

PERBANDINGAN METODE NAÏVE BAYES DAN SVM PADA ULASAN GOOGLE PLAYSTORE MOBILE LEGENDS: BANG BANG

Patrick Gracezando Yehova ¹⁾, Jessica ²⁾, Muhammad Ihsan Jambak ³⁾

^{1,2,3)} Universitas Sriwijaya

Email : 09031282126036@student.unsri.ac.id ¹⁾, 09031282126074@student.unsri.ac.id ²⁾, jambak@unsri.ac.id ³⁾

ABSTRAK

Perkembangan internet di Indonesia meningkatkan minat bermain *game* terbukti dari pertumbuhan pasar *game* hingga mencapai Rp25 triliun pada 2022. Penerapan analisis sentimen pada ulasan pengguna penting untuk membantu bisnis mengambil tindakan yang tepat guna meningkatkan produk, layanan, serta strategi bisnis. Penelitian ini bertujuan membandingkan kinerja melalui tingkat akurasi metode *Support Vector Machine* (SVM) dan *Naïve Bayes*, dalam analisis sentimen pengguna *game Mobile Legends: Bang Bang* pada ulasan Google Playstore. Dari total 796 dataset, sentimen negatif mencapai 73%, sedangkan sentimen positif hanya mencapai 27%. Hal ini menunjukkan ketidakseimbangan data, dengan lebih banyak sentimen negatif. Teknik *upsampling* SMOTE meningkatkan akurasi *Naïve Bayes* (82%) dan SVM (91%). Perbandingan ini menunjukkan kecenderungan yang jelas bahwa algoritma SVM memberikan kinerja yang lebih baik dalam mengklasifikasikan sentimen dari data yang telah dikumpulkan dan diujikan. Berdasarkan kurva ROC *model* klasifikasi *Naïve Bayes* lebih baik dibandingkan SVM dengan nilai AUROC sebesar 0,070.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*, Ulasan, Perbandingan

ABSTRACT

The rise of the internet in Indonesia has led to more people being interested in gaming. This is shown by the gaming market growing to 25 trillion rupiah in 2022. Using sentiment analysis on user reviews is important for businesses to improve their products, services, and strategies. This study wants to compare how well Support Vector Machine (SVM) and Naïve Bayes methods can analyze user feelings about the game Mobile Legends: Bang Bang using Google Playstore reviews. Out of 796 reviews, 73% were negative and only 27% were positive. This uneven data suggests that there are more negative feelings overall. Using the SMOTE technique made Naïve Bayes 82% accurate and SVM 91% accurate. This comparison suggests that the SVM method is better at figuring out feelings from the data we collected and tested. Based on the ROC curve, the Naïve Bayes classification model is better than SVM with an AUROC value of 0.070.

Keywords: Sentiment Analysis, *Naïve Bayes*, *Support Vector Machine*, Review, Comparison

1. PENDAHULUAN

Perkembangan internet di Indonesia sendiri memiliki dampak yang signifikan terhadap minat bermain *game* yang terus meningkat di masyarakat. Hal ini terbukti melalui pernyataan yang diberikan oleh Direktur Aplikasi Permainan, Televisi, dan Radio Kementerian Pariwisata dan Ekonomi Kreatif, Iman Santosa melalui Antaranews bahwa pasar *game* Indonesia pada tahun 2022 dapat mencapai Rp25 triliun (Purnama, 2023). Perkembangan pasar *game* yang besar ini membuktikan bahwa masyarakat Indonesia semakin banyak menghabiskan waktunya untuk bermain *game*. Pada tahun 2022, *We Are Social* melakukan survei yang membuktikan bahwa 94,5% dari total penduduk Indonesia pernah bermain *game* (*We Are Social*, 2022). Melalui survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada 2023 membuktikan bahwa 3.593 dari 8.510 atau sekitar 42,23% pemain *game* online di Indonesia dapat bermain *game* lebih dari 4 jam (Asosiasi Penyelenggara Jasa Internet (APJII), 2023).

Seiring dengan pertumbuhan *market game* yang kian pesat, tentu akan ada pengguna yang membagikan pengalaman mereka melalui ulasan. Ulasan ini berisi opini pengguna mengenai aplikasi, dalam hal ini *game* yang telah dimainkannya, yang biasanya diunggah pada *platform* tempat pengguna mengunduh aplikasi. Ulasan dari pengguna memiliki peran yang sangat penting dalam penerapan analisis sentimen dimana hal tersebut dapat membantu bisnis atau organisasi untuk mengambil tindakan yang tepat guna meningkatkan produk, layanan, serta strategi bisnis (Drus and Khalid, 2019). Analisis sentimen adalah sub bidang penambangan teks yang berfokus pada pengklasifikasian dan penggalan opini dan sentimen dari data teks (Darwis et al., 2021). Analisis sentimen memberikan kemudahan bagi pemilik bisnis untuk mengetahui popularitas mereka dan pendapat pengguna terkait produk yang ditawarkan (El Rahman et al., 2019). Selain itu, ulasan pengguna juga penting untuk mengetahui kelemahan dari produk yang ditawarkan. Studi kasus mengenai *Big Six* (pemasok gas dan listrik

terbesar dan tertua di Britania Raya serta konsumen energi baru) membandingkan data sentimen pada *tweet* konsumennya (Ikoro et al., 2018). Hasil yang didapat menunjukkan sentimen *Big Six* yang negatif dibanding konsumen energi yang baru bergabung. Opini dalam analisis sentimen bersifat subjektif, dimana nilai dari sebuah kata tidak konkret (Wati and Ernawati, 2021).

Penelitian ini berfokus kepada perbandingan antara metode *Support Vector Machine* (SVM) dan *Naïve Bayes*. Penelitian sebelumnya menyoroti perbandingan metode *Naïve Bayes* dan SVM pada analisis sentimen tentang Perppu Cipta Kerja melalui *Twitter* atau yang sekarang dikenal sebagai X (Farhan and Setiaji, 2023a). Penelitian tersebut menjelaskan bahwa akurasi analisis sentimen dengan metode SVM memiliki akurasi yang lebih tinggi, sebesar 78%. Penelitian tersebut juga menemukan bahwa sentimen negatif lebih banyak ketimbang sentimen positif, menandakan bahwa banyak masyarakat yang kurang setuju dengan topik tersebut. Kekurangan penelitian ini adalah data yang masih minim sehingga akurasi menjadi kurang maksimal, pembagian data sentimen juga tidak seimbang, sehingga data training dan testing tidak memberikan hasil yang maksimal.

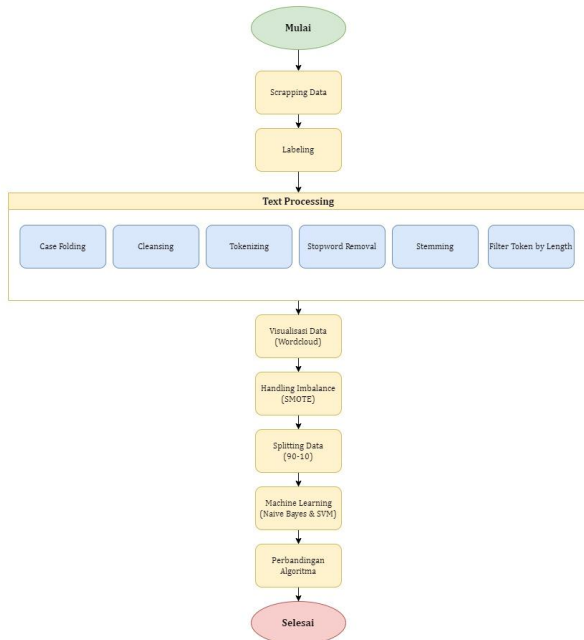
Penelitian yang dilakukan oleh (Atmajaya et al., 2023a) menggunakan metode SVM dan *Naïve Bayes* untuk menentukan sentimen *ChatGPT* pada *Twitter*. Dalam penelitian tersebut menggunakan metode *Naïve Bayes* dan SVM lalu dibandingkan dengan label sentimennya yaitu, *Vader* dan *RoBERTa*. Hasil penelitian tersebut menunjukkan bahwa metode SVM dengan pelabelan sentimen *Vader* memiliki akurasi yang terbaik, sebesar 59% dari hasil data testing 30% dan training 70%. Kelemahan pada penelitian ini adalah tidak memberikan penjelasan mengenai *imbalance* data sehingga tidak dapat menentukan secara maksimal apakah data tersebut benar-benar seimbang.

Pada penelitian ini, penulis akan meningkatkan keseimbangan data sentimen dengan menggunakan menggunakan teknik *Synthetic Minority Oversampling Technique*

(SMOTE) untuk mendapatkan hasil akurasi yang lebih baik.

2. METODE

Metode penelitian mencakup proses lengkap yang dilakukan untuk menganalisis sentimen dari ulasan pengguna *game Mobile Legends: Bang Bang* pada *Google Playstore*. Penggunaan algoritma *Naïve Bayes* dan *Support Vector Machine* pada penelitian bertujuan untuk membandingkan tingkat akurasi dari performa masing-masing algoritma. Dalam proses pengolahan data hingga mendapatkan akurasi digunakan bahasa pemrograman *Python* melalui *platform Google Colab*. Berikut alur yang diimplementasikan dalam penelitian ini.



Gambar 1. Tahapan Metode Penelitian

1. Scrapping Data

Tahap awal dalam penelitian ini adalah mengumpulkan data yang diperlukan dari ulasan pengguna *game* dari *Playstore*. Dataset dikumpulkan menggunakan *Google-play-scraper* (Latif et al., 2019). Data yang terkumpul dari pengambilan data yaitu sebanyak 796 *dataset* dari ulasan pengguna pada tanggal 14 Maret 2024 hingga 25 Maret 2024. *Dataset* tersebut memiliki 11 atribut sesuai pada Gambar 2.

reviewId	userName	userImage	content	score	timestamp
40a2c2c5-8650-4811-a405-0e1115e12592	Khatran Cs	https://play-https://play-	Hero dilnya sh udah bagus tapi semua kebulat	3	6427
528821d7-658c-45a8-ba85-55912d5e1c72	Imones	https://play-https://play-	Semuaanya udah bagus, update-nya juga beda.	5	7880
80a3478c-0f8d-462f-a235-528a922a4853	Nirwana Herlans	https://play-https://play-	Perbaiki masalah jaringan, Selesai maua portia.	1	1218
728508bc-e5ff-4a2d-81e7-e4864c20858	Delta 123	https://play-https://play-	Hero baru cepi Baet menggendakap map menpa.	2	106
2d5c7873-1114-4544-99a1-4064b114789e	Muhammad Abdul Holik Hs	https://play-https://play-	Tolong ni untuk developnya gamenya di petta.	1	670

Gambar 2. Contoh Hasil *Scrapping* Data

2. Pelabelan Data

Pada tahap kedua ini data ulasan *game* akan dikategorikan atau dilabel menjadi dua kategori yakni positif dan negatif melalui atribut *score*. Jika *score* memiliki nilai diatas atau sama dengan 3 maka ulasan diberi label positif (1) dan untuk *score* di bawah 3 maka label negatif (0). Kategori positif sendiri digunakan untuk menggambarkan emosi bahagia, ceria, senang, rileks, yang secara garis besar berhubungan dengan kebajikan. Sedangkan kategori negatif menggambarkan emosi marah, takut, tertekan, serta emosi lain yang menyebabkan penderitaan (Ferdiana et al., 2019).

Tabel 1. Contoh Pelabelan Data

Content	Sentiment
"BANYAK JUNGLER TROLL NGGA DIBASMI! MOONTON TOLONG LAH, MASA MINTA TOLONG BAIK BAIK GA DIGUBRIS, GUA UDAH MAKI MAKI PUN NGGA BERUBAH? PLAYER TIM SENDIRI SELALU DAPET TROL EGOIS! PICK GA JELAS! mending kalo bisa main, udah ga jelas MAIN KEK AYAM"	0
"Asalamualaikum kepada yth moonton, saya ingin nanya anda niat bikin game atau niat rusuh,terus matchmaking nya lama lama, akhirnya pas main musuhnya jago, timnya beban,and ini tidak adil kepada saya,terus saya pas pake hero mmr musuh nya pada keren,jago,sedangkan tim saya bot,beban,dan pas pake hero yg lain bukan mmr,tim bot musuh bot...tolong perbaikiin..dpt tim dark system terus unistal aja lah ML mls saya di kasih menang 3x,di kasih kalah 9x pintar game ML luar biasa"	0
"Developer tolong lah kalok ngasih tim yang sesuai bisa ga sih. Yang jago lawan yang jago yang lemah lawan yang lemah, Gitu Napa sih bukan malah ngasih tim beban atau di campur sama yang boot buat di gendong satu orang . Sekali lagi tolong lah kasih tim yang tau role Hero, item dan yang lainnya, lama lama gue males main ni game kalau belum di benerin nih permasalahan"	0
"Game nya bagus tapi muali dari dark sistemnya buruk sekali ,jaringannya,dipilihkan tim fight yang selalu TOLOL,saat main nge bug, saya main game ini untuk menghibur diri tapi sistemnya sekali lagi bukan manusia yang buat tapi kaya apa yaaaaaa...kalau mau update itu bagus tapi bukan dark sistem kamu yang juga kau update MOONTON.susah karena orang-orang MOONTON tambah senang makanya saya bilang bukan manusia."	0

"Makin ksini makin parah, bug nya sangat mengganggu saat tempur, padahal jaringan wifi lancar aja, belum lagi dapat musuh kelewat sakti, jadi makin susah menangnya, benar² terasa ketidak Adilan.."

3. Preprocessing Data

Tahap *preprocessing* yakni dilakukan pembersihan data sebelum dilakukan proses selanjutnya. Data ulasan *game* akan melewati tahap *preprocessing* untuk menghilangkan karakter-karakter yang tidak relevan (Atmajaya et al., 2023b). Langkah awal *preprocessing* ini diperlukan sebelum menganalisis data, karena tahap ini bertujuan mengubah data dari yang tidak terstruktur menjadi terstruktur sehingga mempermudah proses analisis dan menghindari kesalahan dalam analisis (Muflih et al., 2023; Wijaya et al., 2024). Metode yang digunakan dalam *preprocessing* data yaitu *Case Folding*, *Cleansing*, *Tokenizing*, *Stopwords Removal*, *Stemming*, dan *Filter Token by Length*.

a. Case Folding

Proses ini dilakukan dengan tujuan untuk mempermudah proses pencarian. Pada dasarnya, tidak semua penggunaan huruf kapital dalam teks konsisten digunakan (Alita and Isnain, 2020). Maka, metode klasifikasi *case folding* berperan dalam memproses kata-kata dalam teks kembali dalam bentuk standar (huruf kecil) (Farhan and Setiaji, 2023b).

b. Cleansing

Cleansing merupakan tahap yang bertujuan untuk menghilangkan karakter seperti simbol, angka, tanda baca.

c. Tokenizing

Tahap *tokenizing* bermaksud untuk memisahkan setiap kata yang ada pada dokumen. Proses ini memanfaatkan *library Python* berupa *nltk* yaitu *word_tokenize* sehingga didapat hasil pada tabel 2.

```
Tokenizing

from nltk.tokenize import word_tokenize

from nltk.tokenize import RegexpTokenizer
regexp = RegexpTokenizer('\w+')
my_df['content_token'] = my_df['content'].apply(regexp.tokenize)
my_df.head(3)
```

Gambar 3. Proses Tokenizing

Tabel 2. Hasil Tokenizing Data

Content	content_token
banyak jungler troll ngga dibasmi moonton tolong lah masa minta tolong baik baik ga digubris gua udah maki maki pun ngga berubah player tim sendiri selalu dapet trol egois pick ga jelas mending kalo bisa main udah ga jelas main kek ayam	"banyak,jungler,troll,ng ga,dibasmi,moonton,tolong, lah,masa,minta,tolong,baik, baik,ga,digubris,gua,udah, maki,maki,pun,ngga,beruba h,player,tim,sendiri,selalu,d apet,trol,egois,pick,ga,jelas, mending,kalo,bisa,main,uda h,ga,jelas,main,kek,ayam"
asalamualaikum kepada yth moonton saya ingin nanya anda niat bikin game atau niat rusuh terus matchmaking nya lama lama akhirnya pas main musuhnya jago timnya beban anda ini tidak adil kepada saya terus saya pas pake hero mmr musuh nya pada keren jago sedangkan tim saya bot beban dan pas pake hero yg lain bukan mmr tim bot musuh bot tolong perbaiki dpt tim dark system terus unistal aja lah ml mls saya di kasih menang 3x di kasih kalah 9x pintar game ml luar biasa	"asalamualaikum,kepad a,yth,moonton,saya,ingin,na nya,and,niat,bikin,game,at au,niat,rusuh,terus,matchma king,nya,lama,lama,akhirny a,pas,main,musuhnya,jago,t imnya,beban,and,ini,tidak, adil,kepada,saya,terus,saya, pas,pake,hero,mmr,musuh,n ya,pada,keren,jago,sedangk an,tim,saya,bot,beban,dan,p as,pake,hero,yg,lain,bukan, mmr,tim,bot,musuh,bot,tolo ng,perbaiki,dpt,tim,dark,sy stem,terus,unistal,aja,lah,ml ,mls,saya,di,kasih,menang,3 x,di,kasih,kalah,9x,pintar,ga me,ml,luar,biasa"
developer tolong lah kalok ngasih tim yang sesuai bisa ga sih yang jago lawan yang jago yang lemah lawan yang lemah gitu napa sih bukan malah ngasih tim beban atau di campur sama yang boot buat di gendong satu orang sekali lagi tolong lah kasih tim yang tau role hero item dan yang lainnya lama lama gue males main ni game kalau belum di benerin nih permasalahan	"developer,tolong,lah,k alok,ngasih,tim,yang,sesuai, bisa,ga,sih,yang,jago,lawan, yang,jago,yang,lemah,lawa n,yang,lemah,gitu,napa,sih, bukan,malah,ngasih,tim,beb an,atau,di,campur,sama,yan g,boot,buat,di,gendong,satu orang,sekali,lagi,tolong,lah, kasih,tim,yang,tau,role,hero ,item,dan,yang,lainya,lama,l ama,gue,males,main,ni,gam e,kalau,belum,di,benerin,ni h,permasalahan"

d. Stopwords Removal

Kata-kata berulang yang tidak memiliki makna yang spesifik atau tidak memiliki arti penting dalam struktur kalimat maka akan di *filter* (dibuang) dari data.

Tabel 3. Hasil Stopwords Removal

Content	content_token_stopwor ds
banyak jungler troll ngga dibasmi moonton tolong lah masa minta tolong baik baik ga digubris gua udah maki maki pun ngga berubah player tim sendiri selalu dapet trol egois pick ga jelas mending kalo bisa main udah ga jelas main kek ayam	"jungler,troll,ngga,d ibasmi,moonton,tolong, tolong,ga,digubris,gua,u dah,maki,maki,ngga,ber ubah,player,tim,dapet,tr ol,egois,pick,ga,mendin g,kalo,main,udah,ga,ma in,kek,ayam"
asalamualaikum kepada yth moonton saya ingin nanya anda	"asalamualaikum,yt h,moonton,nanya,niat,bi

niat bikin game atau niat rusuh terus matchmaking nya lama lama akhirnya pas main musuhnya jago timnya beban anda ini tidak adil kepada saya terus saya pas pake hero mmr musuh nya pada keren jago sedangkan tim saya bot beban dan pas pake hero yg lain bukan mmr tim bot musuh bot tolong perbaiki dpt tim dark system terus unistal aja lah ml mls saya di kasih menang 3x di kasih kalah 9x pintar game ml luar biasa

developer tolong lah kalok ngasih tim yang sesuai bisa ga sih yang jago lawan yang jago yang lemah lawan yang lemah gitu napa sih bukan malah ngasih tim beban atau di campur sama yang boot buat di gendong satu orang sekali lagi tolong lah kasih tim yang tau role hero item dan yang lainnya lama lama gue males main ni game kalau belum di benerin nih permasalahan

kin,game,niat,rusuh,matchmaking,nya,pas,main,musuhnya,jago,timnya,beban,adil,pas,pake,hero,mmr,musuh,nya,keren,jago,tim,bot,beban,pas,pake,hero,yg,mmr,tim,bot,musuh,bot,tolong,perbaiki,dpt,tim,dark,system,unistal,aja,ml,mls,kasih,menang,3x,kasih,kalah,9x,pintar,game,ml"

"developer,tolong,kalok,ngasih,tim,sesuai,ga,sih,jago,lawan,jago,lemah,lawan,lemah,gitu,napa,sih,ngasih,tim,beban,campur,boot,gendong,orang,tolong,kasih,tim,tau,role,hero,item,lainnya,gue,males,main,ni,game,benerin,nih,permasalahan"

dan pas pake hero yg lain bukan mmr tim bot musuh bot tolong perbaiki dpt tim dark system terus unistal aja lah ml mls saya di kasih menang 3x di kasih kalah 9x pintar game ml luar biasa

developer tolong lah kalok ngasih tim yang sesuai bisa ga sih yang jago lawan yang jago yang lemah lawan yang lemah gitu napa sih bukan malah ngasih tim beban atau di campur sama yang boot buat di gendong satu orang sekali lagi tolong lah kasih tim yang tau role hero item dan yang lainnya lama lama gue males main ni game kalau belum di benerin nih permasalahan

game nya bagus tapi muali dari dark sistemnya buruk sekali jaringannya dipilhkan tim fight yang selalu tolol saat main nge bug saya main game ini untuk menghibur diri tapi sistemnya sekali lagi bukan manusia yang buat tapi kaya apa yaaaaaa kalau mau update itu bagus tapi bukan dark sistem kamu yang juga kau update moonton susah karena orang moonton tambah senang makanya saya bilang bukan manusia

makin ksini makin parah bug nya sangat mengganggu saat tempur padahal jaringan wifi lancar aja belum lagi dapat musuh kelewat sakti jadi makin susah menangnya benar² terasa ketidak adilan

perbaiki,dpt,tim,dark,system,unistal,aja,ml,mls,kasih,menang,3x,kasih,kalah,9x,pintar,game,ml"

"developer,tolong,kalok,ngasih,tim,sesuai,ga,sih,jago,lawan,jago,lemah,lawan,napa,sih,ngasih,tim,beban,campur,boot,gendong,orang,tolong,kasih,tim,tau,role,hero,item,lainnya,gue,males,main,ni,game,benerin,nih,permasalahan"

"game,nya,bagus,muali,dark,sistemnya,buruk,jaringannya,dipilhkan,tim,fight,tolol,main,nge,bug,main,game,hibur,sistemnya,manusia,kaya,yaaaaaa,update,bagus,dark,sistem,kau,update,moonton,susah,orang,orang,moonton,senang,bilang,manusia"

"ksini,parah,bug,nya,mengganggu,tempur,jaringan,wifi,lancar,aja,musuh,kelewat,sakti,susah,menang,benar²,ketidakadilan"

,perbaiki,dpt,tim,dark,system,unistal,aja,ml,mls,kasih,menang,3x,kasih,kalah,9x,pintar,game,ml"

"developer,tolong,kalok,ngasih,tim,sesuai,ga,sih,jago,lawan,jago,lemah,lawan,napa,sih,ngasih,tim,beban,campur,boot,gendong,orang,tolong,kasih,tim,tau,role,hero,item,lainnya,gue,males,main,ni,game,benerin,nih,masalalah"

"game,nya,bagus,mual,dark,sistem,buruk,jaring,pilih,tim,fight,tolol,main,nge,bug,main,game,hibur,sistem,manusia,kaya,yaaaaaa,update,bagus,dark,sistem,kau,update,moonton,sah,orang,orang,moonton,senang,bilang,manusia"

"ksini,parah,bug,nya,gangguan,tempur,jaringan,wifi,lancar,aja,musuh,lewat,sakti,susah,menang,benar,tidak,adil"

e. Stemming

Tahap *stemming* adalah tahapan menyederhanakan kata-kata dalam dokumen ke bentuk kata dasar. Hal ini karena pada dasarnya kata yang berimbuhan memiliki makna yang sama dengan kata dasarnya.

Tabel 4. Hasil *Stemming*

Content	content_token_stopwords	stemmed
banyak jungler troll ngga dibasmi moonton tolong lah masa minta tolong baik baik ga digubris gua udah maki maki pun ngga berubah player tim sendiri selalu dapet trol egois pick ga jelas mending kalo bisa main udah ga jelas main kek ayam	"jungler,troll,ngga,dibasmi,moonton,tolong,tolong,ga,digubris,gua,udah,maki,maki,ngga,berubah,player,tim,dapet,trol,egois,pick,ga,mending,kalo,main,udah,ga,main,kek,ayam"	"jungler,troll,ngga,basmi,moonton,tolong,ga,gubris,gua,udah,maki,maki,ngga,ubah,player,tim,dapet,trol,egois,pick,ga,mending,kalo,main,udah,ga,main,kek,ayam"
asalamualaikum kepada yth moonton saya ingin nanya anda niat bikin game atau niat rusuh terus matchmaking nya lama lama akhirnya pas main musuhnya jago timnya beban anda ini tidak adil kepada saya terus saya pas pake hero mmr musuh nya pada keren jago sedangkan tim saya bot beban	"asalamualaikum,yth,moonton,nanya,niat,bikin,game,niat,rusuh,matchmaking,nya,pas,main,musuhnya,jago,timnya,beban,adil,pas,pake,hero,mmr,musuh,nya,keren,jago,tim,bot,beban,pas,pake,hero,yg,mmr,tim,bot,musuh,bot,tolong,	"asalamualaikum,yth,moonton,nanya,niat,bikin,game,niat,rusuh,matchmaking,nya,pas,main,musuh,jago,tim,beban,adil,pas,pake,hero,mmr,musuh,nya,keren,jago,tim,bot,beban,pas,pake,hero,yg,mmr,tim,bot,musuh,bot,tolong,

f. Filter Token by Length

Filter token adalah proses dalam preparasi data untuk menghapus beberapa kata dengan jumlah karakter tertentu. Pada penelitian ini,

panjang minimum yang digunakan untuk sebuah karakter yaitu 4 karakter. Artinya bahwa bila sebuah kata memiliki karakter dibawah 4 maka kata tersebut akan dihilangkan (Mustopa et al., 2020) .

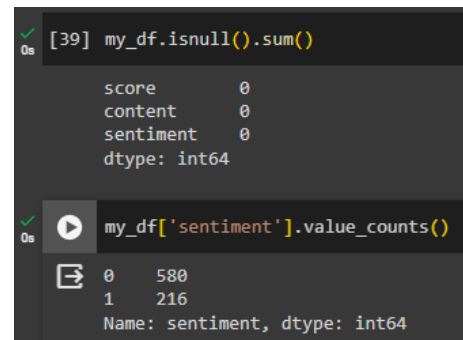
Tabel 5. Hasil Filter Token by Length

stemmed	text_string
"jungler,troll,ngga,basmi,moonton,tolong,tolong,ga,gubris,gu a,udah,maki,maki,ngga,ubah,player,tim,dapet,trol,egois,pick,ga,mending,kalo,main,udah,ga,mai n,kek,ayam"	jungler troll ngga basmi moonton tolong tolong gubris udah maki maki ngga ubah player dapet trol egois pick mending kalo main udah main ayam
"asalamualaikum,yth,moonton,nanya,niat,bikin,game,niat,ru suh,matchmaking,nya,pas,main, musuh,jago,tim,beban,adil,pas,p ake,hero,mmr,musuh,nya,keren,j ago,tim,bot,beban,pas,pake,hero ,yg,mmr,tim,bot,musuh,bot,tolo ng,perbaiki,dpt,tim,dark,syste m,unistal,aja,ml,mls,kasih,mena ng,3x,kasih,kalah,9x,pintar,gam e,ml"	asalamualaikum moonton nanya niat bikin game niat rusuh matchmaking main musuh jago beban adil pake hero musuh keren jago beban pake hero musuh tolong perbaikan dark system unistal kasih menang kasih kalah pintar game
"developer,tolong,kalok,nga sih,tim,sesuai,ga,sih,jago,lawan,j ago,lemah,lawan,lemah,gitu,nap a,sih,ngasih,tim,beban,campur,b oot,gendong,orang,tolong,kasih, tim,tau,role,hero,item,lai,gue,ma les,main,ni,game,benerin,nih,ma salah"	developer tolong kalok ngasih sesuai jago lawan jago lemah lawan lemah gitu napa ngasih beban campur boot gendong orang tolong kasih role hero item males main game benerin masalah
"game,nya,bagus,mual,dark, sistem,buruk,jaring,pilih,tim,fig ht,tolol,main,nge,bug,main,gam e,hibur,sistem,manusia,kaya,yaa aaaa,update,bagus,dark,sistem,k au,update,moonton,susah,orang, orang,moonton,senang,bilang,m anusia"	game bagus mual dark sistem buruk jaring pilih fight tolol main main game hiburan sistem manusia kaya yaaaaaa update bagus dark sistem update moonton susah orang orang moonton senang bilang manusia
"ksini,parah,bug,nya,ganggu ,tempur,jaring,wifi,lancar,aja,mu suh,lewat,sakti,susah,menang,be nar,tidak,adil"	ksini parah ganggu tempur jaring wifi lancar musuh lewat sakti susah menang benar tidak adil

4. Handling Imbalance Data

Kinerja algoritma memiliki tantangan keadaan data yang tidak seimbang dalam distribusi kelas, sehingga membuat pengambilan keputusan dari data mentah menjadi terhambat (He and Garcia, 2009). Dalam penelitian ini, diketahui total data negatif 580 dan data positif 216. Maka, diperlukan pengolahan data menggunakan teknik *Synthetic Minority Over-sampling*

Technique (SMOTE) sebagai salah satu teknik *oversampling* untuk masalah ketidakseimbangan dalam klasifikasi (Chawla et al., 2002).



Gambar 4. Hasil Pengecekan Data Kosong dan Jumlah Masing-masing Sentimen

5. Klasifikasi Naïve Bayes

Metode klasifikasi ini mengacu kepada Teorema Bayes yakni prinsip probabilitas, dengan menghitung frekuensi untuk setiap kata untuk menentukan kemungkinan hasil klasifikasi, yaitu positif dan negatif (Shaban et al., 2021; Tangkelayuk, 2022). Kebutuhan data pelatihan yang sedikit dalam penentuan estimasi parameter dalam proses klasifikasi menjadi keunggulan metode *Naïve Bayes*. Rumus metode *Naïve Bayes*.

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)} \quad (1)$$

Di mana $P(c|x)$ adalah nilai probabilitas yang ingin dihitung, $P(c)$ adalah probabilitas kelas berdasarkan hipotesis, $P(x|c)$ adalah prediktor probabilitas berdasarkan kelas yang diberikan, dan $P(x)$ adalah prediktor probabilitas secara umum (Wijaya et al., 2024).

6. Klasifikasi Support Vector Machine (SVM)

Metode klasifikasi SVM memanfaatkan pembelajaran mesin dan menentukan kelas prediksi dari model hasil pelatihan. Tingkat akurasi model tergantung pada karakteristik dan parameter kernel yang digunakan. Terdapat dua jenis SVM yang berbeda sifatnya, yakni SVM linier serta SVM non-linier. SVM linier memisahkan dua kelas pada *hyperplane* dengan batas *soft*, sementara SVM non-linier memanfaatkan fungsi trik kernel untuk

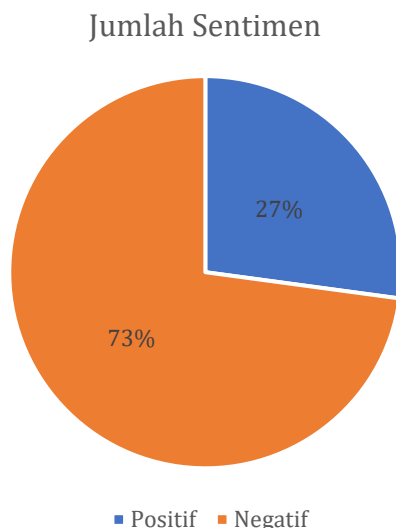
memetakan ke ruang dimensi yang lebih tinggi (Tineges et al., 2020).

7. Kurva ROC

Kurva ROC dibuat untuk perbandingan visual dari model-model klasifikasi (Nur and Oktora, 2020). Kurva ROC lebih informatif dibandingkan tabel klasifikasi, karena memberikan hasil prediksi *cut-off* yang mungkin terjadi. Rentang kurva ROC dari nol sampai satu. Semakin tinggi area dari garis diagonal kurva, maka semakin baik model tersebut dalam memberikan prediksi (Nur and Oktora, 2020).

3. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan berdasarkan data ulasan pengguna *game* Mobile Legends: Bang Bang pada *Google Playstore* periode 14 Maret 2024 hingga 25 Maret 2024 menghasilkan pembobotan sentimen untuk setiap ulasan. Pembobotan kata terbagi menjadi dua, yaitu positif dan negatif. Jumlah sentimen yang diperoleh beragam di setiap kategorinya. Sentimen negatif mencapai 73% dari total 796 *dataset* yang diambil, sementara sentimen positif mencapai 27%. Dalam total 796 ini, terdapat 580 sentimen negatif dan 216 sentimen positif. Detail pembagian ini dapat dilihat pada Gambar 5.



Gambar 5. Diagram Lingkaran Jumlah Sentimen

1. Visualisasi Data

Hasil visualisasi berupa tampilan diagram yang menggambarkan informasi

menggunakan *word cloud* pada sentimen positif dan negatif. *Word cloud* menampilkan kata yang paling sering muncul dengan ukuran yang paling besar, begitupun sebaliknya.



Gambar 6. Visualisasi Data Sentimen Positif

Dari Gambar 6, *word cloud* dengan sentimen positif menunjukkan kata-kata yang banyak muncul yaitu “main”, “jaring”, “game”, “bagus”, “tolong”, “player”, “baik” dan lainnya.



Gambar 7. Visualisasi Data Sentimen Negatif

Sedangkan, untuk sentimen negatif *word cloud* menunjukkan kata-kata yang banyak muncul yaitu “main”, “game”, “jaring”, “musuh”, “player”, “tolong”, “kalah”, dan lainnya.

2. Handling Imbalance Data

Ketidakseimbangan data diatasi menggunakan teknik SMOTE. Tahap ini dilakukan proses menyeimbangkan data menggunakan *Google Colab* dan bahasa pemrograman *Python*. Dengan memanfaatkan *library Python* berupa “*imblearn.over_sampling*” yaitu SMOTE. Dapat dilihat pada gambar 8.

```
# Handling imbalanced using SMOTE
from imblearn.over_sampling import SMOTE
smote = SMOTE()
x_sm, y_sm = smote.fit_resample(X_final, y)
```

Gambar 8. Handling Imbalanced using SMOTE

3. Pengujian Algoritma

a. Klasifikasi *Naïve Bayes*

Hasil pengujian menggunakan algoritma klasifikasi *Naïve Bayes* dari data yang telah dikumpulkan dan diujikan yaitu nilai akurasi sebesar 82% seperti pada tabel 6.

Tabel 6. *Naive Bayes Classification Report*

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.90	0.69	0.78	54
Positif	0.77	0.94	0.85	62
Accuracy			0.82	116
Macro Avg	0.84	0.81	0.81	116
Weighted Avg	0.83	0.82	0.82	116

b. Klasifikasi *Support Vector Machine (SVM)*

Berdasarkan hasil pengujian data yang telah diujikan menggunakan algoritma SVM diperoleh tingkat akurasi sebesar 91% dapat dilihat pada tabel 7.

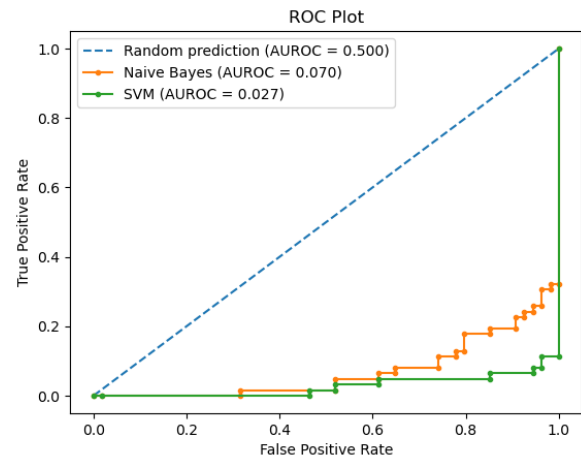
Tabel 7. *SVM Classification Report*

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.84	1.00	0.92	54
Positif	1.00	0.84	0.91	62
Accuracy			0.91	116
Macro Avg	0.92	0.92	0.91	116
Weighted Avg	0.93	0.91	0.91	116

4. Kurva ROC

Gambar 9 menunjukkan bahwa terdapat perbandingan model *Naïve Bayes* dan SVM. Pada kedua model tersebut didapatkan bahwa model *Naïve Bayes* merupakan model klasifikasi yang terbaik. Hal tersebut ditunjukkan melalui keterangan nilai AUROC *Naïve Bayes* yang bernilai 0,070 dan pada

visualisasinya juga mendekati *random classifier* (garis diagonal).



Gambar 9. Kurva ROC

4. PENUTUP

4.1. Kesimpulan

Berdasarkan hasil analisis sentimen ulasan aplikasi *Mobile Legends: Bang Bang* sebanyak 796 data. Data tersebut dibagi menjadi dua jenis sentimen, yaitu positif dan negatif. Klasifikasi sentimen tersebut memperoleh 580 sentimen negatif dan 216 sentimen positif. Ketidakseimbangan data tersebut diperbaiki dengan teknik SMOTE sehingga akurasi data meningkat dari penelitian sebelumnya. Akurasi yang didapatkan dari algoritma *Naïve Bayes* sebesar 82% dan SVM sebesar 91%. Melalui data tersebut dapat disimpulkan bahwa SVM memiliki tingkat ketepatan yang lebih tinggi. Berdasarkan kurva ROC, model klasifikasi *Naïve Bayes* lebih baik dibandingkan SVM dengan nilai AUROC sebesar 0,070.

Topik yang diangkat menjelaskan bahwasannya lebih banyak sentimen negatif. Hal ini dapat menandakan bahwa banyak dari pemain *game* tersebut kurang setuju dengan fitur maupun layanan yang diberikan. Harapan kedepannya agar pengembang maupun *publisher* dapat memberikan layanan terbaik kepada pemain.

Dalam analisis sentimen terhadap ulasan pengguna *game Mobile Legends: Bang Bang*, perbandingan antara *Naïve Bayes* dan SVM menunjukkan beberapa temuan penting. SVM cenderung memiliki kinerja lebih baik dalam mengklasifikasikan sentimen, namun *Naïve*

Bayes memiliki nilai AUROC yang lebih tinggi, menunjukkan kemampuannya dalam menangani distribusi sentimen yang kompleks. Teknik *upsampling* SMOTE berhasil meningkatkan akurasi keduanya, terutama pada *Naïve Bayes*. Hal ini menyarankan bahwa *Naïve Bayes* lebih responsif terhadap perubahan dalam distribusi sentimen dan dapat ditingkatkan kinerjanya dengan SMOTE. Dalam memilih antara SVM dan *Naïve Bayes*, perlu mempertimbangkan kebutuhan spesifik tugas analisis sentimen dan kemampuan masing-masing metode.

4.2. Saran

Penelitian selanjutnya dapat menambahkan jumlah data sehingga pembagian data untuk *testing* dan *training* dapat dimaksimalkan. Tingkat akurasi juga dapat ditambah melalui *hyperparameter tuning*. Serta dapat menggunakan pelabelan sentimen *Vader* maupun *RoBERTa* dan dapat menggunakan metode perbandingan.

5. DAFTAR PUSTAKA

- Alita, D., Isnain, A.R., 2020. Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi* 8, 50–58.
- Asosiasi Penyelenggara Jasa Internet (APJII), 2023. Survei Penetrasi & Perilaku Internet 2023 [WWW Document]. Asosiasi Penyelenggara Jasa Internet (APJII). URL <https://databoks.katadata.co.id/datapublish/2023/06/15/mayoritas-konsumen-game-online-main-lebih-dari-4-jam-sehari> (accessed 3.14.24).
- Atmajaya, D., Febrianti, A., Darwis, H., 2023a. Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *Indonesian Journal of Computer Science* 12. <https://doi.org/10.33022/ijcs.v12i4.3341>
- Atmajaya, D., Febrianti, A., Darwis, H., 2023b. Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *Indonesian Journal of Computer Science* 12.
- Chawla, N. V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16, 321–357.
- Darwis, D., Siskawati, N., Abidin, Z., 2021. Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak* 15, 131–145.
- Drus, Z., Khalid, H., 2019. Sentiment analysis in social media and its application: Systematic literature review. *Procedia Comput Sci* 161, 707–714.
- El Rahman, S.A., AlOtaibi, F.A., AlShehri, W.A., 2019. Sentiment analysis of twitter data, in: 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, pp. 1–4.
- Farhan, N.M., Setiaji, B., 2023a. Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter Mengenai Persoalan Perpu Cipta Kerja. *Indonesian Journal of Computer Science* 12. <https://doi.org/10.33022/ijcs.v12i5.3375>
- Farhan, N.M., Setiaji, B., 2023b. Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter Mengenai Persoalan Perpu Cipta Kerja: Comparison of Naive Bayes and SVM Methods on Twitter Sentiment Regarding the Government Regulations on Job Creation Issue. *Indonesian Journal of Computer Science* 12.
- Ferdiana, R., Jatmiko, F., Purwanti, D.D., Ayu, A.S.T., Dicka, W.F., 2019. Dataset Indonesia untuk Analisis Sentimen. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* 8, 334–339.
- He, H., Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21, 1263–1284.
- Ikoru, V., Sharmina, M., Malik, K., Batista-Navarro, R., 2018. Analyzing sentiments expressed on Twitter by UK energy company consumers, in: 2018 Fifth

- International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, pp. 95–98.
- Latif, R.M.A., Abdullah, M.T., Shah, S.U.A., Farhan, M., Ijaz, F., Karim, A., 2019. Data scraping from google play store and visualization of its content for analytics, in: 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET). IEEE, pp. 1–8.
- Muflih, H.Z., Abdillah, A.R., Hasan, F.N., 2023. Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer* 4, 1613–1621.
- Mustopa, A., Pratama, E.B., Hendini, A., Risdiansyah, D., 2020. Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization, in: 2020 Fifth International Conference on Informatics and Computing (ICIC). IEEE, pp. 1–7.
- Nur, M.R., Oktora, S.I., 2020. Analisis Kurva ROC Pada Model Logit Dalam Pemodelan Determinan Lansia Bekerja Di Kawasan Timur Indonesia. <https://doi.org/https://doi.org/10.29244/ij sa.v4i1.524>
- Purnama, F., 2023. Pendapatan Industri Game Indonesia Capai Rp25 Triliun [WWW Document]. ANTARA. URL <https://www.antaranews.com/berita/3774033/pendapatan-industri-game-indonesia-capai-rp25-triliun> (accessed 3.14.24).
- Shaban, W.M., Rabie, A.H., Saleh, A.I., Abo-Elvoud, M.A., 2021. Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy. *Pattern Recognit* 119, 108110.
- Tangkelayuk, A., 2022. The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)* 9, 1109–1119.
- Tineges, R., Triayudi, A., Sholihati, I.D., 2020. Analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM). *Jurnal Media Informatika Budidarma* 4, 650–658.
- Wati, R., Ernawati, S., 2021. Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas* 6, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>
- We Are Social, 2022. Digital 2022: Another Year of Bumper Growth [WWW Document]. We Are Social. URL <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/> (accessed 3.14.24).
- Wijaya, D., Saputra, R.A., Irwiensyah, F., 2024. Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer* 4, 2369–2380.