendala terhadap signal atau jaringan saat memainkan game. Kata “lag” dapat memberikan gambaran bahwa sistem game seringkali melambat ketika permainan dimulai.



Gambar 9 Awan kata untuk data aspek *Performa* sentimen positif (a) dan sentimen negatif (b)

Awan kata yang dihasilkan dari data aspek Visualisasi dapat dilihat pada Gambar 10. Hasil awan kata pada sentimen positif (a) menunjukkan bahwa dominasi kata yang sering muncul adalah kata “bagus”, “suka”, dan “hero”. Kata-kata tersebut dapat memberikan gambaran bahwa pengguna menyukai hero atau karakter yang terdapat dalam game Mobile Legends: Bang-Bang dan menyatakan hero tersebut sudah bagus. Awan kata pada sentimen negatif (b) menunjukkan kata “pakai”, “skin”, dan “baru” memiliki frekuensi kemunculan terbanyak di dalam ulasan. Kata “skin” menunjukkan fitur untuk merubah tampilan visual pada suatu hero di dalam game. Terlihat juga kata “jelek” dan “mohon” pada awan kata. Kata tersebut dapat memberikan gambaran bahwa pengguna merasa skin yang baru diperoleh dari hasil pembaruan memiliki tampilan yang jelek sehingga dapat menjadi saran untuk segera diperbaiki.



Gambar 10 Awan kata untuk data aspek Visualisasi sentimen positif (a) dan sentimen negatif (b)

Tabel 14 Hasil pembagian data *training* dan data *testing* per aspek

Aspek	Sentimen	80% data <i>training</i>	Total data <i>training</i>	20% data <i>testing</i>	Total data <i>testing</i>
<i>Gameplay</i>	Positif	508	1137	114	285
	Negatif	629		171	
<i>Performa</i>	Positif	298	1228	68	308
	Negatif	930		240	
Visualisasi	Positif	448	546	113	137
	Negatif	98		24	
<i>Player</i>	Positif	29	178	37	45
	Negatif	149		8	

Pemeriksaan ada dan tidaknya ketidakseimbangan pada data perlu dilakukan karena jika dibiarkan kondisi tersebut nantinya akan berpengaruh pada hasil klasifikasi (Amelia *et al.* 2018). Ketidakseimbangan pada data akan ditangani menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Metode SMOTE digunakan untuk melakukan oversampling pada kelas minoritas. Pembangkitan data kelas minoritas ini dilakukan pada data *training*. Hasil Tabel 12 menunjukkan kategori data *training* yang menjadi data minor adalah data dengan sentimen positif pada aspek *Performa* dan *Player* sedangkan data dengan sentimen negatif terdapat pada aspek Visualisasi. Data pada aspek *Gameplay* tidak dicurigai sebagai data tidak seimbang karena perbedaan jumlah ulasan pada kedua kelas sentimen tidak cukup besar. Data dikatakan tidak seimbang apabila suatu kelas data memiliki jumlah amatan yang jauh lebih sedikit dibandingkan kelas lainnya (Bunkhumpornpat *et al.* 2012).

Data hasil pembangkitan melalui SMOTE menghasilkan data pada kelas minoritas menjadi sama banyak dengan kelas mayoritas dengan proporsi akhir 50% pada masing-masing aspek. Presentase data sebelum dan sesudah dilakukan SMOTE dapat dilihat pada Tabel 13.

Tabel 15 Presentase data sebelum dan sesudah dilakukan SMOTE

Aspek	Kelas	80% data <i>training</i>	80% data <i>training</i> (SMOTE)	Data akhir (SMOTE)
<i>Performa</i>	Positif	24.3%	50%	930
	Negatif	75.7%	50%	930
Visualisasi	Positif	82.1%	50%	448
	Negatif	17.9%	50%	448
<i>Player</i>	Positif	16.3%	50%	149
	Negatif	83.7%	50%	149

4.4 Pemodelan Klasifikasi

Pemodelan klasifikasi pada penelitian ini menggunakan metode SVM dengan fungsi *linear kernel*. Parameter yang digunakan dalam fungsi *linear kernel* adalah parameter C dengan nilai $C = 1$. Pemodelan dilakukan pada masing-masing aspek menggunakan data *training* setelah dilakukan proses SMOTE. Akan tetapi, pada aspek *Gameplay*, data yang digunakan untuk pemodelan klasifikasi merupakan data *training* tanpa SMOTE.

Klasifikasi sentimen juga dilakukan dengan menyesuaikan model yang telah terbentuk sebelumnya yang selanjutnya digunakan fungsi *predict* untuk memprediksi label kelas sentimen. Setiap algoritma klasifikasi memiliki kesalahan dalam melakukan klasifikasi maupun prediksi. Nilai prediksi model tersebut dapat diketahui kesalahannya dalam klasifikasi melalui hasil *confusion matrix*. Data yang digunakan dalam proses ini merupakan data *testing* dari masing-masing aspek.

4.4.1 Aspek *Gameplay*

Prediksi klasifikasi aspek *Gameplay* menggunakan data *testing* sebanyak 285 ulasan. Hasil prediksi klasifikasi sentimen pada aspek *Gameplay* terlihat pada *confusion matrix* yang terdapat dalam Tabel 16. Berdasarkan nilai *confusion matrix* pada aspek *Gameplay* didapatkan bahwa banyak data yang benar terklasifikasi sebagai ulasan positif sebanyak 84 ulasan. Data yang benar terklasifikasi sebagai ulasan negatif sebanyak 154 ulasan.

Tabel 16 *Confusion matrix* aspek *Gameplay*

Kelas aktual	Kelas prediksi	
	Positif	Negatif
Positif	84	30
Negatif	17	154

4.4.2 Aspek *Performa*

Prediksi klasifikasi aspek *Performa* menggunakan data *testing* sebanyak 308 ulasan. Hasil prediksi klasifikasi sentimen pada aspek *Gameplay* terlihat pada *confusion matrix* yang terdapat dalam Tabel 17. Berdasarkan nilai *confusion matrix* pada aspek *Performa* didapatkan bahwa banyak data yang benar terklasifikasi sebagai ulasan positif sebanyak 53 ulasan. Data yang benar terklasifikasi sebagai ulasan negatif sebanyak 230 ulasan.

Tabel 17 *Confusion matrix* aspek *Performa*

Kelas aktual	Kelas prediksi	
	Positif	Negatif
Positif	53	15
Negatif	10	230

4.4.3 Aspek Visualisasi

Prediksi klasifikasi aspek Visualisasi menggunakan data *testing* sebanyak 137 ulasan. Hasil prediksi klasifikasi sentimen pada aspek *Gameplay* terlihat pada *confusion matrix* yang terdapat dalam Tabel 18. Berdasarkan nilai *confusion matrix* pada aspek Visualisasi didapatkan bahwa banyak data yang

benar terklasifikasi sebagai ulasan positif sebanyak 108 ulasan. Data yang benar terklasifikasi sebagai ulasan negatif sebanyak 11 ulasan.

Tabel 18 *Confusion matrix* aspek Visualisasi

Kelas aktual	Kelas prediksi	
	Positif	Negatif
Positif	108	5
Negatif	13	11

4.4.4 Aspek Player

Prediksi klasifikasi aspek *Player* menggunakan data *testing* sebanyak 45 ulasan. Hasil prediksi klasifikasi sentimen pada aspek *Gameplay* terlihat pada *confusion matrix* yang terdapat dalam Tabel 19. Berdasarkan nilai *confusion matrix* pada aspek *Player* didapatkan bahwa banyak data yang benar terklasifikasi sebagai ulasan positif sebanyak 6 ulasan. Data yang benar terklasifikasi sebagai ulasan negatif sebanyak 37 ulasan.

Tabel 19 *Confusion matrix* aspek Player

Kelas aktual	Kelas prediksi	
	Positif	Negatif
Positif	6	2
Negatif	0	37

4.5 Evaluasi Model

Proses evaluasi dilakukan untuk mengetahui evaluasi kinerja dari metode SVM dalam mengklasifikasikan sentimen pada masing-masing aspeknya. Hasil performansi klasifikasi menggunakan metode SVM dapat dilihat pada Tabel 20.

Tabel 20 Hasil performansi klasifikasi SVM

Aspek	Kelas	<i>F1-score</i>	<i>Macro average F1-score</i>	Akurasi
<i>Gameplay</i>	Positif	0,78	82%	84%
	Negatif	0,87		
<i>Performa</i>	Positif	0,81	88%	92%
	Negatif	0,95		
Visualisasi	Positif	0,92	74%	87%
	Negatif	0,55		
<i>Player</i>	Positif	0,86	92%	96%
	Negatif	0,97		

Hasil performansi tersebut diperoleh bahwa nilai akurasi masing-masing aspek di atas 80%. Akurasi menunjukkan ketepatan model dan fungsi yang digunakan terhadap klasifikasi yang dihasilkan dari data ulasan. Model mampu mengklasifikasi data uji dengan tepat sebesar 84% pada aspek *Gameplay*, sebesar 92% pada aspek *Performa*, sebesar 87% pada aspek Visualisasi, dan sebesar 96% pada aspek *Player*.

Nilai *F1-score* dapat dilihat pada masing-masing kelas sentimen setiap aspek sehingga untuk mengetahui nilai *F1-score* secara keseluruhan dapat dilihat pada nilai *macro average F1-score*. Aspek Visualisasi memiliki nilai *macro average F1-score* yang paling rendah dibandingkan dengan ketiga aspek lainnya, yaitu sebesar 74%. Sementara itu aspek *Gameplay*, *Performa*, dan *Player* memiliki nilai *macro average F1-score* di atas 80%. Secara keseluruhan hasil performansi tersebut menunjukkan bahwa pemodelan klasifikasi menggunakan SVM pada masing-masing aspek yang telah dilakukan sudah cukup baik.

Evaluasi model juga dilakukan dengan melihat nilai AUC. Hasil nilai AUC masing-masing aspek dapat dilihat pada Tabel 21.

Tabel 21 Nilai AUC masing-masing aspek

Aspek	Nilai AUC
<i>Gameplay</i>	0,819
<i>Performa</i>	0,869
Visualisasi	0,707
<i>Player</i>	0,875

Aspek *Gameplay* memiliki nilai AUC sebesar 0,819. Aspek *Performa* memiliki nilai AUC sebesar 0,869. Aspek Visualisasi memiliki nilai AUC sebesar 0,707. Aspek *Player* memiliki nilai AUC sebesar 0,875. Menurut Gorunescu (2011), aspek *Gameplay*, *Performa*, dan *Player* memiliki nilai AUC diatas 0,80 yang menunjukkan bahwa ketiga aspek tersebut memiliki tingkat klasifikasi baik. Sementara itu, nilai AUC dari aspek Visualisasi yaitu kurang dari 0,80 yang menunjukkan bahwa tingkat klasifikasi cukup baik.

V SIMPULAN DAN SARAN

5.1 Simpulan

Analisis sentimen berbasis aspek dilakukan pada data ulasan aplikasi game online Mobile Legends: Bang-Bang periode Januari 2022 sampai Maret 2022. Kategori aspek yang digunakan dalam analisis adalah aspek *Gameplay*, *Performa*, *Visualisasi*, dan *Player*. Aplikasi game online Mobile Legends: Bang-Bang cenderung memiliki sentimen negatif dari aspek *Gameplay*, *Performa*, dan *Player*. Ketiga aspek tersebut menjadi urgensi untuk segera diperbaiki dan ditingkatkan kembali oleh pihak Moonton sebagai pengembang game. Sementara itu dari aspek *Visualisasi* cenderung memiliki sentimen positif. Pengguna banyak memberikan ulasan positif mengenai ketertarikannya terhadap hero atau karakter yang terdapat di dalam game sehingga hal tersebut perlu dipertahankan. Algoritma SVM digunakan sebagai metode klasifikasi. Performansi hasil klasifikasi menggunakan SVM diukur berdasarkan nilai akurasi, *F1-score*, dan nilai AUC yang dihasilkan model. Aspek *Gameplay*, *Performa*, dan *Player* memiliki tingkat klasifikasi yang lebih baik dibandingkan aspek *Visualisasi*. Sementara itu, secara umum algoritma SVM menghasilkan performa yang cukup baik pada klasifikasi masing-masing aspek.

5.2 Saran

Saran yang ingin disampaikan yaitu dapat dicoba menggunakan data ulasan berbahasa Inggris dengan pengguna game berasal dari negara lain sehingga dapat ditemukan informasi yang menarik sebagai perbandingan dengan pengguna Indonesia. Pelabelan sentimen secara manual sebaiknya dilakukan oleh para ahli agar hasil klasifikasi dapat lebih akurat.

DAFTAR PUSTAKA

- Amelia OD, Soleh AM, Rahardianto S. 2018. Pemodelan support vector machine data tidak seimbang keberhasilan studi mahasiswa magister IPB. *Xplore*. 2(1):33-40.
- Bunkhumpornpat C, K Sinapiromsaran, C Lursinsap. 2012. DbSMOTE: density-based synthetic minority over-sampling technique. *Application Intelligence*. 36:664-684. doi: 10.1007/s10489-011-0287-y.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 16(2002):321-357. doi: doi.org/10.29244/xplore.v2i1.76.
- Christianini N, Shawe-Taylor J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge (UK): Cambridge University Press.
- Hutami WP. 2021. Penerapan support vector machine dengan SMOTE untuk klasifikasi sentimen pemberitaan omnibus law pada situs cnnindonesia.com [skripsi]. Bogor: IPB University.
- Ilmawan BL, Mude MA. 2020. Perbandingan metode klasifikasi support vector machine dan naïve bayes untuk analisis sentimen pada ulasan tekstual di google play store. *ILKOM Jurnal Ilmiah*. 12(2):154-161.
- Iskandar JW, Nataliani Y. 2021. Perbandingan Naïve Bayes, SVM, dan k-NN untuk analisis sentimen *gadget* berbasis aspek. *Jurnal Resti*. 5(6):1120-1126. doi: 10.29207/resti.v5i6.3588.
- Istiadi, Rahman AY. 2020. Optimasi parameter *support vector machine* berbasis algoritma genetika pada klasifikasi teks pengaduan masyarakat. Seminar Nasional Hasil Riset CIASTECH 2020; 2020 Desember 2; Malang, Indonesia. hlm:481-488.
- Joachims T. 1997. *Probabilistic analysis of the rocchio algorithm with TFIDF for text categorization*. Tennessee (US): Morgan Kaufmann Publishers.
- Lebho MA, Lerik MDC, Wijaya RPC, Littik SKA. 2020. Perilaku kecanduan game online ditinjau dari kesepian dan kebutuhan berafiliasi pada remaja. *Journal of Health and Behavioral Science*. 2(2):202-212.
- Liu B. 2012. *Sentimen analysis and opinion mining*. Morgan & Claypool Publisher.
- Prasetyo E. 2014. *Data mining mengolah data menjadi informasi menggunakan matlab*. Yogyakarta (ID): Andi Publisher.
- Prasetyo BA, Subagyo. 2021. Analisis sentimen pengguna twitter untuk teks berbahasa Indonesia terhadap penyedia layanan home fix broadband. Seminar Nasional Teknik Industri Universitas Gadjah Mada; 2021 September 23; Yogyakarta, Indonesia. hlm:18-23.
- Septian JA, Fahrudin TM, Nugroho A. 2019. Analisis sentimen pengguna twitter terhadap polemik persepakbolaan indonesia menggunakan pembobotan TF-IDF dan *k-nearest neighbor*. *Journal of Intelligent Systems and Computation*. 1(1):43-49.
- Simorangkir H, Lhaksana KM. 2018. Analisis sentimen pada twitter untuk game online mobile legends dan arena of valor dengan metode naïve bayes classifier. *e-Proceeding of Engineering*. 5(3):8131-8140.

- Umma NF, Warsito B, Maruddin DAI. 2021. Klasifikasi status kemiskinan rumah tangga dengan algoritma C5.0 di kabupaten pematang. *Jurnal Gaussian*. 10(2):221-229.
- Yogatama IKS, Kharisma AP, Fanani L. 2019. Analisis faktor-faktor yang memengaruhi minat pemain dalam permainan MOBA (studi kasus: mobile legends: bang-bang!). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 3(3):2558-2566.

RIWAYAT HIDUP

Penulis dilahirkan di Banyumas pada tanggal 9 Maret 2000. Penulis merupakan anak pertama dari dua bersaudara dari pasangan Bapak Sudarmo dan Ibu Siti Toyibah. Tahun 2018, penulis menyelesaikan pendidikan menengah atas di SMAN 1 Purwokerto. Pada tahun yang sama penulis diterima sebagai mahasiswa program sarjana (S-1) di Institut Pertanian Bogor (IPB University) Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA).

Selama mengikuti program S-1, penulis pernah berpartisipasi pada lomba Jambore Statistika IX Universitas Mulawarman tahun 2020 dan meraih Juara Favorit Infografis. Dalam kepanitiaan, penulis berpartisipasi sebagai staf Divisi Kestari Statistika Ria ke-14 pada tahun 2019, tim Design dan Multimedia Satria Data pada tahun 2020, dan kepala divisi Multimedia dan Publikasi Tanoto Student Research Awards (TSRA) National Competition pada tahun 2020. Dalam organisasi, penulis bergabung sebagai anggota Biro Media dan Informasi Gamma Sigma Beta IPB periode 2019/2020. Penulis juga berkesempatan melakukan praktik lapang di Lembaga Administrasi Negara tahun 2021. Saat ini, penulis aktif menjadi bagian dari Komunitas Pendidikan Simpul Angan.