

**ANALISIS SENTIMEN TERHADAP DEVELOPER  
MOBILE LEGENDS DI TWITTER MENGGUNAKAN  
METODE MULTINOMIAL NAÏVE BAYES**

**SKRIPSI**

Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
Program Studi Informatika



Disusun Oleh:  
Damianus Halilintar  
185314086

**PROGRAM STUDI INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS SANATA DHARMA  
YOGYAKARTA  
2023**

**SENTIMENT ANALYSIS OF MOBILE LEGENDS  
DEVELOPER ON TWITTER USING MULTINOMIAL NAÏVE  
BAYES METHOD**

**THESIS**

Presented as Partial Fulfillment of Requirments  
To Obtain *Sarjana Komputer* Degree  
in Informatics Department.



Created by:

Damianus Halilintar

185314086

**INFORMATICS STUDY PROGRAM  
INFORMATICS DEPARTMENT  
FACULTY OF SCIENCE AND TECHNOLOGY  
SANATA DHARMA UNIVERSITY  
YOGYAKARTA  
2023**

**HALAMAN PERSETUJUAN PEMBIMBING**

**SKRIPSI**

**“ANALISIS SENTIMEN TERHADAP  
DEVELOPER MOBILE LEGENDS DI TWITTER  
MENGUNAKAN METODE MULTINOMIAL NAÏVE BAYES”**

Disusun oleh:

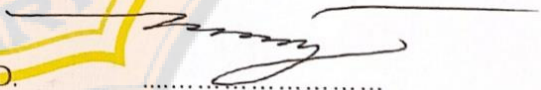
Damianus Halilintar

185314086

Telah Disetujui Oleh:

Dosen Pembimbing,

Cyprianus Kuntoro Adi, S.J. M.A., M.Sc., Ph.D.

  
27 Juli 2023

**HALAMAN PENGESAHAN**

**ANALISIS SENTIMEN TERHADAP  
DEVELOPER MOBILE LEGENDS DI TWITTER  
MENGUNAKAN METODE MULTINOMIAL NAÏVE BAYES**

Dipersiapkan dan ditulis oleh:

DAMIANUS HALILINTAR

185314086

Telah dipertahankan di depan panitia dan penguji

Pada tanggal, 13 Juli 2023

Dan dinyatakan telah memenuhi syarat.

Susunan Panitia Penguji

Nama Lengkap

Ketua

Sekretaris

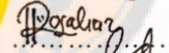
Anggota

: Rosalia Arum Kumalasanti, M.T.

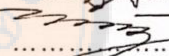
: Ir. Kartono Pinaryanto, S.T., M.Cs

: Cyprianus Kuntoro Adi, S.J. M.A., M.Sc., Ph.D.

Tanda Tangan







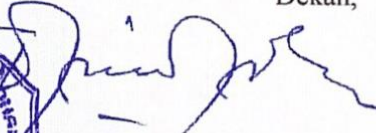
Yogyakarta, 27 Juli 2023

Fakultas Sains dan Teknologi

Universitas Sanata Dharma

Dekan,





Drs. H. Sriwindono, M.Kom., Ph.D.

## HALAMAN PERSEMBAHAN

Jangalah gelisah hatimu; Percayalah kepada Allah, Percayalah juga kepada-Ku

Yohanes 14:1



**PERNYATAAN LEMBAR KEASLIAN KARYA**

Saya menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis ini tidak memuat karya atau bagian karya orang lain, kecuali yang telah disebutkan dalam kutipan dan daftar pustaka dengan mengikuti ketentuan sebagaimana layaknya karya ilmiah.

Apabila dikemudian hari ditemukan indikasi plagiarisme dalam naskah ini, saya bersedia menanggung segala sanksi sesuai peraturan perundang-undang yang berlaku.

Yogyakarta, 12 Juli 2023

Penulis,



Damianus Halilintar



## LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini, saya mahasiswa Universitas Sanata Dharma:

Nama : Damianus Halilintar

NIM : 185314086

Demi pengembangan ilmu pengetahuan, saya memberikan karya ilmiah ini kepada Perpustakaan Universitas Sanata Dharma Karya Ilmiah yang berjudul:

### **“ANALISIS SENTIMEN TERHADAP DEVELOPER MOBILE LEGENDS DI TWITTER MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES”**

Beserta perangkat yang diperlukan (bila ada). Dengan demikian saya memberikan hak kepada Perpustakaan Universitas Sanata Dharma baik untuk menyimpan, mengalihkan dalam bentuk media lain, mengolah dalam bentuk pangkalan data, mendistribusikannya secara terbatas, dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya atau memberikan royalti kepada saya selama tetap mencantumkan nama saya sebagai penulis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Yogyakarta, 12 Juli 2023

Penulis,



Damianus Halilintar



## ABSTRAK

*E-sport* di Indonesia sudah diakui salah satu cabang olahraga Nasional salah satunya merupakan *game online* yang dikembangkan oleh *Developer Mobile legends*. Dengan tujuan perusahaan *E-sport* mengembangkan industrinya untuk dapat bersaing dan meningkatkan layanan berserta kualitas mereka. Namun faktanya masih banyak terdapat kekurangan dari *Developer Mobile Legends*, beberapa kekurangan yang timbul dari pemain yang memainkan permainan ini diantara *bug* yang terjadi didalam permainan ketika sudah berada didalam permainan, *bug* kekuatan karakter, *bug skill*, dan *bug item*. Oleh karena itu topik mengenai *Developer Mobile Legends* menjadi perbincangan dikalangan pemain. Banyak pemain yang kemudian menyapaikan pro dan kontra terhadap *Developer Mobile Legends* yang dituangkan didalam media sosial salah satunya Twitter. Penelitian ini akan menggunakan model *Multinomial Naïve Bayes* dalam melakukan analisis sentimen terhadap *Developer Mobile Legends* yang dilakukan berdasarkan hasil dari *tweet*. Pengujian akan dilakukan dengan model *k-fold cross validation* pada model *Multinomial Naïve Bayes* untuk memperoleh hasil terbaik. Dari pengujian ini hasil rata-rata terbaik pada *12-fold* dengan *accuracy* sebesar 80.82%, *recall* sebesar 80.84%, dan *precision* sebesar 81.07%.

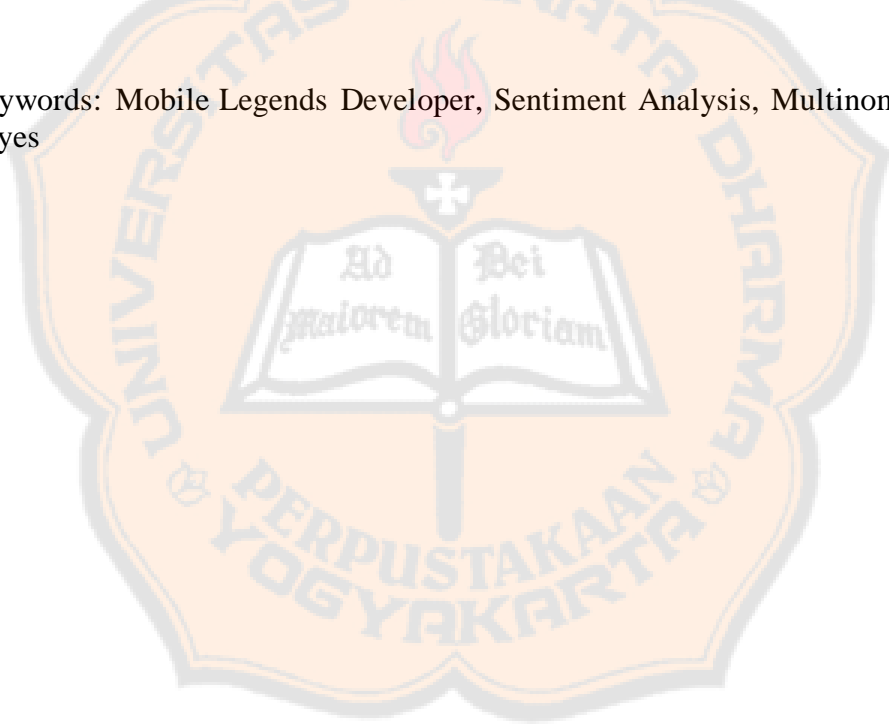
Kata Kunci: *Developer Mobile Legends*, Analisis Sentimen, *Multinomial Naïve Bayes*



## ABSTRACT

E-Sport in Indonesia has been recognized as one of the branches of national sports, one of which is an online game developed Mobile legends. With the aim of e-sport companies developing their industry so they can compete and improve their services and quality. But the fact is that there are still many shortcomings from Mobile Legends, some of the deficiencies that arise from players who play this game include bugs that occur in the game when they are in the game, character strength bugs, skill bugs, and item bugs. Therefore the topic regarding Mobile Legends is a topic of conversation among players. Many players have conveyed the pros and cons of the Mobile Legends as outlined on social media, one of which was Twitter. This research will use the Multinomial Naïve Bayes model in conducting sentiment analysis of Mobile Legends Developers based on the results of tweets. Testing will be carried out using the k-fold cross validation model on Multinomial Naïve Bayes model to get the best result. From this test the best average results are 12-fold with accuracy 80.82%, 80.84% recall, and 81.07% precision.

Keywords: Mobile Legends Developer, Sentiment Analysis, Multinomial Naïve Bayes



## KATA PENGANTAR

Segala puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan penulisan skripsi dengan judul “**ANALISIS SENTIMEN TERHADAP DEVELOPER MOBILE LEGENDS DI TWITTER MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES**” sebagai syarat untuk memperoleh gelar sarjana komputer dari program studi Informatika Universitas Sanata Dharma Yogyakarta.

Selama penulisan skripsi penulis mendapatkan bantuan, dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak, sehingga pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Tuhan Yang Maha Esa. Yang telah memberikan rahmat dan berkat-Nya untuk penulis, sehingga penulis dapat menyelesaikan skripsi ini.
2. Kedua orang tua terkasih, Ibu Anastasia dan Bapak Samuel Hinggirandja yang selalu memberikan doa, dukungan, dan semangat bagi penulis dalam menyelesaikan skripsi ini.
3. Romo Cyprianus Kuntoro Adi, S.J., M.A., M.Sc., Ph.D selaku dosen pembimbing skripsi yang telah bersedia memberikan bimbingan, masukan, serta waktu kepada penulis selama menyelesaikan skripsi.
4. Bapak Robertus Adi Nugroho S.T., M.Eng. selaku Ketua Program Studi Informatika Universitas Sanata Dharma.
5. Bapak Ir. Drs. Haris Sriwindono, M.Kom., Ph.D selaku Dekan Fakultas Sains dan Teknologi Universitas Sanata Dharma.
6. Seluruh dosen Universitas Sanata Dharma yang telah mendidik dan memberikan ilmu pengetahuan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini.
7. Olga Angely, Theodora patricia, Erik, Bagas, Daniel, Diki, Galang, Nando, Yesa, Drajat, dan Wahyu yang memberikan dukungan serta berbagi ilmu disaat penulis mengalami kesulitan saat mengerjakan skripsi.
8. Seluruh pihak yang tidak bisa penulis sebutkan satu persatu yang selalu mendukung dan mendo'akan selama penulis mengerjakan skripsi.

Penulis menyadari bahwa masih banyak kekurangan yang terdapat dalam penulisan skripsi ini. Sehingga saran dan kritik yang membangun sangat diharapkan dalam menyempurnakannya. Dan harapan penulis semoga skripsi ini dapat memberikan pengetahuan tambahan bagi pembaca.

Yogyakarta, 12 Juli 2023

Penulis,



Damianus Halilintar



## DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING.....	i
HALAMAN PENGESAHAN .....	ii
HALAMAN PERSEMBAHAN .....	iii
PERNYATAAN LEMBAR KEASLIAN KARYA .....	iv
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH .....	v
ABSTRAK.....	vi
ABSTRACT .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	8
2.1 Tinjauan Pustaka .....	8
2.2 Analisis Sentimen.....	11
2.3 Mobile Legends: Bang Bang.....	12
2.4 Vader Lexicon.....	13
2.5 Pre-processing.....	14
2.6 TF-IDF (Term Frequency Inverse Document Frequency).....	16
2.7 Synthetic Minority Over-sampling Technique (SMOTE).....	16
2.8 Metode Klasifikasi .....	17
2.8.1 Naïve Bayes Classifier.....	17
2.9 K-fold Cross Validation .....	20
2.10 Confusion Matrix .....	21

BAB III METODE PENELITIAN .....	23
3.1 Gambaran Umum Penelitian.....	23
3.2 Pengumpulan Data .....	25
3.3 Pre-processing.....	27
3.3.1 Case Folding .....	27
3.3.2 Tokenizing .....	29
3.3.3 Normalisasi .....	30
3.3.4 Stopword Removal .....	32
3.3.5 Stemming .....	34
3.4 Labeling Tweet.....	36
3.5 Perhitungan TF-IDF .....	38
3.6 SMOTE .....	40
3.7 K-fold Cross Validation .....	40
3.8 Multinomial Naïve Bayes .....	42
3.9 Pengukuran Performa.....	45
3.10 Skenario Pengujian.....	46
3.11 Peralatan Penelitian.....	47
BAB IV HASIL DAN ANALISIS .....	48
4.1 Crawling .....	48
4.2 Pre-processing.....	49
4.2.1 Case Folding .....	49
4.2.2 Tokenizing .....	50
4.2.3 Normalisasi .....	51
4.2.4 Stopword Removal .....	52
4.2.5 Stemming .....	53
4.3 Labeling Data.....	54
4.4 Pembobotan Kata .....	55
4.5 SMOTE .....	56
4.6 K-fold Cross Validation .....	57
4.7 Multinomial Naïve Bayes .....	57

4.8 Confusion Matrix .....	57
4.9 Pengujian 12-fold Cross Validation.....	58
4.9.1 Iterasi fold ke-1.....	60
4.9.2 Iterasi fold ke-2.....	61
4.9.3 Iterasi fold ke-3.....	61
4.9.4 Iterasi fold ke-4.....	62
4.9.5 Iterasi fold ke-5.....	63
4.9.6 Iterasi fold ke-6.....	63
4.9.7 Iterasi fold ke-7.....	64
4.9.8 Iterasi fold ke-8.....	64
4.9.9 Iterasi fold ke-9.....	65
4.9.10 Iterasi fold ke-10.....	66
4.9.11 Iterasi fold ke-11.....	66
4.9.12 Iterasi fold ke-12.....	67
4.10 Analisa Hasil .....	68
BAB V KESIMPULAN DAN SARAN.....	70
5.1 Kesimpulan .....	70
5.2 Saran.....	70
DAFTAR PUSTAKA.....	72
LAMPIRAN .....	74

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait .....	9
Tabel 2. 2 Confusion Matriks.....	21
Tabel 3. 1 Potongan Data Tweet.....	25
Tabel 3. 2 Case Folding .....	28
Tabel 3. 3 Tokenizing .....	29
Tabel 3. 4 Normalisasi .....	31
Tabel 3. 5 Stopword Removal.....	33
Tabel 3. 6 Stemming .....	35
Tabel 3. 7 Contoh Translate dan Labeling Data Tweet.....	37
Tabel 3. 8 Contoh Dokumen .....	39
Tabel 3. 9 Contoh Data .....	43
Tabel 3. 10 Prior Probability .....	43
Tabel 3. 11 Conditional Probability .....	44
Tabel 3. 12 Posterior Probability .....	45
Tabel 3. 13 Pengujian K-fold.....	47
Tabel 4. 1 Contoh Hasil Crawling.....	49
Tabel 4. 2 Pengujian 12-fold.....	58



## DAFTAR GAMBAR

Gambar 2. 1 Easy To Use .....	17
Gambar 3. 1 Gambaran Umum Sistem .....	24
Gambar 3. 2 Langkah Preprocessing.....	27
Gambar 3. 3 Contoh Data Pada Dokumen Teks .....	39
Gambar 3. 4 Contoh Perhitungan Nilai IDF .....	39
Gambar 3. 5 Contoh Perhitungan Nilai TF-IDF .....	39
Gambar 3. 6 Sebelum Dilakukan SMOTE.....	41
Gambar 3. 7 Sesudah Dilakukan SMOTE .....	41
Gambar 3. 8 Contoh 10-fold Cross Validation .....	42
Gambar 3. 9 Diagram Confusion Matriks.....	46
Gambar 4. 1 Source Code Crawling.....	49
Gambar 4. 2 Source Code Case Folding .....	50
Gambar 4. 3 Contoh hasil Case Folding .....	50
Gambar 4. 4 Source Code Tokenizing .....	51
Gambar 4. 5 Contoh Hasil Tokenizing.....	51
Gambar 4. 6 Source Code Normalisasi .....	52
Gambar 4. 7 Contoh Hasil Normalisasi .....	52
Gambar 4. 8 Source Code Stopword Removal .....	52
Gambar 4. 9 Contoh Hasil Stopword Removal.....	53
Gambar 4. 10 Source Code Stemming.....	53
Gambar 4. 11 Contoh Hasil Stemming .....	54
Gambar 4. 12 Source Code Labeling Data.....	54
Gambar 4. 13 Contoh Hasil Labeling Data.....	55
Gambar 4. 14 Source Code TF-IDF .....	55
Gambar 4. 15 Contoh Hasil TF-IDF .....	56
Gambar 4. 16 Source Code SMOTE.....	56
Gambar 4. 17 Contoh hasil SMOTE.....	56
Gambar 4. 18 Source Code K-fold.....	57
Gambar 4. 19 Source Code Multinomial Naïve Bayes.....	57
Gambar 4. 20 Source Code Confusion Matrix.....	58
Gambar 4. 21 Grafik Pengujian K-fold.....	68

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

*E-sport* sendiri di Indonesia sudah diakui salah satu cabang olahraga Nasional dan masuk dalam Undang-Undang No.3/2005 tentang Sistem Keolahragaan Nasional yang bertanggung jawab kepada Menpora RI dan Peraturan Pemerintah No.16/2007 tentang Penyelenggaraan Keolahragaan. *E-sport* tidak bertanding secara fisik melainkan mementingkan strategi, salah satunya *Game Online Mobile Legends* (Erwin Julius, 2016). IeSPA merupakan wadah resmi yang didirikan pemerintah sebagai asosiasi yang menaungi *Mobile Legends*, serta membantu pemain agar dapat mengembangkan bakat dan kemampuan dalam dunia *Game Online* kompetitif (Erwin Julius, 2016). *Mobile Legends* merupakan *Game Online* yang berasal dari Tiongkok atau China. *Game online* ini dikembangkan oleh *Developer* asal Negeri Tirai Bambu dan diekspansikan pada tahun 2016. Dimana *game* ini dimainkan oleh 2 tim dan masing-masing tim berisi 5 pemain yang memainkan 1 karakter. Untuk memenangkan permainan masing-masing tim harus saling menghancurkan *tower* lawan sekaligus mempertahankan *tower* sendiri.

Twitter merupakan salah satu media sosial yang sering digunakan hingga kini, Twitter dapat membantu manusia untuk saling terhubung dan berbagi macam-macam informasi satu sama lain. Twitter sendiri berdiri sejak bulan Maret tahun 2006. Pengguna twitter sekarang memiliki lebih dari 541 juta pengguna yang aktif diberbagai penjuru dunia dapat dilihat

pada <https://www.internetlivestats.com/twitter-statistics/>. Twitter memberikan pengguna fasilitas untuk menulis, membaca, serta mengirimkan informasi kepada forum umum. Twitter juga membuat fasilitas agar perusahaan dan pengguna dapat saling berinteraksi secara tidak langsung. Perusahaan dapat memberikan informasi terbaru kepada pengguna dengan waktu yang singkat, sehingga perusahaan mampu mengevaluasi layanan-layanan yang telah diberikan. Perusahaan *E-sport Mobile Legends* mengembangkan industrinya untuk dapat bersaing dan meningkatkan layanan beserta kualitas perusahaan mereka (Setiawan & Setyohadi, 2017). Media sosial Twitter menjadi sarana untuk menuangkan Banyak keresahan para pemain terhadap permainan *Mobile Legends*. Keresahan pemain mulai *bug* yang terjadi didalam permainan ketika sudah berada didalam permainan seperti ketika satu karakter melakukan penyerangan terhadap monster didalam permainan dan monster tersebut akan keluar dari tempat dia berada dan menghampiri karakter tersebut untuk menyerang balik, namun ketika karakter berhenti menyerang dan pergi maka monster tersebut akan berhenti menyerang dan seharusnya kembali ketempat monster itu berada, tetapi monster tersebut tidak kembali dan terdiam ditempat terakhir dia melakukan serangan. *Bug* karakter seperti salah satu karakter menggunakan *skill* yang seharusnya diarahkan secara manual kepada target tetapi bisa mengenai target tanpa harus mengarahkan kepada target. *Bug skill* seperti dari setiap karakter mempunyai *skill* dan rentan waktu untuk digunakan kembali *skill* tersebut, namun ada beberapa karakter yang tidak memiliki rentan waktu penggunaannya sehingga bisa

digunakan secara terus-menerus. *Bug* item seperti ada beberapa item tertentu jika digunakan oleh salah satu karakter tertentu akan membuatnya menjadi sangat kuat dibandingkan jika digunakan oleh karakter yang lainnya. Dari permasalahan tersebut dapat berpengaruh terhadap perkembangan *Developer Mobile Legends*, serta peminat dari permainan ini akan menjadi menurun (Valentina & Purnama Sari Elvivalentina, 2019).

Berbagai penelitian terkait analisis sentimen telah dilakukan, salah satunya penelitian Joko Winahyu dkk, penelitian terkait aplikasi web analisis sentimen menggunakan algoritma *Multinomial Naïve Bayes*. Penelitian ini bertujuan untuk mengetahui opini pengguna terkait kata kunci Aplikasi Web yang mana hasilnya dapat dijadikan evaluasi dalam peningkatan aplikasi web. Hasil pengujian yang diperoleh dengan menggunakan model *Multinomial Naïve Bayes* mendapatkan akurasi sebesar 69%. Penelitian analisis sentimen juga dilakukan oleh Nabila Surya Wardani dkk. Penelitian terkait analisis sentimen pemindahan ibu kota negara dengan menggunakan algoritma *Naïve Bayes* untuk model *Bernoulli* dan *Multinomial*, tujuan penelitian ini untuk mengetahui klasifikasi sentimen masyarakat terhadap pemindahan ibu kota negara. Penelitian ini membagi kelas menjadi dua kategori yaitu kelas positif dan kelas negatif. Hasil penelitian ini menghasilkan performa sensitifitas 90.1% untuk *Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes* menghasilkan performa sensitifitas sebesar 93.45%.

Dari beberapa referensi penelitian diatas terkait analisis sentimen dengan menggunakan *Multinomial Naïve Bayes*, maka dalam penelitian ini

analisis sentimen digunakan untuk menemukan informasi berharga yang dibutuhkan dari data yang tidak terstruktur, sehingga dilakukan pendekatan *machine learning* menggunakan model *Multinomial Naïve Bayes* dengan proses penyelesaian menggunakan probabilitas. Berdasarkan dari paparan diatas, penelitian ini akan melakukan analisis sentimen terhadap *Developer Mobile legends* pada media sosial Twitter menggunakan metode *Multinomial Naïve Bayes*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas dapat dibuat rumusan masalah sebagai berikut:

1. Bagaimana pendekatan *Multinomial Naïve Bayes* mampu melakukan klasifikasi sentimen terkait *Developer Mobile Legends* dengan baik?
2. Berapa akurasi yang didapatkan dari metode *Multinomial Naïve Bayes* dalam proses analisis sentimen?

## 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini untuk:

1. Mengetahui pendekatan *Multinomial Naïve Bayes* mampu melakukan klasifikasi sentimen terkait *Developer Mobile Legends* dengan baik.
2. Mengetahui hasil akurasi dari metode *Multinomial Naïve Bayes*.

#### 1.4 Manfaat Penelitian

Manfaat yang diberikan pada penelitian ini adalah Membantu untuk menganalisis sentimen data Twitter mengenai *Developer Mobile Legends* dengan menggunakan metode *Multinomial Naïve Bayes*.

#### 1.5 Batasan Masalah

1. Data yang digunakan pada penelitian hanya menggunakan data dari Twitter yaitu *tweet*.
2. Data yang digunakan di penelitian hanya menggunakan Bahasa Indonesia.
3. “MPL Indonesia” akan menjadi target analisis sentimen yang digunakan sebagai data yang mengandung *tweet*.
4. Menggunakan tweet yang diunggah pada Agustus 2022.

#### 1.6 Sistematika Penulisan

##### BAB I PENDAHULUAN

Pendahuluan berisikan dasar-dasar penulisan skripsi yang memberikan informasi awal kepada pembaca mengenai penelitian yang ditulis, dengan tujuan membangun kerangka penelitian sehingga pembaca dapat memahami bagaimana penelitian tersebut berhubungan dengan penelitian-penelitian lain. Dalam penelitian ini pendahuluan yang digunakan meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

##### BAB II LANDASAN TEORI

Berisikan teori-teori yang digunakan dalam penelitian tugas akhir. Teori ini digunakan sebagai seperangkat proposisi yang terintergrasi secara logis dengan data yang diamati untuk menjelaskan fenomena dari permasalahan yang diamati. Dalam penelitian ini landasan teori yang digunakan meliputi tinjauan pustaka, *analisis sentimen*, *Mobile Legend: Bang Bang*, *Vader Lexicon*, *Pre-processing*, TF-IDF, SMOTE, metode klasifikasi, *Naïve Bayes Classifier*, *K-Fold Cross Validation*, dan *Confusion Matriks*.

### BAB III METODE PENELITIAN

Metode ini memberikan gambaran rancangan penelitian dari investigasi yang telah diperoleh dengan teknik yang sistematis dan teliti untuk mendapatkan data dengan tujuan mengubah data dengan konsep dalam metode *Multinomial Naïve Bayes*. Meliputi gambaran umum penelitian, pengumpulan data, *pre-processing*, *labeling tweet*, perhitungan TF-IDF, SMOTE, *Kfold Cross Validation*, *Multinomial Naïve Bayes*, pengukuran performa, peralatan penelitian.

### BAB IV HASIL DAN ANALISA

Hasil dan analisis berisikan penjelasan serta interpretasi atas hasil penelitian yang dilakukan berdasarkan metode yang telah diuraikan pada bab sebelumnya. Sehingga memberikan solusi terhadap pertanyaan dalam penelitian.



## BAB V KESIMPULAN DAN SARAN

Bab ini merupakan penafsiran hasil inti pada penelitian serta saran yang berisikan opini dari penulis dari hasil peneletian terhadap penelitian yang akan dilakukan selanjutnya.



## BAB II

### LANDASAN TEORI

Pada bab berisi penjelasan teori-teori yang digunakan dalam klasifikasi sentimen, penjelasan tinjauan pustaka, analisis sentimen, *Mobile Legends: Bang Bang*, *Vader Lexicon*, *Pre-Processing*, TF-IDF, SMOTE, Metode Klasifikasi, *Naïve Bayes Classifier*, *K-Fold Cross Validation*, dan *Confusion Matrix*.

#### 2.1 Tinjauan Pustaka

Penulisan penelitian ini dilaksanakan tidak terlepas dari referensi peneliti pada penelitian sebelumnya. Berdasarkan penelitian sebelumnya topik pembahasannya tentang analisis sentimen pada Twitter. Mujib Ridwan, Hadi Suryono, dan M. Sarosa melaksanakan penelitian dengan membuat proposal, mengumpulkan data, mengolah data, melakukan implementasi *Naïve Bayes Classifier*, pengujian, dan hasil analisis. Pada saat pengelolaan pada *data mining*, seperti membersihkan data, mengintegrasikan data, menyeleksi data, melakukan transformasi terhadap data, dan membentuk *dataset* pada penelitian yang digunakan menghasilkan nilai *precision* 83%, *recall* 50%, dan akurasi 70% (Mujib Ridwan, 2013).

Joko Winahyu dan Imam Suharjo melakukan penelitian aplikasi web analisis sentimen dengan algoritma *Multinomial Naïve Bayes*. Tahapan *pre-processing* terdiri dari pembersihan data yaitu mendeteksi dan mengoreksi (atau menghapus) catatan yang rusak atau tidak akurat. *Remove duplicates* untuk menghapus data duplikat, *select attribute* untuk memilih *subset* dari

*attributes*, dan *subPreprocessing*. Dilanjutkan dengan sentimen analisis adalah proses untuk memahami dan mengelompokkan emosi (positif, negatif, dan netral), dan dilakukan klasifikasi menggunakan *Multinomial Naïve Bayes* menghasilkan tingkat akurasi 69% (Winahyu & Suharjo, 2021). Adapun beberapa penelitian terkait yang dilakukan peneliti sebelumnya dengan penelitian penulis seperti metode, topik. Dapat dilihat pada Tabel 2.1 Penelitian Terkait.

Tabel 2. 1 Penelitian Terkait

Peneliti	Judul	Model	Hasil
Akhmad Deviyanto, M. Didik R. Wahyudi	Penerapan Analisis Sentimen Pada Pengguna Twitter	<i>K-Nearest Neighbor</i>	Akurasi 67,2%, presisi 56,94%, dan <i>recall</i> 78,24%
Eka Putri Nirwandani, Indriati, Randy Cahya Wihandika	Analisis sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online	Metode <i>Modified Term Frequency Scheme</i> dan <i>Naïve Bayes</i>	Akurasi 83%, <i>recall</i> 86%, presisi 76%, dan <i>f-measure</i> 77,70%
Meilani T. H. Bunga, Bertha S. Djahi, S.Kom, M.T, Y.	Klasifikasi Status Kredit Nitra Binaan Di PT. Angkasa	<i>Multinomial Naïve Bayes</i>	Akurasi 86,56%, presisi 73%, <i>recall</i> 73%, dan <i>F-1 Measure</i>

Nabuasa, S.Kom, M.Cs	Pura I Program Kemitraan		73%
Ida Bagus Gede Widnyana Putra, Made Sudarma, I Nyoman Satya Kumara	Klasifikasi Teks Bahasa Bali	Metode <i>Supervised Learning Naïve Bayes Classifier</i>	Akurasi 95,22%
Angela Puput Giovani, Ardiansyah, Tuti haryani, Laela Kurniawati, Windu Gata	Analisis Sentimen Aplikasi Ruang Guru Di Twitter	Metode Naïve Bayes, Support Vector Machine, K- Nearest Neighbour, dan feature selection dengan PSO	didapatkan nilai tertinggi menggunakan algoritma PSO berbasis SVM dengan akurasi 78,55%, AUC 0,853
Billy Gunawan, Helen Sasty Pratiwi, Enda Esyudha Pratama	Sistem Analisis Sentimen Pada Ulasan Produk	Metode Naïve Bayes	Akurasi 77,78%, recall 93,33%, presisi 77,78%

## 2.2 Analisis Sentimen

Analisis Sentimen merupakan proses memahami, mengekstraksi, dan memproses data secara otomatis agar memperoleh sentimen pada kalimat opini (Damanik & Setyohadi, 2021). Tugas dasar dalam analisis sentimen adalah mengelompokan polaritas dari teks yang ada dalam dokumen kalimat, atau pendapat. Polaritas mempunyai arti bahwa teks memiliki aspek positif, netral, atau negatif. Analisis sentimen merupakan bagian dari *text mining* yang memanfaatkan opini, tanggapan, dan komentar berupa teks yang diolah secara komputasi (Damanik & Setyohadi, 2021). Secara umum analisis sentimen terdiri dari tiga level yaitu:

### 1. Tingkat Dokumen

Pada tingkat ini dokumen mengklasifikasikan seluruh dokumen agar mengetahui opini yang bernilai positif atau negatif. Dimana mengasumsikan setiap dokumen mengungkapkan opini tentang satu entitas, tingkat ini tidak akan sesuai jika diterapkan untuk mengevaluasi atau membandingkan beberapa entitas pada dokumen.

### 2. Tingkat Kalimat

Pada tingkat ini kalimat akan menentukan opini yang diungkapkan pada setiap kalimat bernilai positif, negatif, dan netral. Ketika kalimat bernilai netral berarti tidak ada opini. Subjektivitas tidak sama dengan sentimen banyak kalimat yang dapat menunjukkan opini.

### 3. Tingkat Entitas dan Aspek

Pada tingkat ini menganalisa dengan tidak melihat susuna bahasa sebuah dokumen, paragraf, kalimat, klausa, atau frasa, melaikan langsung pada sebuah opini. Karena pendapat terdiri dari positif dan negatif serta target (opini). Tujuannya adalah menemukan sentimen terhadap entitas pada setiap aspek yang dibahas.

### 2.3 Mobile Legends: Bang Bang

*Mobile Legends: Bang Bang* adalah permainan MOBA (*Multiplayer Online Battle Arena*) dari banyaknya permainan MOBA (*Multiplayer Online Battle Arena*) yang ada dan banyak dimainkan oleh *gamer* di Indonesia. Permainan ini pertama kali diluncurkan pada tahun 2016 dan langsung menjadi permainan yang sangat laris sampai sekarang. *Mobile Legends: Bang Bang* permainan yang dibuat oleh Moonton yang merupakan *developer* yang berasal dari China ([gamedaim.com](http://gamedaim.com)). Sehingga harus paham cara bermain *Mobile Legends: Bang Bang*. Salah satunya menguasai *hero*. *Hero* merupakan karakter didalam permainan ini, sebelum permainan dimulai pemain akan memilih masing-masing satu *hero* (karakter) ([kumparan.com](http://kumparan.com)). Total *hero* (karakter) dalam permainan saat ini adalah 106 *hero*. *Developer Mobile Legends: Bang Bang* setiap tahunnya merilis setidaknya 15 *hero* baru ([esportsku.com](http://esportsku.com)).

*Hero* (karakter) pada permainan ini memiliki peran seperti *Marksman* (mempunyai *damage* dan *attack speed* tinggi, tetapi memiliki total HP (*Health Point*) yang kecil dan rentan), *Tank* (mampu menahan *damage* lebih banyak karena mempunyai jumlah HP (*Health Point*) paling

tinggi. Tugasnya untuk menjaga *hero* lainnya dan melakukan inisiasi ketika perang), *Mage* (*hero* yang memiliki *magic damage* tinggi, tetapi memiliki jumlah HP (*Health Point*) yang kecil. Tugasnya untuk mengganggu pergerakan musuh dan mengganggu bahkan membunuh musuh), *Fighter* (*hero* yang mempunyai keseimbangan ofensif dan defensif. Tugasnya mengganggu musuh pada awal permainan), *Assassin* (*hero* yang memiliki mobilitas tinggi dan mampu membunuh musuh dengan cepat. Tugasnya membunuh musuh pada awal permainan), dan *Support* (tugasnya melindungi dan membantu tim dalam sepanjang permainan) (duniagames.co.id).

## 2.4 Vader Lexicon

*Vader* adalah singkatan dari *Valence Aware Dictionary For Social Reasoning*, model yang digunakan untuk analisis sentimen yang mampu menentukan keragaman data melalui intensitas kekuatan emosional yang ada sesuai dengan kamus *Lexicon*. *Vader* diperkenalkan pada tahun 2014 oleh C.J Hutto dan Eric Gilbert. Metode yang didasarkan pada pendekatan *human-centric*, menggabungkan analisis kualitatif dan validasi empiris (Asri et al., 2022). Kamus *lexicon* dapat digunakan untuk menilai sentimen frasa dan kalimat, tanpa perlu melihat yang lain. Sentimen dikategorikan seperti (negatif, netral, positif) (Amira Sumitro et al., 2021).

Penilaian polaritas menggabungkan fitur kamus leksikal dengan skor sentimen dari 5 kriteria yaitu tanda seru, huruf besar, tingkat susunan kata, pergeseran polaritas karena kata “tapi” serta menggunakan fitur *tri-gram* untuk memeriksa keberadaan negasi. Keuntungan menggunakan *vader*



*polarity detection* adalah tersedia kamus yang berisi nilai dari setiap kata. Kriteria pengelompokan negatif, positif, dan netral, jika hasil *compound* lebih dari 0 maka dimasukkan kategori positif, jika hasil terletak sama dengan 0 maka termasuk kategori netral, dan jika hasil *compound* dibawah 0 maka termasuk kategori negatif.

## 2.5 Pre-processing

*Pre-processing* merupakan tahap awal dalam analisis sentimen untuk mempersiapkan data teks sebelum digunakan pada proses berikutnya. Tahap ini bertujuan untuk mengubah data teks menjadi bentuk yang lebih baik sehingga menghasilkan informasi teks dengan kualitas yang baik (Fanissa et al., 2018). Tujuan dilakukan *pre-processing* memilih setiap kata dari suatu dokumen sehingga merubahnya menjadi kata dasar yang memiliki arti sempit (Amira Sumitro et al., 2021). Ada beberapa tahapan yang dilakukan sebagai berikut:

### 1. Case Folding

Pada proses *Case Folding* digunakan untuk menyamakan semua huruf menjadi huruf kecil (*lowercase*). Fungsinya untuk menghilangkan ketidakkonsistenan penggunaan huruf kapital dan huruf kecil dari 'a' sampai dengan 'z' pada data.

### 2. Tokenizing

Pada proses *Tokenizing* digunakan untuk memenggal masukan *String* pada setiap kata penyusunannya biasanya disebut *token* atau *term* dengan menggunakan tanda petik dan spasi. Fungsinya untuk memisahkan data yang menyusun pada suatu dokumen.

### 3. Normalisasi

Pada proses Normalisasi digunakan untuk kata yang tidak baku akan diubah menjadi kata yang baku dan sesuai dengan kamus besar Bahasa Indonesia. Kata yang diubah menjadi kata baku dilakukan dengan cara mengoreksi *token* pada kalimat dengan menggunakan kamus normalisasi. Normalisasi bertujuan untuk menghilangkan kerangkapan data, mengurangi kompleksitas, dan mempermudah modifikasi data.

### 4. *Stopword Removal*

Pada proses *Stopword Removal* digunakan untuk menghapus atau menghilangkan kata-kata yang memiliki informasi rendah dan tidak relevan terhadap hasil. Contohnya seperti 'dan', 'di', 'dari'. Proses *stopword removal* dengan cara mengumpulkan kata yang paling sering muncul.

### 5. *Stemming*

Pada proses *Stemming* digunakan untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan kata yang memiliki imbuhan yang terdiri dari awalan, sisipan, akhiran pada kata turunan. Contohnya 'memiliki' menjadi kata dasar 'milik'. *Stemming* bertujuan untuk mengganti bentuk suatu kata dijadikan kata dasar yang sesuai dengan morfologi Bahasa Indonesia yang baik dan benar.

## 2.6 TF-IDF (Term Frequency Inverse Document Frequency)

TF-IDF merupakan perhitungan yang menggambarkan berapa pentingnya sebuah kata (*term*) pada dokumen dan korpus. Proses ini untuk melakukan penilaian bobot relevansi *term* dari seluruh dokumen pada korpus. *Term frequency* dihitung menggunakan persamaan (1), *term frequency ke-i* muncul dalam dokumen *ke-j*. sedangkan *inverse document frequency* merupakan logaritma dari rasio jumlah dokumen dalam korpus pada persamaan (2) yaitu jumlah dokumen yang memiliki term (Yutika et al., 2021). Dengan mengalikan keduanya nilai akan didapatkan dengan formulasi pada persamaan (3).

$$tf(i) = \frac{freq\ i\ (d\ j)}{\sum_{i=1}^k freq\ i\ (d\ j)} \quad (1)$$

$$idf\ i = \log \frac{|D|}{|\{d:ti \in d\}|} \quad (2)$$

$$(tf - idf)_{ij} = tf\ i\ (d_j).idf\ i \quad (3)$$

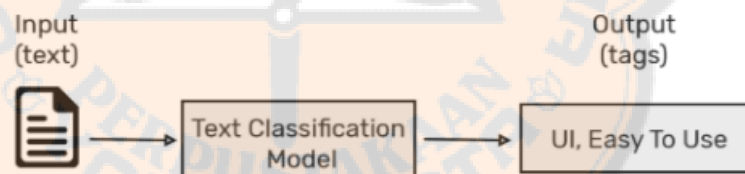
## 2.7 Synthetic Minority Over-sampling Technique (SMOTE)

Metode *Synthetic Minority Over-sampling Technique* (SMOTE) merupakan metode yang diterapkan dalam menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* dengan cara sampling ulang sampel kelas minoritas. *Synthetic Minority Over-sampling Technique* (SMOTE) adalah salah satu turunan dari *oversampling*. SMOTE pertama kali diperkenalkan oleh Nithes V. Chawla. Pendekatan ini bekerja dengan membuat replikasi

dari data minoritas. Replikasi tersebut dikenal dengan data sintesis (*syntetic data*) (Anis Nikmatul Kasanah et al., 2017).

## 2.8 Metode Klasifikasi

Klasifikasi berasal dari latin “*classic*” Klasifikasi adalah proses pengelompokan, yaitu mengumpulkan benda (entitas) yang sama kemudian memisahkan benda (entitas) yang tidak sama. Dapat dikatakan batasan klasifikasi merupakan usaha menata alam pengetahuan ke dalam tata urutan sistematis. Klasifikasi adalah pengelompokan sistematis daripada sejumlah obyek, gagasan, dan benda-benda lain kedalam kelas atau golongan tertentu didasarkan pada ciri yang sama (Mujib Ridwan, 2013). Klasifikasi mampu mengambil sebuah teks dijadikan input, menganalisis isi, dan secara otomatis menetapkan penanda (*tag*) yang benar, seperti *Easy to Use* (monkeyLearn).



Gambar 2. 1 *Easy To Use*

### 2.8.1 Naïve Bayes Classifier

*Naïve Bayes Classifier* (NBC) adalah metode klasifikasi berdasarkan *Teorema Bayes*. *Naïve Bayes* menggunakan metode probabilitas dan statistik untuk memprediksi peluang masa depan berdasarkan pengalaman atau pengetahuan dimasa lalu. Ciri utama NBC yaitu asumsi yang sangat kuat (naif) akan independensi setiap kondisi

sebuah kejadian (Thomas Bayes, abad ke-18). Setiap kelas keputusan *Naïve Bayes* menghitung probabilitas yang mempunyai syarat kelas keputusan yaitu benar. Probabilitas akan dihitung sebagai frekuensi dari tabel keputusan. Persamaan *Teorema Bayes* (Mujib Ridwan, 2013) :

$$P(H|X) = \frac{P(X|H)}{P(X)} P(H) \quad (4)$$

**Keterangan:**

X = Data dengan *class* yang belum diketahui

H = Hipotesis data yang merupakan *class* spesifik

$P(H|X)$  = Probabilitas hipotensis H berdasar kondisi X (*Posteriori Probabilitas*)

$P(H)$  = Probabilitas hipotesis H (*prior probabilitas*)

$P(X|H)$  = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$  = Probabilitas X

*Multinomial Naïve Bayes* merupakan salah satu model dari *Naïve Bayes* yang digunakan untuk mengklasifikasikan teks. *Multinomial Naïve Bayes* adalah metode *supervised learning*. Agar dapat melakukan pelatihan data harus dilabel, model ini menghitung probabilitas dari setiap kata yang muncul pada setiap label. Misalkan ada sebuah dokumen  $d$  dan himpunan kelas  $c$ . Untuk memperhitungkan kelas dari dokumen  $d$ , maka dapat dihitung dengan persamaan (5).

$$P(c|\text{term dokumen } d) = P(c). P(t_1|c). P(t_2|c). P(t_3|c) \dots P(t_n|c) \quad (5)$$

**Keterangan:**

$P(c)$  = Probabilitas *prior* dari kelas

$ctn$  = Kata dokumen  $d$  ke- $n$

$P(c|term \text{ dokumen } d)$  = probabilitas suatu dokumen termasuk kelas  $c$

$P(tn|c)$  = Probabilitas kata ke- $n$  dengan diketahui kelas  $c$

Probabilitas *prior* kelas  $c$  ditentukan dengan persamaan (6).

$$P(c) = \frac{Nc}{N} \quad (6)$$

**Keterangan:**

$Nc$  = Jumlah kelas  $c$  pada seluruh dokumen

$N$  = jumlah seluruh dokumen

Probabilitas kata ke- $n$  ditentukan menggunakan teknik *laplacian smoothing* ditentukan dalam persamaan (7).

$$P(t_n|c) = \frac{\text{count}(t_n, c) + 1}{\text{count}(c) + |V|} \quad (7)$$

**Keterangan:**

$\text{Count}(tn, c)$  = Jumlah *term*  $tn$  diseluruh data pelatihan dengan kategori  $c$

$\text{count}(c)$  = Jumlah *term* pada data pelatihan dengan kategori  $c$

$V$  = Jumlah keseluruhan *term* pada data pelatihan

Rumus *multinomial* yang digunakan untuk pembobotan kata TF-IDF ditentukan dalam persamaan (8).

$$P(t_n|c) = \frac{W_{ct}+1}{(\sum_{W' \in V} W'_{ct} + B')} \quad (8)$$

**Keterangan:**

$W_{ct}$  = Nilai pembobotan TFIDF atau  $W$  dari *term*  $t$  pada kategori  $c$

$\sum_{W' \in V} W'_{ct}$  = Total  $W$  keseluruhan term pada kategori  $c$

$B'$  = Jumlah kata unik  $W$  pada seluruh dokumen

## 2.9 K-fold Cross Validation

*K-Fold Cross Validation* merupakan metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua segmen: yang pertama untuk mempelajari atau melatih model dan lainnya digunakan untuk memvalidasi model, sehingga mendapatkan akurasi yang baik. Dengan cara membagi *dataset* sebanyak  $k$  *subset*. Satu dari *subset* dijadikan data uji dan  $k-1$  *subset* sisanya digunakan untuk proses data latih. Proses ini dilakukan sebanyak  $k$  kali sehingga setiap *subset* akan menjadi data uji dari model. mendapatkan  $k$  buah nilai performa, nilai semua performa akan dicari rata-ratanya dan nilai dengan rata-rata yang tinggi akan dijadikan sebuah model. *K-fold cross validation* memiliki kelebihan dapat mengklasifikasi *dataset* lebih efisien, juga memiliki kelemahan dalam proses komputasi yang lebih besar karena melakukan proses sebanyak  $k$  kali (Rhomadhona & Permadi, 2019).



## 2.10 Confusion Matrix

*Confusion Matrix* merupakan metode untuk melakukan perhitungan akurasi pada konsep *data mining*. *Confusion Matrix* diartikan sebagai alat yang berfungsi untuk menganalisis apakah *classifier* tersebut baik untuk mengenali *tuple* pada kelas berbeda. Nilai *True-Positive* dan *True-Negative* memberikan informasi data yang bernilai benar, untuk *False-Positive* dan *False-Negative* memberikan informasi salah dalam melakukan data. Akurasi merupakan perbandingan data terklasifikasi benar dengan keseluruhan data. *Precision* merupakan gambaran data kategori positif yang diklasifikasikan dengan benar dibagi dengan total data yang diklasifikasi positif. *Recall* berujuk pada persenan data kategori yang terklasifikasi dengan benar oleh sistem (Wardani et al., 2020).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} * 100\% \quad (10)$$

$$\text{Recall} = \frac{TP}{TP+FN} * 100\% \quad (11)$$

Tabel 2. 2 *Confusion Matriks*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP ( <i>True</i> Positif)	FP ( <i>False</i> Positif)
Negatif	FN ( <i>False</i> Negative)	TN ( <i>True</i> Negative)

Pada Tabel 2.2 *Confusion Matriks* TP (*True Positive*) merupakan jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif, FP (*False*

*Positive*) merupakan jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif, FN (*False Negative*) merupakan jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif, dan TN (*True Negative*) merupakan jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif.



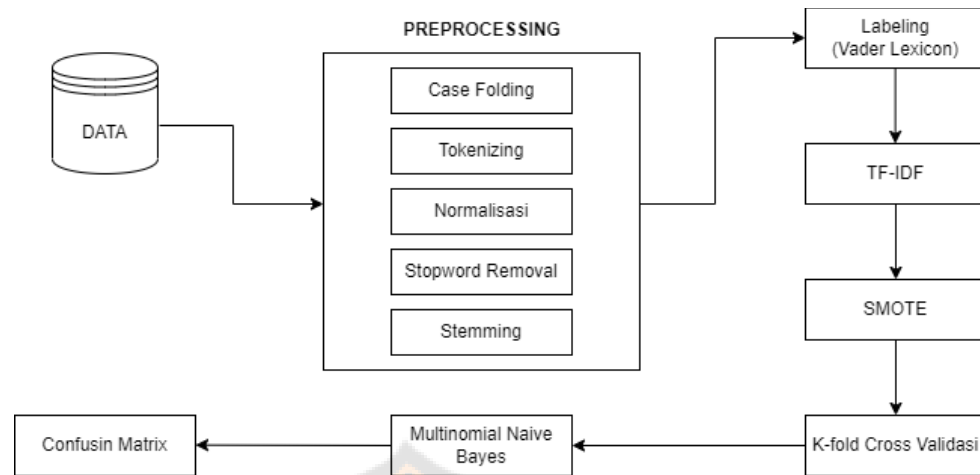
## BAB III

### METODE PENELITIAN

Bab ini akan menguraikan mengenai rencana langkah-langkah yang akan dilakukan dalam penelitian. Termasuk bagaimana cara mendapatkan data, cara mengolah data, cara analisis data, dan cara pengujian.

#### 3.1 Gambaran Umum Penelitian

Bagian ini akan menjelaskan gambaran umum dalam penelitian dan perancangan sistem dalam melakukan analisis sentimen terhadap *Developer Mobile Legends* dengan menggunakan metode *Multinomial Naïve Bayes*. Tahap awal dari sistem yang akan dibangun adalah melakukan *scrapping tweet* dari media sosial Twitter, tahap selanjutnya proses *preprocessing*. Ada beberapa tahapan pada *preprocessing* yaitu *case folding*, *tokenizing*, normalisasi, *stopword removal* dan *stemming*. Data yang sudah melewati tahap *preprocessing* akan diberikan label sentimen dengan menggunakan *Vader Lexicon*. Setelah diberi label sentiment akan dilakukan pembobotan data menggunakan metode pembobotan TF-IDF. Tahap selanjutnya adalah melakukan SMOTE guna menyeimbangkan data. Setelah data diseimbangkan dilanjutkan dengan pemodelan menggunakan *Multinomial Naïve Bayes*, didalam pemodelan ini dilakukan juga proses validasi data dengan menggunakan *K-fold Cross validation*, untuk melakukan variasi dengan mengganti nilai *k*. Hasil pegujian ini kemudian akan dievaluasi dengan menggunakan *confusion matriks* untuk mengetahui *accuracy*, *precision* dan *recall* dari hasil pemodelan. dapat dilihat pada Gambar 3.1 Gambaran Umum Sistem.



Gambar 3. 1 Gambaran Umum Sistem

Pada Gambar 3.1 merupakan alur penelitian dalam mengolah data dalam penelitian, berikut proses yang dilakukan:

1. Data mentah akan melewati tahap *preprocessing*, ada lima tahap yaitu *case folding*, *tokenizing*, *normalisasi*, *stopword removal* dan *stemming*
2. Setelah melewati tahap *preprocessing* data akan diberi labeling menggunakan *vader*
3. Setelah melewati tahap labeling data siap digunakan untuk dilakukan pembobotan kata menggunakan TF-IDF
4. Kemudian data diseimbangkan menggunakan SMOTE
5. Data yang sudah diseimbangkan akan dihitung menggunakan rumus *Multinomial Naive Bayes*
6. Dilakukan proses validasi data menggunakan *k-fold* dengan melakukan variasi nilai *k*
7. Selanjutnya akan dilakukan evaluasi sistem dari metode *Multinomial Naïve Bayes* menggunakan *Confusion Matrix*.

### 3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan *tweet* yang didapatkan dari media sosial Twitter yang hanya menggunakan Bahasa Indonesia, dengan menggunakan *Google Colab*. Kata kunci atau *hashtag* *tweet* yang digunakan dalam pencarian adalah “MPL Indonesia” pada Agustus 2022. Sehingga mendapatkan 4000 data *tweet* dan diunduh dalam bentuk csv. Kemudian dari 4000 data *tweet* tersebut dilakukan pembersihan data tahap awal dengan menghilangkan data secara manual, yaitu data yang tidak berkaitan dengan kata kunci, seperti data yang mengandung iklan jualan *diamonds Mobile Legends*, data promo harga joki *rank* akun *Mobile Legends*, data informasi diadakan turnamen *Mobile Legends*, data yang menggunakan bahasa daerah dan bahasa asing. Setelah dilakukan pembersihan tahap awal sehingga menjadi 2223 data *tweet* yang mengandung label positif, netral, dan negatif. Data tersebut diperoleh dengan menggunakan fasilitas Twitter yaitu *Twitter API (Application Programming Interface)*. Data *tweet* diambil menggunakan *library* bahasa pemrograman *python* yaitu *snsrape*. Tabel 3.1 Potongan Data *Tweet*, yang digunakan untuk penelitian *sentiment analysis*.

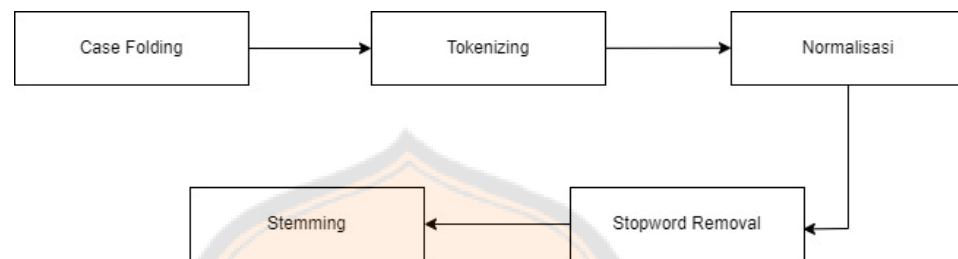
Tabel 3. 1 Potongan Data *Tweet*

No	Date Time	Tweet
1	2022-08-31 23:37:02+00:00	Jumat diajak nonton di MPL Arena langsung yak. Siapin death note buat nyatet nyatet ini mah
2	2022-08-31 23:26:44+00:00	5 Hero Langganan Ban di Week 3 Regular Season MPL Season 10 <a href="https">https</a>
3	2022-08-31 23:10:08+00:00	@el_wjj ehh u pada ngikutin MPL gk?
4	2022-08-31 23:07:40+00:00	@mlbbfess seingetku ini tu pas release hero, lama bgt, mpl sisen 6-7 kalo

		gasalah inget
5	2022-08-31 23:03:35+00:00	@islameyra @mlbbfess Bukan kak, room tourney tu macem room mpl gitu cmiiw
6	2022-08-31 22:37:06+00:00	Galau bgt anjir ke mpl arena bawa motor apa pp apa staycation sekalian anjaay wkwkwk
7	2022-08-31 22:07:14+00:00	@athnashield kita ke jkt aja yuu liat mpl terus ke pik with ayang 🥰👍
8	2022-08-31 21:41:05+00:00	@yuzzhong Nonton bareng sih mpl
9	2022-08-31 21:36:59+00:00	@FWBESS Outfit ke MPL Arena nih
10	2022-08-31 21:07:08+00:00	@bigetroneports BA mulu, itu MDL sama MPL bapuk.
11	2022-08-31 20:37:04+00:00	Mau komen RA bosan. 😞
12	2022-08-31 20:07:11+00:00	pengen mpl date deh ama bibub
13	2022-08-31 19:36:59+00:00	@pajahanep @ndagels nonton mpl id pake inggris 😎
14	2022-08-31 19:06:48+00:00	@teamrrqofficial Bantuin mpl dong rrq sena. 🤔
15	2022-08-31 18:34:56+00:00	kalian masih ingat ini gak sih pas mpl malay s7 todak pake jungler miya game 1 ama jungler layla game 2 dan menang 2-0 lawan red esport. ini ngakak bgt kalo di rewatch pls. 🤔🤔 MI! https
16	2022-08-31 18:07:07+00:00	btr ga yg mpl ga mdl, struggle semua 🤔
17	2022-08-31 17:51:22+00:00	ganti belletron aja
18	2022-08-31 17:36:54+00:00	Ramai Perseteruan RRQ dan EVOS Usai El Clasico MPL Season 10, Xin: Kalah ya Harus Terima <a href="https://t.co/T419EIK1NO">https://t.co/T419EIK1NO</a>
19	2022-08-31 16:32:24+00:00	Stay private hanya untuk orang2 yang tidak bangga dengan pasangannya
20	2022-08-31 16:13:42+00:00	Bulan tercipta ketika batu seukuran planet Mars menabrak Bumi, tidak lama setelah tata surya mulai terbentuk sekitar 4,5 miliar tahun lalu.

### 3.3 Pre-processing

Pada proses *pre-processing* data akan diolah untuk menghilangkan *noise* dan kata baku serta kalimat yang tidak berguna sehingga menjadi lebih terstruktur. Dapat dilihat pada Gambar 3.2 Langkah *Preprocessing*.



Gambar 3. 2 Langkah *Preprocessing*

#### 3.3.1 Case Folding

Tahap *Case Folding* akan mengubah seluruh kata yang memiliki huruf kapital menjadi huruf kecil, berfungsi untuk menjadi format yang sama untuk setiap katanya.

Langkah-langkah *case folding* sebagai berikut:

1. Baca setiap bari pada dokumen
2. Mengubah setiap huruf kapital menjadi huruf kecil
3. Menyimpan dokumen yang telah diubah menjadi huruf kecil

Contoh proses *case folding* pada Tabel 3.2 *Case Folding*.

Tabel 3. 2 Case Folding

<i><b>Tweet</b></i>	<i><b>Case Folding</b></i>
Pembaruan terbaru banyak karakter diubah jadi lemah dan dibuat menjadi lebih kuat gak sabar mau coba.	pada pembaruan banyak karakter yang berubah, gak sabar gue untuk bermain.
Yve jadi aneh sekarang, lebih suka yang dulu.	yve jadi aneh sekarang lebih suka dulu
Main sendiri untuk ningkatin malah ketemu tim bodoh parah sih cok.	main sendiri untuk ningkatin malah ketemu tim bodoh parah sih cok.
Seru banget main pas malam, menang terus.	seru banget main pas malam, menang terus.
Moonton parah banget cok bug beatrix sakit banget. Wajib ban sih.	moonton parah banget cok bug beatrix sakit banget. wajib ban sih.
Esmeralda susah banget lawan dyroth dimenit awal, hero darahnya tebal jadi kayak kertas cok.	esmeralda susah banget lawan dyroth dimenit awal, hero darahnya tebal jadi kayak kertas cok.
Kadita nyeselin banget. Selalu nunggu dirumput langsung mati.	kadita nyeselin banget. selalu nunggu dirumput langsung mati.



Kualitas gamenya biasa aja	kualitas gamenya biasa aja
----------------------------	----------------------------

### 3.3.2 Tokenizing

Tahap *Tokenizing* merupakan proses membagi kalimat atau memisahkan kalimat menjadi beberapa bagian kata yang disebut *token*. Serta menghilangkan emoji dan tanda baca pada data. Langkah-langkah *tokenizing* sebagai berikut:

1. Baca setiap baris pada dokumen
2. Mengambil *token* dari kalimat dengan menggunakan spasi untuk memisahkan *token* satu dengan *token* lainnya
3. Menghapus tanda baca, simbol, emoji, *mention*, *hashtag*, pada data
4. Simpan dokumen yang telah diubah menjadi *token*.

Contoh *tokenizing* dapat dilihat pada Tabel 3.3 *Tokenizing*.

Tabel 3. 3 *Tokenizing*

Hasil <i>Case Folding</i>	<i>Tokenizing</i>
pada pembaruan banyak karakter yang berubah, gak sabar gue untuk bermain.	'pada', 'pembaruan', 'banyak', 'karakter', 'yang', 'berubah', 'gak', 'sabar', 'gue', 'untuk', 'bermain'
yve jadi aneh sekarang lebih suka dulu	'yve', 'jadi', 'aneh', 'sekarang', 'lebih', 'suka', 'yang', 'dulu'
main sendiri untuk ningkatin malah	'main', 'sendiri', 'untuk', 'ningkatin', 'malah', 'ketemu',

ketemu tim bodoh parah sih cok.	'tim', 'bodoh', 'parah', 'sih', 'cok'
seru banget main pas malam, menang terus.	'seru', 'banget', 'main', 'pas', 'malam', 'menang', 'terus'
moonton parah banget cok bug beatrix sakit banget. wajib ban sih.	'moonton', 'parah', 'banget', 'cok', 'bug', 'beatrix', 'sakit', 'banget', 'wajib', 'ban', 'sih'
esmeralda susah banget lawan dyroth dimenit awal, hero darahnya tebal jadi kayak kertas cok.	'esmeralda', 'susah', 'banget', 'lawan', 'dyroth', 'dimenit', 'awal', 'hero', 'darahnya', 'tebal', 'jadi', 'kayak', 'kertas', 'cok'
kadita nyeselin banget. selalu nunggu dirumput langsung mati.	'kadita', 'nyeselin', 'banget', 'selalu', 'nunggu', 'dirumput', 'langsung', 'mati'
kualitas gamenya biasa aja	'kualitas', 'gamenya', 'biasa', 'aja'

### 3.3.3 Normalisasi

Tahap normalisasi berfungsi untuk mengubah kata yang salah ketik dan berlebihan menjadi kata baku. Normalisasi bertujuan untuk menghilangkan kerangkapan data, mengurangi kesulitan, dan memudahkan perubahan kata. Contohnya seperti kata 'ndak' yang seharusnya 'tidak' dan 'mndi' yang seharusnya 'mandi'.

Langkah-langkah normalisasi sebagai berikut:

1. Baca baris disetiap dokumen apakah sesuai dengan standar kamus kata, maka *token* tidak diubah
2. Jika tidak sama dengan kamus maka *token* akan diubah

Contoh normalisasi dapat dilihat pada Tabel 3.4 Normalisasi.

Tabel 3. 4 Normalisasi

Hasil <i>Tokenizing</i>	Normalisasi
'pada', 'pembaruan', 'banyak', 'karakter', 'yang', 'berubah', 'gak', 'sabar', 'gue', 'untuk', 'bermain'	'pada', 'baru', 'banyak', 'karakter', 'yang', 'ubah', 'tidak', 'sabar', 'aku', 'untuk', 'main'
'yve', 'jadi', 'aneh', 'sekarang', 'lebih', 'suka', 'yang', 'dulu'	'yve', 'jadi', 'aneh', 'sekarang', 'lebih', 'suka', 'yang', 'dulu'
'main', 'sendiri', 'untuk', 'ningkatin', 'malah', 'ketemu', 'tim', 'bodoh', 'parah', 'sih', 'cok'	'main', 'sendiri', 'untuk', 'tingkat', 'malah', 'ketemu', 'tim', 'bodoh', 'parah', 'sih', 'cok'
'seru', 'banget', 'main', 'pas', 'malam', 'menang', 'terus'	'seru', 'banget', 'main', 'pas', 'malam', 'menang', 'terus'
'moonton', 'parah', 'banget', 'cok', 'bug', 'beatrix', 'sakit', 'banget', 'wajib', 'ban', 'sih'	'moonton', 'parah', 'banget', 'cok', 'eror', 'beatrix', 'sakit', 'banget', 'wajib', 'larang', 'sih'
'esmeralda', 'susah', 'banget',	'esmeralda', 'susah', 'banget',

‘lawan’, ‘dyroth’, ‘dimenit’, ‘awal’, ‘hero’, ‘darahnya’, ‘tebal’, ‘jadi’, ‘kayak’, ‘kertas’, ‘cok’	‘lawan’, ‘dyroth’, ‘dimenit’, ‘awal’, ‘karakter’, ‘darahnya’, ‘tebal’, ‘jadi’, ‘kayak’, ‘kertas’, ‘cok’
‘kadita’, ‘nyeselin’, ‘banget’, ‘selalu’, ‘nunggu’, ‘dirumput’, ‘langsung’, ‘mati’	‘kadita’, ‘kesal’, ‘banget’, ‘selalu’, ‘nunggu’, ‘dirumput’, ‘langsung’, ‘mati’
‘kualitas’, ‘gamenya’, ‘biasa’, ‘aja’	‘kualitas’, ‘permainannya’, ‘biasa’, ‘aja’

### 3.3.4 Stopword Removal

Tahap *Stopword Removal* berfungsi untuk menghilangkan kata yang tidak bermakna atau kata umum. Contohnya “dan”, “di”, “yang” merupakan kata yang memiliki informasi rendah dari sebuah dokumen.

Langkah-langkah *stopword removal* sebagai berikut:

1. Baca baris pada setiap dokumen
2. Bandingkan *token* dari dokumen dengan kamus *stopword*
3. Jika dokumen terdapat *stoplist*, maka *token* dihapus
4. Jika kata dalam dokumen tidak ada pada kamus *stopword*, maka kata akan disimpan

Contoh *stopword removal* pada Tabel 3.5 *Stopword Removal*.

Tabel 3. 5 *Stopword Removal*

<b>Hasil Normalisasi</b>	<b><i>Stopword Removal</i></b>
‘pada’, ‘baru’, ‘banyak’, ‘karakter’, ‘yang’, ‘ubah’, ‘tidak’, ‘sabar’, ‘aku’, ‘untuk’, ‘main’	‘pada’, ‘baru’, ‘banyak’, ‘karakter’, ‘ubah’, ‘sabar’, ‘main’
‘yve’, ‘jadi’, ‘aneh’, ‘sekarang’, ‘lebih’, ‘suka’, ‘yang’, ‘dulu’	‘yve’, ‘jadi’, ‘aneh’, ‘sekarang’, ‘lebih’, ‘suka’, ‘dulu’
‘main’, ‘sendiri’, ‘untuk’, ‘tingkat’, ‘malah’, ‘ketemu’, ‘tim’, ‘bodoh’, ‘parah’, ‘sih’, ‘cok’	‘main’, ‘sendiri’, ‘tingkat’, ‘malah’, ‘ketemu’, ‘tim’, ‘bodoh’, ‘parah’
‘seru’, ‘banget’, ‘main’, ‘pas’, ‘malam’, ‘menang’, ‘terus’	‘seru’, ‘main’, ‘malam’, ‘menang’, ‘terus’
‘moonton’, ‘parah’, ‘banget’, ‘cok’, ‘eror’, ‘beatrice’, ‘sakit’, ‘banget’, ‘wajib’, ‘larang’, ‘sih’	‘moonton’, ‘parah’, ‘eror’, ‘beatrice’, ‘sakit’, ‘wajib’, ‘larang’
‘esmeralda’, ‘susah’, ‘banget’, ‘lawan’, ‘dyroth’, ‘dimenit’, ‘awal’, ‘karakter’, ‘darahnya’, ‘tebal’, ‘jadi’, ‘kayak’, ‘kertas’, ‘cok’	‘esmeralda’, ‘susah’, ‘lawan’, ‘dyroth’, ‘dimenit’, ‘awal’, ‘karakter’, ‘darahnya’, ‘tebal’, ‘kertas’
‘kadita’, ‘kesal’, ‘banget’, ‘selalu’, ‘nunggu’, ‘dirumput’,	‘kadita’, ‘kesal’, ‘selalu’, ‘nunggu’, ‘dirumput’, ‘langsung’, ‘mati’

'langsung', 'mati'	
'kualitas', 'permainannya', 'biasa', 'aja'	'kualitas', 'main', 'biasa'

### 3.3.5 Stemming

Tahap *stemming* berfungsi untuk mengubah kata pada suatu dokumen menjadi kata dasar yang lebih mudah dipahami, berdasarkan kamus *stemming*. Proses ini juga menghilangkan kata imbuhan dan akhiran suatu kata.

langkah-langkah *stemming* sebagai berikut:

1. Baca baris disetiap dokumen
2. Bandingkan dengan kamus *stemming*
3. Jika sama dengan kamus *stemming*, maka kata tersebut merupakan kata dasar
4. Jika tidak sama maka imbuhan kata akan dihapus.

Contoh *stemming* pada Tabel 3.6 *Stemming*

Tabel 3. 6 *Stemming*

<b>Hasil Stopword Removal</b>	<b>Stemming</b>
‘pada’, ‘baru’, ‘banyak’, ‘karakter’, ‘ubah’, ‘sabar’, ‘main’	‘baru’, ‘karakter’, ‘ubah’, ‘sabar’, ‘main’
‘yve’, ‘jadi’, ‘aneh’, ‘sekarang’, ‘lebih’, ‘suka’, ‘dulu’	‘yve’, ‘jadi’, ‘aneh’, ‘sekarang’, ‘lebih’, ‘suka’, ‘dulu’
‘main’, ‘sendiri’, ‘tingkat’, ‘malah’, ‘ketemu’, ‘tim’, ‘bodoh’, ‘parah’	‘main’, ‘sendiri’, ‘tingkat’, ‘malah’, ‘ketemu’, ‘tim’, ‘bodoh’, ‘parah’
‘seru’, ‘main’, ‘malam’, ‘menang’, ‘terus’	‘seru’, ‘main’, ‘malam’, ‘menang’, ‘terus’
‘moonton’, ‘parah’, ‘eror’, ‘beatrice’, ‘sakit’, ‘wajib’, ‘larang’	‘moonton’, ‘parah’, ‘eror’, ‘beatrice’, ‘sakit’, ‘wajib’, ‘larang’
‘esmeralda’, ‘susah’, ‘lawan’, ‘dyroth’, ‘dimenit’, ‘awal’, ‘karakter’, ‘darahnya’, ‘tebal’, ‘kertas’	‘esmeralda’, ‘susah’, ‘lawan’, ‘dyroth’, ‘menit’, ‘awal’, ‘karakter’, ‘darah’, ‘tebal’, ‘kertas’
‘kadita’, ‘kesal’, ‘selalu’, ‘nunggu’, ‘dirumput’, ‘langsung’, ‘mati’	‘kadita’, ‘kesal’, ‘selalu’, ‘nunggu’, ‘rumput’, ‘langsung’, ‘mati’
‘kualitas’, ‘main’, ‘biasa’	‘kualitas’, ‘main’, ‘biasa’

### 3.4 Labeling Tweet

Labeling *tweet* akan menggunakan *tweet* berjumlah 2223, data *tweet* yang akan digunakan belum mempunyai label sentimen, sehingga perlu diberi label terlebih dahulu terhadap data. Modul *Vader* hanya dapat digunakan untuk teks berbahasa Inggris, Sehingga *tweet* harus terlebih dahulu diterjemahkan dari bahasa Indonesia ke dalam bahasa Inggris. Proses *translate* data menggunakan *google translate*, sebelum *translate* dilakukan dilakukan pengecekan pada kata, sehingga kata sesuai dengan EYD bahasa Indonesia. data yang sudah diterjemahkan menjadi bahasa Inggris telah siap untuk proses labeling dengan menggunakan *library vader* pada *python*. Setelah dilakukan pelabelan terhadap 2223 data *tweet* didapat label positif berjumlah 1003 data, label netral berjumlah 806 data, dan label negatif berjumlah 414 data, dari hasil labeling yang didapatkan tidak seimbang karena perbedaan label netral dan label negatif terhadap label positif terbilang sangat jauh, sehingga perlu dilakukan peyeimbangan terhadap data. Contoh *translate* data *tweet* dapat dilihat pada Tabel 3.7.



Tabel 3. 7 Contoh *Translate* dan *Labeling Data Tweet*

<b><i>Tweet</i> sebelum ditranslate</b>	<b><i>Tweet</i> sesudah ditranslate</b>	<b>Compound</b>	<b>Label</b>
'baru', 'karakter', 'ubah', 'sabar', 'main'	'new', 'character', 'change', 'patient', 'play'	0.2263	POSITIF
'yve', 'jadi', 'aneh', 'sekarang', 'lebih', 'suka', 'dulu'	'yve', 'so', 'weird', 'now', 'more', 'like', 'was'	-0.1280	NEGATIF
'main', 'sendiri', 'tingkat', 'malah', 'ketemu', 'tim', 'bodoh', 'parah'	'play', 'alone', 'level', 'even', 'meet', 'team', 'stupid', 'severe'	-0.4391	NEGATIF
'seru', 'main', 'malam', 'menang', 'terus'	'fun', 'play', 'night', 'win', 'keep'	0.8643	POSITIF
'moonton', 'parah', 'eror', 'beatrix', 'sakit', 'wajib', 'larang'	'moonton', 'severe', 'error', 'beatrix', 'sick', 'mandatory', 'banned'	-0.8119	NEGATIF
'esmeralda', 'susah', 'lawan', 'dyroth', 'menit', 'awal', 'karakter', 'darah',	'esmeralda', 'hard', 'against', 'dyroth', 'minute', 'beginning',	-0.0498	NEGATIF

'tebal', 'kertas'	'character', 'blood', 'thick', 'paper'		
'kadita', 'kesal', 'selalu', 'nunggu', 'rumput', 'langsung', 'mati'	'kadita', 'annoyed', 'always', 'waiting', 'grass', 'immediately', 'dead'	-0.7713	NEGATIF
'kualitas', 'main', 'biasa'	'quality', 'play', 'ordinary'	0.0000	NETRAL

### 3.5 Perhitungan TF-IDF

Setelah melakukan tahap *pre-processing*, pada tahap ini akan dilakukan perhitungan kata atau *term*. Kata dari *tweet* akan dihitung jumlah kemunculannya pada setiap dokumen. Langkah-langkah dalam pembobotan TFIDF dilakukan sebagai berikut:

Tabel 3. 8 Contoh Dokumen

Dokumen I	Dokumen II	Dokumen III
Akhirnya ada permainan mobile baru	Teman aku bilang sih namanya ada legendnya	Iya betul sekali ada legendnya, mobile legend
Jumlah dokumen = 3		

## 1. Menghitung semua kata yang muncul pada dokumen teks

Token	Kata Kunci (kk)	TF		
		D1	D2	D3
main	0	1	0	0
mobile	1	1	0	1
baru	0	1	0	0
nama	0	0	1	0
legend	1	0	1	2
betul	0	0	0	1

Gambar 3. 3 Contoh Data Pada Dokumen Teks

## 2. Menghitung nilai IDF

Token	Kata Kunci (kk)	TF			df	D/df
		D1	D2	D3		
main	0	1	0	0	1	3
mobile	1	1	0	1	2	1.5
baru	0	1	0	0	1	3
nama	0	0	1	0	1	3
legend	1	0	1	2	2	1.5
betul	0	0	0	1	1	3

Gambar 3. 4 Contoh Perhitungan Nilai IDF

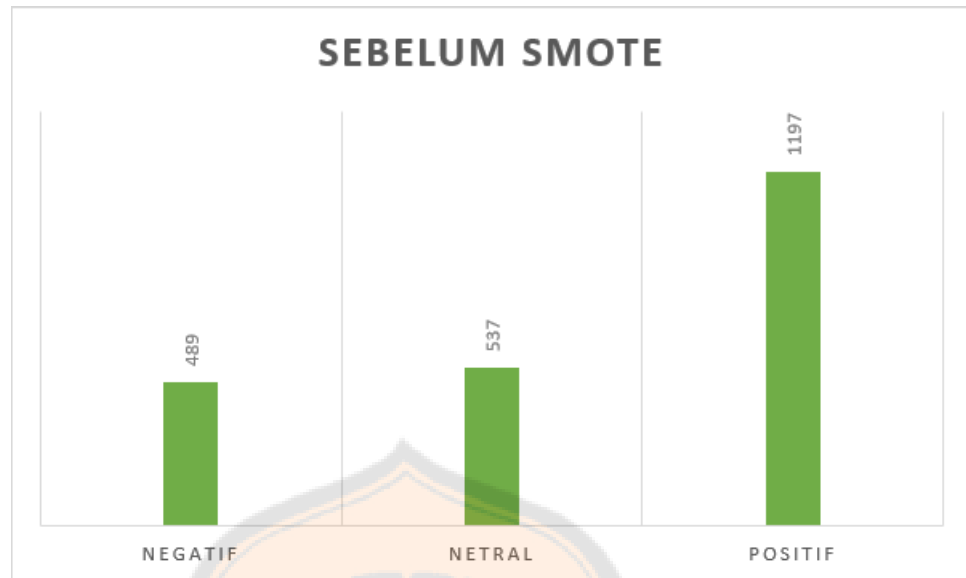
## 3. Menghitung nilai TF-IDF

Token	Kata Kunci (kk)	TF			df	D/df	IDF (log D/df)	W			
		D1	D2	D3				kk	D1	D2	D3
main	0	1	0	0	1	3	0.477121255	0	0.477	0	0
mobile	1	1	0	1	2	1.5	0.176091259	0.176	0.176	0	0.176
baru	0	1	0	0	1	3	0.477121255	0	0.477	0	0
nama	0	0	1	0	1	3	0.477121255	0	0	0.477	0
legend	1	0	1	2	2	1.5	0.176091259	0.176	0	0.176	0.352
betul	0	0	0	1	1	3	0.477121255	0	0	0	0.477

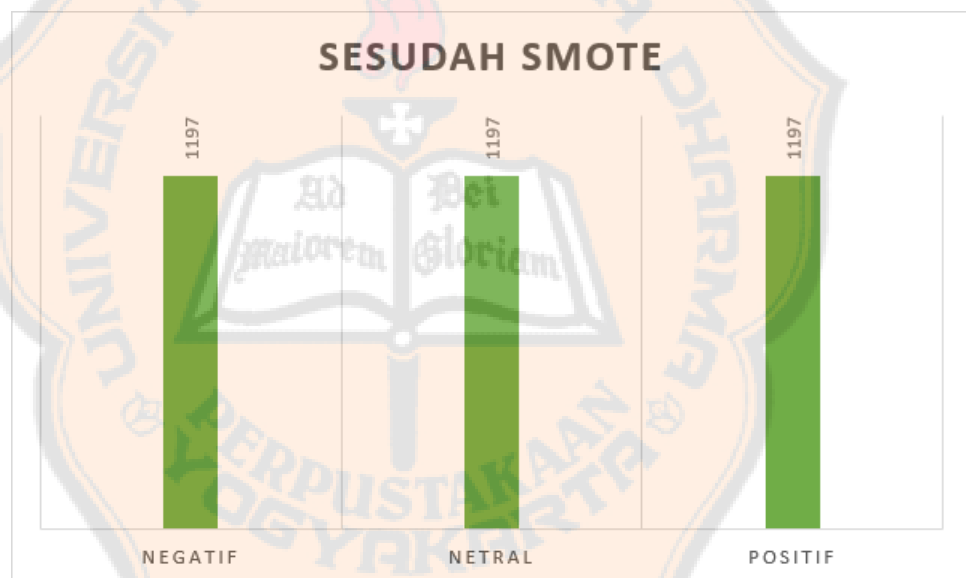
Gambar 3. 5 Contoh Perhitungan Nilai TF-IDF

### 3.6 SMOTE

Pada tahap ini data akan dilakukan resampling untuk menangani kumpulan data yang tidak seimbang dengan oversampling SMOTE. Dimana jumlah data 2223 dengan label berjumlah 1003 data label positif, label netral berjumlah 806 data, dan label negatif berjumlah 414 data. Sehingga dilakukan penyeimbangan data menjadi 3009 dengan label positif berjumlah 1003, label netral berjumlah 1003 dan label negatif berjumlah 1003. Jumlah data pada kelas positif (kelas mayoritas) sangat besar dibandingkan kelas netral dan negatif (kelas minoritas). Kelas yang tidak seimbang tidak dapat bekerja dengan baik dalam klasifikasi, sehingga kelas minoritas sering disalah klasifikasikan sebagai kelas mayoritas dan membuat hasil klasifikasi bisa menjadi kecil. Dengan dilakukan penyeimbangan terhadap kelas akan membuat hasil klasifikasi bekerja dengan baik dan mendapatkan hasil akurasi yang lebih baik. Proses SMOTE dapat dilihat pada Gambar 3.6 Sebelum Dilakukan SMOTE dan Gambar 3.7 Sesudah Dilakukan SMOTE.



Gambar 3. 6 Sebelum Dilakukan SMOTE



Gambar 3. 7 Sesudah Dilakukan SMOTE

### 3.7 K-fold Cross Validation

*K-Fold Cross Validation* merupakan salah satu pengujian *cross validation* yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai *k-fold*. Salah satu kelompok *k-fold* tersebut akan

dijadikan sebagai data *testing* dan sisanya kelompok lainnya akan dijadikan sebagai data *training* dengan jumlah data awal 2223 *tweet* setelah *balacing* 3591. Contoh *k-fold cross validation* dapat dilihat pada Gambar 3.8 contoh *10-fold Cross Validation*.

Iterasi										
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

Data Testing
  Data Training

Gambar 3. 8 Contoh *10-fold Cross Validation*

Berikut langkah-langkah pengujian data dengan *10-fold cross validation*.

1. Dataset yang digunakan akan dibagi menjadi 10 bagian, yaitu D1, D2, D3, ... dan Dt,  $t = (1, 2, 3, \dots)$  yang digunakan untuk data *testing* dan dataset sisanya dijadikan sebagai data *training*.
2. data dilakukan klasifikasi menggunakan *Multinomial Naïve Bayes*.

### 3.8 Multinomial Naïve Bayes

*Multinomial Naive Bayes* akan menentukan label dari *tweet* berdasarkan probabilitas kemunculan kata dari proses sebelumnya. Data yang sudah melewati tahap *preprocessing* tersebut akan digunakan sebagai data *training*. Dalam perhitungan manual akan melakukan percobaan mengambil 1 data *testing* dan data yang lain akan digunakan sebagai data *training*. Contoh data *testing* ditunjukkan pada Tabel 3.9.

Tabel 3. 9 Contoh Data

	<i>Tweet</i>	<b>Label</b>
D1	baru karakter ubah sabar main	Positif
D2	yve jadi aneh sekarang lebih suka dulu	Negatif
D3	kualitas main biasa	Netral

Tahap klasifikasi *Multinomial Naive Bayes* akan menentukan label dari *tweet* dan melewati beberapa tahap klasifikasi sebagai berikut:

#### 1. *Prior Probability*

*Prior probability* diambil dari 3 *tweet* yang terdiri dari 1 *tweet* positif, 1 *tweet* netral dan 1 *tweet* negatif. Pada tahap ini data *testing* akan dihitung nilai *prior probability* seperti pada Tabel 3.10 *Prior Probability*.

Tabel 3. 10 *Prior Probability*

<b>Label</b>	<i>Prior Probability</i>
Positif	1/3
Netral	1/3
Negatif	1/3

## 2. Conditional Probability

Pada tahap ini data *testing* akan dihitung nilai *conditional probability* seperti pada Tabel 3.11 *Conditional probability*.

Tabel 3. 11 *Conditional Probability*

Kata	Conditional probability		
	Positif	Netral	Negatif
Baru	0.0909090909090909	0.0526315789473684	0.0416666666666667
Karakter	0.0909090909090909	0.0526315789473684	0.0416666666666667
Ubah	0.0909090909090909	0.0526315789473684	0.0416666666666667
Sabar	0.0909090909090909	0.0526315789473684	0.0416666666666667
Main	0.0909090909090909	0.0526315789473684	0.0833333333333333
Yve	0.0454545454545455	0.0526315789473684	0.0833333333333333
Jadi	0.0454545454545455	0.0526315789473684	0.0833333333333333
Aneh	0.0454545454545455	0.0526315789473684	0.0833333333333333
Sekarang	0.0454545454545455	0.0526315789473684	0.0833333333333333
Lebih	0.0454545454545455	0.0526315789473684	0.0833333333333333
Suka	0.0454545454545455	0.0526315789473684	0.0833333333333333
Dulu	0.0454545454545455	0.0526315789473684	0.0833333333333333
Kualitas	0.0454545454545455	0.1052631578947368	0.0416666666666667
Main	0.0909090909090909	0.1052631578947368	0.0416666666666667
biasa	0.0454545454545455	0.1052631578947368	0.0416666666666667



### 3. Posterior Probability

Pada tahap ini akan dihitung nilai *posterior probability* seperti pada

Tabel 3.12 *Posterior probability*.

Tabel 3. 12 *Posterior Probability*

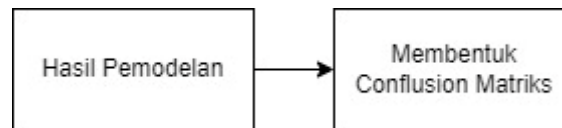
Label	<i>Posterior probability</i>
Positif	$P(\text{positif}) \times p(\text{baru}   \text{positif}) \times p(\text{main}   \text{positif})$ $= 0,3 \times 0.090 \times 0.090$ $= 0.000243$
Netral	$P(\text{netral}) \times p(\text{baru}   \text{netral}) \times p(\text{main}   \text{netral})$ $= 0,3 \times 0.05 \times 0.05$ $= 0.00075$
Negatif	$P(\text{negatif}) \times p(\text{baru}   \text{negatif}) \times p(\text{main}   \text{negatif})$ $= 0,3 \times 0.04 \times 0.08$ $= 0.00096$

berdasarkan tabel 3.10 nilai *posterior probability* tertinggi adalah milik label negatif dengan hasil 0.00096 sehingga data *testing* dapat dikatakan sebagai label negatif.

### 3.9 Pengukuran Performa

Setelah mendapatkan hasil klasifikasi dari menggunakan metode *Multinomial Naïve Bayes*, maka akan dilakukan pengujian akurasi, *presicion*, *recall* hasil klasifikasi. Pengujian akurasi, *presicion*, *recall*

menggunakan *confusion matriks*, yang digunakan untuk mengetahui *persentase* sistem dalam mengklasifikasi *tweet*. Tahap pengukuran performa dapat dilihat pada Gambar 3.9 *Diagram Confusion Matriks*.



Gambar 3. 9 *Diagram Confusion Matriks*

### 3.10 Skenario Pengujian

Dalam penelitian ini menerapkan metode *Multinomial Naïve Bayes* dalam melakukan analisis sentimen terkait *Developer Mobile Legends* di Indonesia. Langkah pertama melakukan *preprocessing* dengan tujuan untuk menghilangkan *noise* agar data yang tidak terstruktur menjadi lebih terstruktur. Tahap selanjutnya melakukan labeling terhadap *dataset* untuk memperoleh nilai sentiment dari *tweet* tersebut. Setelah data diberi labeling data akan dilakukan TFIDF untuk memberikan bobot pada setiap kata. Kemudian dilanjutkan dengan menyeimbangkan data dengan menggunakan SMOTE. Setelah data diseimbangkan dilakukan pemodelan dengan menggunakan *Multinomial Naïve Bayes* dan dilakukan validasi dengan menggunakan *K-Fold Cross Validation* untuk melakukan pengujian dengan cara melakukan variasi terhadap nilai *k* dapat dilihat pada Tabel 3.13 Pengujian *K-fold*.

Tabel 3. 13 Pengujian *K-fold*

Pengujian ke-	<i>K-fold</i>
1	3
2	5
3	7
4	9
5	10
6	11
7	12
8	13

### 3.11 Peralatan Penelitian

#### 1. Perangkat Lunak (Software)

Sistem Operasi : *Windows 10 Home Single Language 64-bit*

Bahasa Pemrograman : *Python (Google Colab)*

#### 2. Perangkat Keras (Hardware)

System Model : ASUS GL 503VD

Processor : INTEL(R) CORE(TM) i7-7700HQ

Harddisk : 1 TB

Memory : 16 GB

## BAB IV

### HASIL DAN ANALISIS

Pada penelitian ini digunakan data sebanyak 2223 yang terdiri dari 1003 data label positif, 806 data label netral dan 414 data label negatif. Langkah pertama yaitu melakukan *praproses* data dengan tujuan untuk menghilangkan *noise* agar data yang tidak terstruktur menjadi lebih terstruktur. Selanjutnya melakukan labeling pada *dataset* yang digunakan untuk memperoleh nilai *sentiment* dari *tweet*. Data yang telah dilabeling selanjutnya dilakukan *tf-idf* untuk memberikan bobot pada setiap kata.

Selanjutnya dilakukan *over-sampling* (SMOTE) untuk dilakukan *balancing* pada data agar proses pemodelan mendapatkan hasil yang maksimal. Selanjutnya akan dilakukan pembagian data menggunakan *Kfold* dengan jumlah *k* yang telah ditentukan. Kemudian setelah dilakukan pembagian data, dilakukan pemodelan dengan menggunakan *Multinomial Naïve Bayes*. Hasil pengujian ini kemudian dievaluasi dengan menggunakan *confusion matriks* untuk mengetahui metode *Multinomial Naïve Bayes* bekerja dengan baik dalam melakukan analisis sentimen terhadap *tweet Developer Mobile Legends* di Indonesia.

#### 4.1 Crawling

Pengambilan data diambil dengan menggunakan *library python* yaitu *snsrape*. Implementasi program dan hasil dapat dilihat pada Gambar 4.1 *Source Code Crawling* Data dan Tabel 4.1 Contoh Hasil *Crawling*.

```

for tweet in sntwitter.TwitterSearchScraper(query=query).get_items():
    if len(tweets) == limit:
        break
    else:
        tweets.append([tweet.date, tweet.user.username, tweet.content])

```

Gambar 4. 1 *Source Code Crawling*Tabel 4. 1 *Contoh Hasil Crawling*

No	<i>Tweet</i>
1	Jumat diajak nonton di MPL Arena langsung yak. Siapin death note buat nyatet nyatet ini mah
2	5 Hero Langgan Ban di Week 3 Regular Season MPL Season 10 https
3	@el_wjj eh u pada ngikutin MPL gk?
4	@mlbbfess seingetku ini tu pas release hero, lama bgt, mpl sisen 6-7 kalo gasalah inget
5	@islameyra @mlbbfess Bukan kak, room tourney tu macem room mpl gitu cmiiw
6	Galau bgt anjir ke mpl arena bawa motor apa pp apa staycation sekalian anjaay wkwkwk
7	@athnashield kita ke jkt aja yuu liat mpl terus ke pik with ayang 🥰👍
8	@yuzzhong Nonton bareng sih mpl
9	@FWBESS Outfit ke MPL Arena nih
10	"@bigetroneports BA mulu, itu MDL sama MPL bapak. Mau komen RA bosen. 😞"

## 4.2 Pre-processing

Proses *preprocessing* dilakukan menggunakan data dengan jumlah 2223 yang belum diolah dan diunduh dari *Google Colab* melalui proses *crawling* dengan format file .csv.

### 4.2.1 Case Folding

Pada tahap ini data akan diubah menjadi huruf kecil (*case folding*), menggunakan fungsi `series.str.lower()` pada *pandas*. Implementasi program dan hasil dapat dilihat pada Gambar 4.2 *Source Code Case Folding* dan Gambar 4.3 *Contoh Hasil Case Folding*.

```
#Case Folding
#menggunakan fungsi Series.str.lower() pada pandas
data_tweet['tweet'] = data_tweet['tweet'].str.lower()
```

Gambar 4. 2 Source Code Case Folding

Case Folding Result :

```
0      jumat diajak nonton di mpl arena langsung yak....
1      5 hero langganan ban di week 3 regular season ...
2      @el_wjj ehh u pada ngikutin mpl gk?
3      @mlbbfess seingetku ini tu pas release hero, l...
4      @islameyra @mlbbfess bukan kak, room tourney t...
...
2218   haruskah kita membantu nana untuk menemukan mo...
2219   apakah support damage masih laku di update bar...
2220   #mobilelegends esme mm,classic memang adalah t...
2221   jadi gini rasanya ke wc sambil main mobile leg...
2222   mobile legends ikut diblok kominfo gasi?
Name: tweet, Length: 2223, dtype: object
```

Gambar 4. 3 Contoh hasil Case Folding

#### 4.2.2 Tokenizing

Pada tahap ini akan memisahkan kalimat pada data menjadi beberapa bagian yang disebut *term* dan menghapus *tab*, *new line*, *ans back slice*, *non ASCII* (*emoticon*, *mention*, *link*, *URL*, *hastag*, dan *chinese word*), angka, tanda baca, spasi kosong, *multiple whitespace* ke *single whitespace*, dan *single char*. Implementasi program dan hasil dapat dilihat pada Gambar 4.4 Source Code Tokenizing dan Gambar 4.5 Contoh Hasil Tokenizing.

```
#Tokenizing
def remove_tweet_special(tweet):
    #Menghapus tab, new line, ans back slice, non ASCII(emoticon, mention, link, URL, hastag, chinese word)
    tweet = tweet.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\ ', " ")
    tweet = tweet.encode('ascii', 'replace').decode('ascii')
    tweet = ' '.join(re.sub("([@#][A-Za-z0-9+])|(\w+:\w+/\w+)", " ", tweet).split())
    return tweet.replace("http://", " ").replace("https://", " ")

#Menghapus Angka
def remove_number(tweet):
    return re.sub(r"\d+", "", tweet)
```

```

#Menghapus tanda baca
def remove_punctuation(tweet):
    return tweet.translate(str.maketrans("", "", string.punctuation))

#Menghapus multiple whitespace ke single whitespace
def remove_whitespace_multiple(tweet):
    return re.sub('\s+', ' ', tweet)

#menghapus single char
def remove_single_char(tweet):
    return re.sub(r"\b[a-zA-Z]\b", "", tweet)

#NLTK word tokenize
def word_tokenize_wrapper(tweet):
    return word_tokenize(tweet)

```

Gambar 4. 4 Source Code Tokenizing

Tokenizing Result :

```

0      [jumat, diajak, nonton, di, mpl, arena, langsu...
1      [hero, langganan, ban, di, week, regular, seas...
2      [wjj, eh, pada, ngikutin, mpl, gk]
3      [seingetku, ini, tu, pas, release, hero, lama,...
4      [bukan, kak, room, tourney, tu, macem, room, m...

2218   [haruskah, kita, membantu, nana, untuk, menemu...
2219   [apakah, support, damage, masih, laku, di, upd...
2220   [esme, mmclassic, memang, adalah, tempat, yg, ...
2221   [jadi, gini, rasanya, ke, wc, sambil, main, mo...
2222   [mobile, legends, ikut, diblok, kominfo, gasi]
Name: token_tweets, Length: 2223, dtype: object

```

Gambar 4. 5 Contoh Hasil Tokenizing

### 4.2.3 Normalisasi

Pada tahap ini, kalimat pada data yang sudah menjadi *term* akan dilakukan pengecekan untuk mengubah kata tidak baku dan kata singkat menjadi kata baku Bahasa Indonesia. Kamus yang digunakan pada tahap normalisasi ini merupakan kamus yang dibuat oleh penulis penelitian ini yang sudah disesuaikan berdasarkan dari data yang digunakan pada penelitian ini menurut Kamus Besar Bahasa Indonesia. Implementasi program dan hasil dapat dilihat pada Gambar 4.6 Source Code Normalisasi dan Gambar 4.7 Contoh Hasil Normalisasi.

```

for index, row in normalized_word.iterrows():
    if row[0] not in normalized_word_dict:
        normalized_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalized_word_dict[term] if term in normalized_word_dict else term for term in document]

```

Gambar 4. 6 *Source Code* Normalisasi

```

Normalisasi Result:
0      [jumat, diajak, nonton, di, mpl, arena, langsu...
1      [karakter, langganan, melarang, di, week, regu...
2      [wjj, eh, pada, ikut, mpl, tidak]
3      [ingat, ini, itu, pas, melepaskan, karakter, l...
4      [bukan, kakak, ruang, turnamen, itu, macam, ru...

...
2218   [haruskah, kita, membantu, nana, untuk, menemu...
2219   [apa, mendukung, rusak, masih, laku, di, mempe...
2220   [esme, mmclassic, memang, adalah, tempat, yang...
2221   [jadi, gini, rasanya, ke, wc, sambil, main, mo...
2222   [mobile, legends, ikut, blokir, kominfo, tidak]
Name: tweet_normalized, Length: 2223, dtype: object

```

Gambar 4. 7 Contoh Hasil Normalisasi

#### 4.2.4 Stopword Removal

Pada tahap ini, setelah melewati tahap normalisasi akan dilakukan penghilangan kata yang tidak memiliki arti, menggunakan *library nltk.corpus*. Kamus *Stopword Removal* yang digunakan pada penelitian ini merupakan kamus yang dibuat oleh penulis penelitian ini yang sudah disesuaikan berdasarkan kata pada data penelitian ini. Implementasi program dan hasil dapat dilihat pada Gambar 4.8 *Source Code Stopword Removal* dan Gambar 4.9 Contoh Hasil *Stopword Removal*.

```

# get stopwords indonesia
list_stopwords = stopwords.words('indonesian')

def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

```

Gambar 4. 8 *Source Code Stopword Removal*



StopWord Result :

```

0      [jumat, diajak, nonton, mpl, arena, langsung, ...
1      [karakter, langganan, melarang, reguler, mpl]
2      [mpl]
3      [melepaskan, karakter, mpl, musim, tidak salah]
4      [kakak, ruang, turnamen, ruang, mpl, gitu]
...
2218   [haruskah, membantu, nana, menemukan, molina, ...
2219   [mendukung, rusak, laku, memperbarui, teman]
2220   [esme, indah]
2221   [main, mobile, legends]
2222   [mobile, legends, blokir]
Name: tweet_tokens_WSW, Length: 2223, dtype: object

```

Gambar 4. 9 Contoh Hasil *Stopword Removal*

#### 4.2.5 Stemming

Pada tahap ini, setelah *stopword removal* akan dilakukan perubahan kata menjadi kata dasar yang mudah dipahami sesuai kamus *stemming*, menggunakan *library Sastrawi.stemmer.StemmerFactory*. Implementasi program dan hasil dapat dilihat pada Gambar 4.10 *Source Code Stemming* dan Gambar 4.11 Contoh Hasil *Stemming*.

```

stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return stemmer.stem(term)
term_dict = {}

for document in data_tweet['tweet_tokens_WSW']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ' '

```

Gambar 4. 10 *Source Code Stemming*

```

Stemming Result:
0      [jumat, ajak, nonton, mpl, arena, langsung, ma...
1      [karakter, langgan, larang, reguler, mpl]
2      [mpl]
3      [lepas, karakter, mpl, musim, tidak salah]
4      [kakak, ruang, turnamen, ruang, mpl, gitu]
...
2218   [harus, bantu, nana, temu, molina, anchang, tsu...
2219   [dukung, rusak, laku, baru, teman]
2220   [esme, indah]
2221   [main, mobile, legends]
2222   [mobile, legends, blokir]
Name: tweet_tokens_stemmed, Length: 2223, dtype: object

```

Gambar 4. 11 Contoh Hasil *Stemming*

### 4.3 Labeling Data

Sebelum dilakukan proses labeling data akan diterjemahkan kedalam bahasa Inggris terlebih dahulu, menggunakan *google translate*. Labeling data menggunakan bahasa pemrograman *python* yaitu *vader* yang terdapat dalam *library* NLTK. Implementasi program dan hasil dapat dilihat pada Gambar 4.12 *Source Code* Labeling Data dan Gambar 4.13 Contoh Hasil Labeling Data.

```

doc['compound'] = doc['scores'].apply(lambda score_dict: score_dict['compound'])
doc['sentimen_type'] = ''
print('')
doc.loc[doc.compound>0, 'sentyimen_type'] = 'POSITIVE'
doc.loc[doc.compound<0, 'sentyimen_type'] = 'NEGATIVE'
doc.loc[doc.compound==0, 'sentyimen_type'] = 'NEUTRAL'

```

Gambar 4. 12 *Source Code* Labeling Data

		scores	compound
0	{'neg': 0.178, 'neu': 0.822, 'pos': 0.0, 'comp...		-0.5994
1	{'neg': 0.198, 'neu': 0.604, 'pos': 0.198, 'co...		0.0000
2	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...		0.0000
3	{'neg': 0.0, 'neu': 0.749, 'pos': 0.251, 'comp...		0.7479
4	{'neg': 0.121, 'neu': 0.604, 'pos': 0.275, 'co...		0.4215
...		...	...
2218	{'neg': 0.0, 'neu': 0.847, 'pos': 0.153, 'comp...		0.4019
2219	{'neg': 0.176, 'neu': 0.691, 'pos': 0.133, 'co...		-0.2168
2220	{'neg': 0.0, 'neu': 0.698, 'pos': 0.302, 'comp...		0.5994
2221	{'neg': 0.0, 'neu': 0.893, 'pos': 0.107, 'comp...		0.2023
2222	{'neg': 0.296, 'neu': 0.704, 'pos': 0.0, 'comp...		-0.2732

Gambar 4. 13 Contoh Hasil Labeling Data

#### 4.4 Pembobotan Kata

Setelah melewati tahap labeling data akan dilakukan pembobotan kata pada data, nilai bobot didapatkan dari hasil perhitungan kemunculan setiap kata lalu dikalikan dengan *idf*, menggunakan *library sklearn.feature\_extraction.text*. Implementasi program dan hasil dapat dilihat pada Gambar 4.14 *Source Code* TF-IDF dan Gambar 4.15 Contoh Hasil TF-IDF.

```
#TFIDF
vektorizer = TfidfVectorizer()
response = vektorizer.fit_transform(data_tweet4['tweet'])
```

Gambar 4. 14 *Source Code* TF-IDF

TF-IDF Result:

(0, 1469)	0.7301156411833473
(0, 309)	0.34658767211591457
(0, 1263)	0.2797916326182347
(0, 1100)	0.19875941808409872
(0, 103)	0.22425159111169454
(0, 1350)	0.07434701211682057
(0, 1451)	0.14584842930480874
(0, 26)	0.28343781737141854
(0, 861)	0.2675714056146797
(1, 1732)	0.5827757135252473
(1, 1108)	0.44430793183064515
(1, 1095)	0.5678055539014415
(1, 901)	0.3501446590594794
(1, 1350)	0.13400932851227834

Gambar 4. 15 Contoh Hasil TF-IDF

#### 4.5 SMOTE

Pada tahap ini, setelah TF-IDF akan dilakukan penyeimbangan data terhadap kelas minoritas sampai sama dengan kelas mayoritas dengan menduplikasi secara acak sampel kelas minoritas menggunakan *library imblearn.over\_sampling*. Implementasi program dan hasil dapat dilihat pada Gambar 4.16 *Source Code SMOTE* dan Gambar 4.17 Contoh Hasil SMOTE.

```
sm = SMOTE(random_state=2)
xSmote, ySmote= sm.fit_resample(x,y.ravel())
```

Gambar 4. 16 *Source Code SMOTE*

```
Sebelum OverSampling, count of label 'POSITIVE': 1197
Sebelum OverSampling, count of label 'NEUTRAL': 537
Sebelum OverSampling, count of label 'NEGATIVE': 489
```

```
Setelah OverSampling, the shape of train_x: (3591, 2324)
Setelah OverSampling, the shape of train_y: (3591,)
```

```
Setelah OverSampling, count of label 'POSITIVE': 1197
Setelah OverSampling, count of label 'NEUTRAL': 1197
Setelah OverSampling, count of label 'NEGATIVE': 1197
```

Gambar 4. 17 Contoh hasil SMOTE

#### 4.6 K-fold Cross Validation

Tahap ini dilakukan menggunakan model *Kfold* dari *library sklearn.model\_selection*. Pembagian data sesuai dengan jumlah *k* pada skenario pengujian yang ditentukan untuk mendapatkan model terbaik. Implementasi program dapat dilihat pada Gambar 4.18 *Source Code K-fold*.

```
kfold = KFold(k, random_state = 0, shuffle = True)
for train_index, test_index in kfold.split(xSmote):
    x_train, x_test = xSmote[train_index], xSmote[test_index]
    y_train, y_test = ySmote[train_index], ySmote[test_index]
```

Gambar 4. 18 *Source Code K-fold*

#### 4.7 Multinomial Naïve Bayes

Setelah ditentukan pembagian data sesuai dengan jumlah *k* pada skenario pengujian, maka dilakukan klasifikasi menggunakan metode *Multinomial Naïve Bayes*, *library* yang digunakan *sklearn.naive\_bayes*. Implementasi program dapat dilihat pada Gambar 4.19 *Source Code Multinomial Naïve Bayes*.

```
model = MultinomialNB().fit(x_train, y_train)
predik = model.predict(x_test)
```

Gambar 4. 19 *Source Code Multinomial Naïve Bayes*

#### 4.8 Confusion Matrix

Setelah dilakukan proses klasifikasi, maka dilakukan pengujian akurasi, *precision* dan *recall*. Pengujian akurasi, *precision* dan *recall* dilakukan menggunakan metode *confusion matrix* dari *library sklearn.metrics*. Implementasi program dapat dilihat pada Gambar 4.20 *Source Code Confusion Matrix*.

```

acc = accuracy_score(y_test, predik)*100
pre = precision_score(y_test, predik, average = 'macro')*100
rec = recall_score(y_test, predik, average = 'macro')*100
con_matrix = confusion_matrix(y_test, model.predict(x_test))

```

Gambar 4. 20 Source Code Confusion Matrix

#### 4.9 Pengujian 12-fold Cross Validation

Pengujian dengan *k-fold cross validation* dilakukan dengan tujuan agar mengetahui pengaruh akurasi metode *Multinomial Naïve Bayes* yang dihasilkan oleh sistem. Dengan membagi nilai *k* sebanyak *12-fold*, dataset awal berjumlah 2223 setelah dilakukan *balancing dataset* berjumlah 3009 akan dibagi menjadi 12 bagian diinisialisasi dengan nama K1, K2, K3, K4, K5, K6, K7, K8, K9, K10, K11, K12 dengan masing dataset 251 dan 250. Pada *fold* awal akan dikombinasi dengan 11 bagian data berbeda yang digabungkan serta dijadikan sebagai data latih, untuk sisanya dijadikan sebagai data uji. Proses pelatihan dan pengujian dilakukan sampai *fold* kesepuluh. Hasil pengujian dapat dilihat pada Tabel 4.2 Pengujian *12-fold*.

Tabel 4. 2 Pengujian *12-fold*

Iterasi <i>Fold</i> <i>ke</i>	Data Latih	Data Uji	Rata-rata tiap Pengujian		
			<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
1	K2, K3, K4, K5, K6, K7, K8, K9, K10, K11, K12	K1	84.46%	84.45%	84.54%
2	K1, K3, K4, K5, K6, K7, K8, K9,	K2	82.47%	82.77%	82.26%

	K10, K11, K12				
3	K1, K2, K4, K5, K6, K7, K8, K9, K10, K11, K12	K3	79.28%	79.24%	79.67%
4	K1, K2, K3, K5, K6, K7, K8, K9, K10, K11, K12	K4	76.89%	76.76%	77.08%
5	K1, K2, K3, K4, K6, K7, K8, K9, K10, K11, K12	K5	75.29%	75.96%	76.04%
6	K1, K2, K3, K4, K5, K7, K8, K9, K10, K11, K12	K6	82.07%	82.61%	81.90%
7	K1, K2, K3, K4, K5, K6, K8, K9, K10, K11, K12	K7	81.27%	80.94%	80.48%
8	K1, K2, K3, K4, K5, K6, K7, K9, K10, K11, K12	K8	83.26%	83.47%	83.58%
9	K1, K2, K3, K4, K5, K6, K7, K8, K10, K11, K12	K9	82.47%	81.81%	81.48%
10	K1, K2, K3, K4, K5, K6, K7, K8,	K10	78.4%	79.53%	78.37%

	K9, K11, K12				
11	K1, K2, K3, K4, K5, K6, K7, K8, K9, K10, K12	K11	82.39%	83.17%	82.98%
12	K1, K2, K3, K4, K5, K6, K7, K8, K9, K10, K11	K12	81.6%	82.16%	81.70%
Nilai Rata-Rata keseluruhan <i>fold</i>			80.82%	81.07%	80.84%

#### 4.9.1 Iterasi fold ke-1

Pengujian iterasi *fold* ke-1 dengan menggunakan 251 data uji (data ke 1 sampai data ke 251). Sistem mendapat data yang benar sebanyak 70 data positif, 63 data netral, dan 79 data negatif. Total yang berhasil dilabeling sistem dengan benar sebanyak 212 data. Ditemukan hasil sistem yang gagal dalam melabeli data yang seharusnya berlabel positif namun sistem melabeli sebagai label netral 9 data, dan label negatif 11 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun sistem melabeli sebagai label positif sebanyak 8 data, dan label negatif sebanyak 4 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun sistem melabeli sebagai label positif sebanyak 2 data, dan label netral sebanyak 5 data. Didapatkan evaluasi *accuracy* 84.46%, dengan kemampuan sistem dalam menemukan semua dokumen yang relevan 84.54% *recall*, dan *precision* 84.45% untuk tingkat ketepatan inputan antara informasi yang diinputkan dengan hasil klasifikasi.



#### 4.9.2 Iterasi fold ke-2

Pengujian iterasi *fold* ke-2 dengan menggunakan 251 data uji (data ke 252 sampai data ke 502). Sistem mendapat data yang benar sebanyak 60 data positif, 69 data netral, dan 78 data negatif. Total yang dilabeling sistem dengan benar sebanyak 207 data. Ditemukan hasil sistem yang gagal dalam melabeli dengan benar yang seharusnya berlabel positif namun sistem melabeli sebagai label netral sebanyak 7 data, dan berlabel negatif sebanyak 12 data. Sistem juga gagal melabeli yang seharusnya berlabel netral namun sistem melabeli sebagai label positif sebanyak 7 data, dan berlabel negatif sebanyak 9 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun sistem melabeli sebagai positif sebanyak 6 data, dan label netral sebanyak 3 data. Didapatkan evaluasi *accuracy* 82.47%, *recall* 82.26% dalam menemukan semua dokumen yang relevan, dan *precision* 82.77% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.9.3 Iterasi fold ke-3

Pengujian iterasi *fold* ke-3 dengan menggunakan 251 data uji (data ke 503 sampai data ke 753). Sistem mendapat data yang benar sebanyak 72 data positif, 51 data netral, dan 76 data negatif. Total yang dilabeli sistem dengan benar sebanyak 199 data. Ditemukan hasil sistem yang gagal dalam melabeli dengan benar yang seharusnya berlabel positif namun sistem melabeli sebagai label netral sebanyak 14 data, dan berlabel negatif sebanyak 15 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun sistem melabeli sebagai positif sebanyak 6 data, dan berlabel

negatif sebanyak 9 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun sistem melabeli sebagai positif sebanyak 3 data, dan berlabel netral sebanyak 5 data. Didapatkan evaluasi *accuracy* 79.28%, *recall* 79.67% dalam menemukan semua dokumen yang relevan, dan *precision* 79.24% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.9.4 Iterasi fold ke-4

Pengujian iterasi *fold* ke-4 dengan menggunakan 251 data uji (data ke 754 sampai dengan 1004). Sistem mendapat data yang benar sebanyak 57 data positif, 76 data netral, dan 60 data negatif. Total yang dilabeli sistem dengan benar sebanyak 193 data. Ditemukan hasil sistem yang gagal dalam melabeli dengan benar yang seharusnya berlabel positif namun sistem melabeli sebagai label netral sebanyak 16 data, dan label negatif sebanyak 8 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun sistem melabeli sebagai label positif sebanyak 13 data, dan label negatif sebanyak 9 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun sistem melabeli sebagai positif sebanyak 7 data, dan label netral sebanyak 5 data. Didapatkan evaluasi *accuracy* 76.89%, *recall* 77.08% dalam menemukan semua dokumen yang relevan, dan *precision* 76.76% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.9.5 Iterasi fold ke-5

Pengujian iterasi *fold* ke-5 menggunakan 251 data uji (data ke 1005 sampai dengan 1255). Sistem mendapat data yang benar sebanyak 61 data positif, 61 data netral, dan 67 data negatif. Total data yang dilabeli oleh sistem yang benar sebanyak 189 data. Ditemukan juga hasil sistem yang gagal melabeli dengan benar yang seharusnya label positif namun dilabeli sebagai netral sebanyak 13 data, dan dilabeli sebagai negatif sebanyak 17 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabel sebagai positif sebanyak 10 data, dan dilabeli sebagai negatif sebanyak 14 data. Dan sistem juga gagal melabeli data yang seharusnya berlabel negatif namun sistem melabeli sebagai positif sebanyak 4 data, dan label netral sebanyak 4 data. Didapatkan evaluasi *accuracy* 75.29%, *recall* 76.04% dalam menemukan semua dokumen yang relevan, dan *precision* 75.96% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.9.6 Iterasi fold ke-6

Pengujian iterasi *fold* ke-6 menggunakan 251 data uji (data ke 1256 sampai dengan 1506). Sistem mendapatkan data yang benar sebanyak 54 data positif, 69 data netral, dan 83 data negatif. Total data yang dilabeli oleh sistem dengan benar sebanyak 206 data. Ditemukan juga hasil sistem yang gagal melabeli dengan benar yang seharusnya berlabel positif namun dilabeli sebagai label netral sebanyak 6 data, dan label negatif sebanyak 9 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai label positif sebanyak 9 data, dan label negatif

sebanyak 15 data. Sistem juga gagal melabeli data yang seharusnya negatif namun sistem melabeli sebagai label positif sebanyak 3 data, dan berlabel netral sebanyak 3 data. Didapatkan evaluasi *accuracy* 82.07%, *recall* 81.90% dalam menemukan semua dokumen yang relevan, dan *precision* 82.61% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.9.7 Iterasi fold ke-7

Pengujian iterasi *fold* ke-7 menggunakan 251 data uji (data ke 1507 sampai dengan 1757). Sistem mendapatkan data dengan benar sebanyak 64 positif, 52 netral, dan 88 negatif. Total data yang dilabeli sistem dengan benar sebanyak 205 data. Ditemukan juga hasil sistem yang gagal dalam melabeli dengan benar yang seharusnya berlabel positif namun dilabeli sebagai label netral sebanyak 9 data, dan dilabel negatif sebanyak 13 data. Sistem juga gagal melabeli data yang seharusnya netral namun sistem melabeli sebagai label positif sebanyak 13 data, dan label negatif sebanyak 4 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai label positif sebanyak 4 data, dan label netral sebanyak 4 data. Didapatkan *accuracy* 81.27%, *recall* 80.48%, dan *precision* 80.94%.

#### 4.9.8 Iterasi fold ke-8

Pengujian iterasi *fold* k-8 menggunakan 251 data uji (data ke 1758 sampa dengan 2008). Sistem mendapatkan data dengan benar sebanyak 63 data positif, 75 data netral, dan 71 data negatif. Total data yang dilabeli sistem dengan benar sebanyak 209 data. Ditemukan juga hasil sistem yang

gagal melabeli data dengan benar yang seharusnya berlabel positif namun dilabeli sebagai label netral sebanyak 7 data, dan label negatif sebanyak 8 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai label positif sebanyak 7 data, dan label negatif sebanyak 13 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai label positif sebanyak 6 data, dan label netral sebanyak 1 data. Didapatkan *accuracy* 83.26%, *recall* 83.58%, dan *precision* 83.47%.

#### 4.9.9 Iterasi fold ke-9

Pengujian iterasi *fold* ke-9 menggunakan 251 data uji (data ke 2009 sampai dengan 2259). Sistem mendapatkan data yang benar sebanyak 49 data positif, 75 data netral, dan 83 data negatif. Total data yang dilabeli oleh sistem dengan benar sebanyak 207 data. Ditemukan juga hasil sistem yang gagal melabeli data dengan benar yang seharusnya berlabel positif namun sistem melabeli sebagai label netral sebanyak 12 data, dan label negatif sebanyak 10 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai positif sebanyak 10 data, dan label negatif sebanyak 3 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai label positif sebanyak 5 data, dan netral sebanyak 4 data. Didapatkan *accuracy* 82.47%, *recall* 81.48%, dan *precision* 81.81%.

#### 4.9.10 Iterasi fold ke-10

Pengujian iterasi *fold* ke-10 menggunakan 250 data uji (data ke 2260 sampai dengan 2509). Sistem mendapat data yang benar sebanyak 55 data positif, 65 data netral, dan 76 data negatif. Total data yang dilabeli oleh sistem dengan benar sebanyak 196 data. Ditemukan juga hasil sistem yang gagal dalam melabeli data yang seharusnya data berlabel positif namun dilabeli sebagai label netral sebanyak 10 data, dan label negatif sebanyak 19 data. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai label positif sebanyak 5 data, dan label negatif sebanyak 11 data. Dan sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai label positif sebanyak 5 data, dan label netral sebanyak 4 data. Didapatkan *accuracy* 78.4%, *recall* 78.37%, dan *precision* 79.53%.

#### 4.9.11 Iterasi fold ke-11

Pengujian iterasi *fold* ke-11 menggunakan 250 data uji (data ke 2510 sampai dengan 2759). Sistem mendapat data yang benar sebanyak 61 data positif, 73 data netral, dan 72 data negatif. Total data yang dilabeli oleh sistem dengan benar sebanyak 206 data. Ditemukan juga hasil sistem yang gagal dalam melabeli data yang seharusnya berlabel positif namun dilabeli sebagai netral sebanyak 14 data, dan sebanyak 14 data negatif. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai positif sebanyak 6 data, dan negatif sebanyak 5 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai netral sebanyak 5 data. Didapatkan evaluasi *accuracy* 82.39%, *recall*

82.98% dalam menemukan semua dokumen yang relevan, dan *precision* 83.17% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

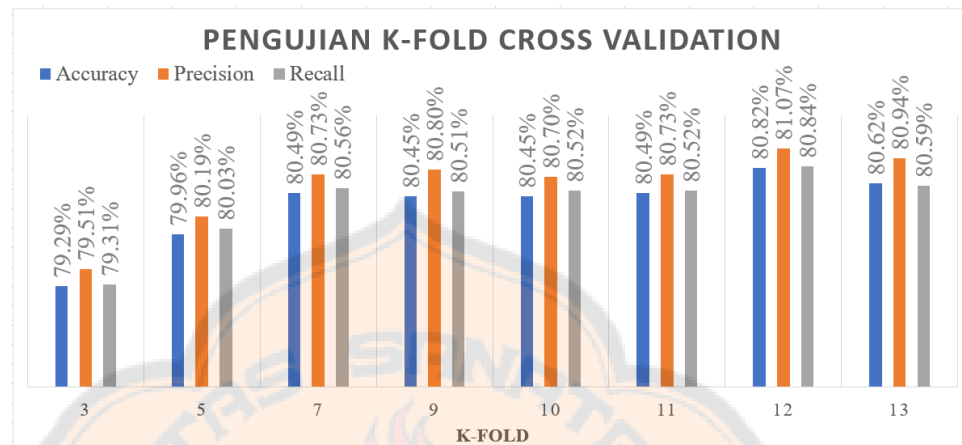
#### 4.9.12 Iterasi fold ke-12

Pengujian iterasi *fold ke-12* menggunakan 250 data uji (data ke 2760 sampai dengan 3009). Sistem mendapat data yang benar sebanyak 57 data positif, 69 data netral, dan 78 data negatif. Total data yang dilabeli oleh sistem dengan benar sebanyak 204 data. Ditemukan juga hasil sistem yang gagal dalam melabeli data yang seharusnya berlabel positif namun dilabeli sebagai netral sebanyak 12 data, dan sebanyak 15 data negatif. Sistem juga gagal melabeli data yang seharusnya berlabel netral namun dilabeli sebagai positif sebanyak 8 data, dan negatif sebanyak 7 data. Sistem juga gagal melabeli data yang seharusnya berlabel negatif namun dilabeli sebagai positif sebanyak 1 data, dan netral sebanyak 3 data. Didapatkan evaluasi *accuracy* 81.6%, *recall* 80.84% dalam menemukan semua dokumen yang relevan, dan *precision* 81.07% untuk tingkat ketepatan input antara informasi yang diinputkan dengan hasil klasifikasi.

#### 4.10 Analisa Hasil

Hasil pengujian K-Fold Cross Validation dapat dilihat pada Gambar 4.21

Grafik Pengujian *K-fold*.



Gambar 4. 21 Grafik Pengujian *K-fold*

Nilai optimal didapatkan pada percobaan *12-fold*. Pada percobaan *12-fold* diperoleh akurasi tertinggi pada iterasi *fold ke-1* sebesar 84.46%, presisi 84.45%, dan *recall* 84.54%. Diperoleh *tweet* yang terklasifikasi dengan benar *True Negatif* sebanyak 911, *True Netral* sebanyak 798, *True Positif* sebanyak 723. Diperoleh juga *tweet* yang gagal terklasifikasi dengan benar *False Negatif* sebanyak 254, *False Netral* sebanyak 175, *False Positif* sebanyak 148. *Confusion matrix* menghasilkan tingkat *accuracy* rata-rata sebesar 80.82%, rata-rata *precision* 81.07%, dan rata-rata *recall* 80.84%. Dengan 3009 data *tweet* yang diuji, terdapat 2432 data *tweet* yang benar diklasifikasikan, dan terdapat 577 data *tweet* yang gagal dikalsifikasi dengan benar oleh *Multinomial Naïve Bayes*.

Hasil ini dipengaruhi oleh nilai variasi *k-fold* pada model *Multinomial Naïve Bayes*. Dengan bervariasinya *k-fold* dalam proses



pembagian data, maka hasil yang diperoleh pun bervariasi. Hasil ini berdasar pada teori *Multinomial Naïve Bayes* yang merupakan model *probabilitas* turunan dari teori *bayes* yang dapat memberikan hasil yang baik dalam melakukan klasifikasi pada teks. Dimana model *Multinomial Naïve Bayes* mengasumsikan dimana setiap data harus memiliki label agar dapat dilakukan pelatihan. Model ini menghitung *probabilitas* dari setiap kata yang muncul pada setiap label.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

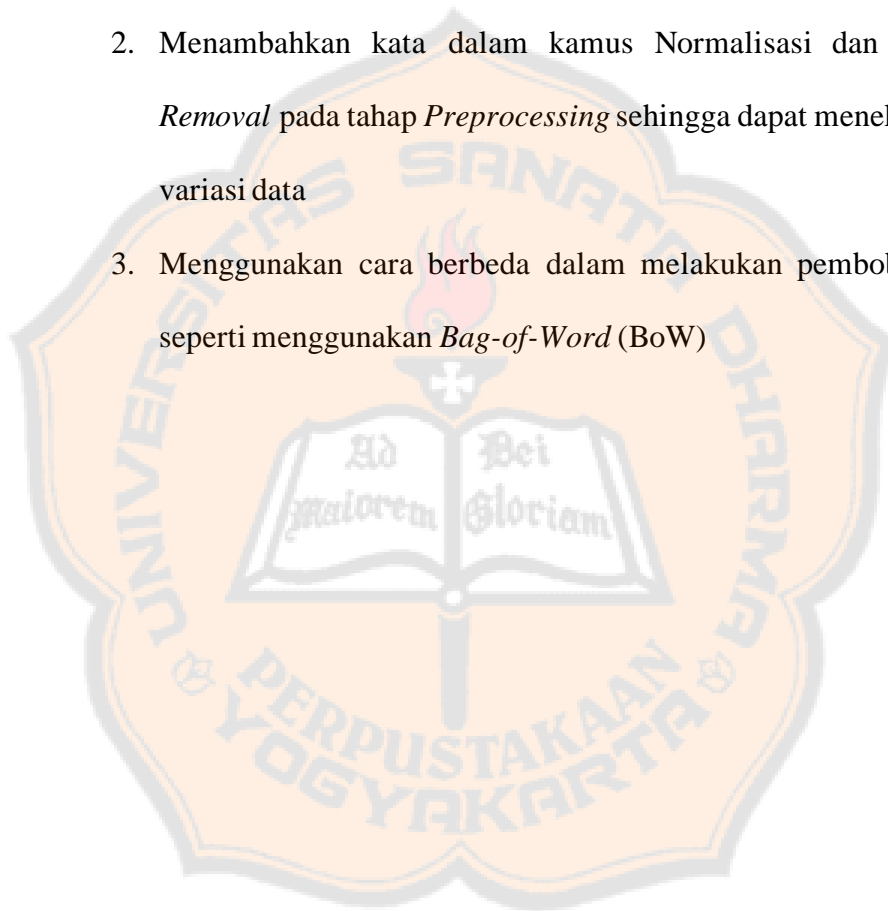
Berdasarkan hasil penelitian Algoritma *Multinomial Naïve Bayes* menggunakan *K-fold cross Validation* dengan melakukan beberapa variasi *k-fold* dalam pengujian dengan nilai  $k = 3, 5, 7, 9, 10, 11, 12, 13$  diperoleh akurasi sistem dari setiap variasi *k-fold* yang diuji. Berdasarkan hasil pengujian menggunakan *K-Fold Cross Validation* diperoleh model dengan akurasi tertinggi dari variasi pengujian *12-fold* menghasilkan rata-rata akurasi sebesar 80.82%, *recall* sebesar 80.84% untuk kemampuan sistem dalam menemukan semua dokumen yang relevan, dan *precision* sebesar 81.07% untuk tingkat ketepatan inputan antara informasi yang diinputkan dengan hasil klasifikasi. Hasil tersebut didapatkan dengan menggunakan *dataset* awal yang diperoleh dari Twitter sebanyak 4000 data kemudian dilakukan pembersihan tahap awal dengan menghilangkan data yang tidak berkaitan dengan kata kunci, sehingga data menjadi 2223 *tweet* terdiri dari 1003 *tweet* positif, 806 *tweet* netral, dan 414 *tweet* negatif, dari hasil pelabelan tersebut sehingga data *tweet* menjadi tidak seimbang kemudian dilakukan penyeimbangan data menggunakan SMOTE membuat *dataset* menjadi 3009 *tweet* yang terdiri dari 1003 *tweet* positif, 1003 *tweet* netral, dan 1003 *tweet* negatif. Pada pengujian *12-fold* didapatkan sebanyak 911 *tweet* diklasifikasi negatif dengan benar, 798 *tweet* yang diklasifikasi netral dengan benar, dan 723 *tweet* yang diklasifikasi positif dengan benar. Sementara sebanyak 255 *tweet* gagal diklasifikasi dengan benar negatif,

sebanyak 175 *tweet* gagal diklasifikasi netral dengan benar, dan sebanyak 148 *tweet* gagal diklasifikasi positif dengan benar.

## 5.2 Saran

Saran yang dapat diberikan dari hasil penelitian ini adalah:

1. Menggunakan cara berbeda dalam melakukan labeling data seperti menggunakan LSTM (*Long Short-Term Memory*)
2. Menambahkan kata dalam kamus Normalisasi dan *Stopword Removal* pada tahap *Preprocessing* sehingga dapat menekan jumlah variasi data
3. Menggunakan cara berbeda dalam melakukan pembobotan data seperti menggunakan *Bag-of-Word* (BoW)



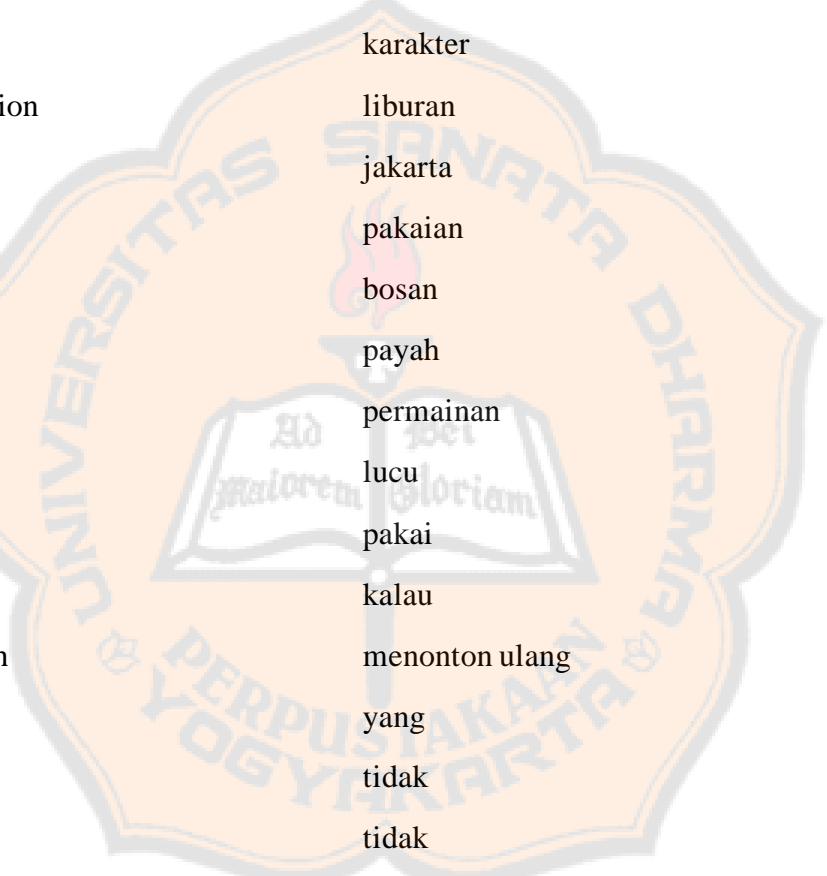
## DAFTAR PUSTAKA

- Amira Sumitro, P., Iskandar Mulyana, D., & Saputro, W. (2021). Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. *Jurnal Informatika Dan Teknologi Komputer*, 2(2), 50–56.
- Anis Nikmatul Kasanah, Muladi, & Pujianto Utomo. (2017). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *JURNAL RESTI*, 1(3), 196–201.
- Asri, Y., Suliyanti, W. N., Kuswardani, D., & Fajri, M. (2022). Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile. *PETIR:Jurnal Pengkajian Dan Penerapan Teknik Informatika*, 15(2), 264–275.
- Bagus, I., Widnyana, G., Klasifikasi, P. :, Bahasa..., T., Gede, I. B., Putra, W., Nyoman, I., & Kumara, S. (2016). KLASIFIKASI TEKS BAHASA BALI DENGAN METODE SUPERVISED LEARNING NAIVE BAYES CLASSIFIER. *Teknologi Elektro*, 15(2).
- Bunga, M. T. H., Djahi, B. S., Kom, S., Nabuasa, Y. Y., Cs, M., Komputer, J. I., Sains, F., Teknik, D., & Cendana, U. N. (2018). MULTINOMIAL NAIVE BAYES UNTUK KLASIFIKASI STATUS KREDIT MITRA BINAAN DI PT. ANGKASA PURA I PROGRAM KEMITRAAN. *J-ICON*, 6(2), 30–34.
- Damanik, F. J., & Setyohadi, D. B. (2021). Analysis of public sentiment about COVID-19 in Indonesia on Twitter using multinomial naive bayes and support vector machine. *IOP Conference Series: Earth and Environmental Science*, 704(1).
- Deviyanto, A., Didik Wahyudi, M. R., & Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, T. (2018). PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *Jurnal Informatika Sunan Kalijaga*, 3(1), 1–13.
- Erwin Julius, S. P. H. P. E. D. (2016). Perancangan Interior Fasilitas E-SportsArena. *JURNAL INTRA* , 4(2), 672–681.
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). *Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking* (Vol. 2, Issue 8).
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. *Jurnal Teknoinfo*, 14(2), 115-224X.

- Gunawan Billy, Pratiwi Sasty Hele, Esyudha Enda, & Pratama Enda Esyudha Pratama. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 4(2), 17–29.
- Mujib Ridwan, H. S. dan M. S. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 7.
- Putri Nirwandani, E., & Cahya Wihandika, R. (2021). *Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naïve Bayes* (Vol. 5, Issue 3).
- Rhomadhona, H., & Permadi, J. (2019). Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation. *Jurnal Sains Dan Informatika*, 5(2), 108–117.
- Setiawan, R. A., & Setyohadi, D. B. (2017). Analisis Komunikasi Sosial Media Twitter sebagai Saluran Layanan Pelanggan Provider Internet dan Seluler di Indonesia. *Journal of Information Systems Engineering and Business Intelligence*, 3(1).
- Valentina, E., & Purnama Sari Elvivalentina, W. (2019). *Studi Komunikasi Verbal dan Non Verbal Game Mobile Legends: Bang Bang*.
- Wardani, N. S., Prahutama, A., & Kartikasari, P. (2020). ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA DENGAN KLASIFIKASI NAÏVE BAYES UNTUK MODEL BERNOULLI DAN MULTINOMIAL. *JURNAL GAUSSIAN*, 9, 237–249.
- Winahyu, J., & Suharjo, I. (2021). Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, 10(2), 206–2014.
- Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 422–430

## LAMPIRAN

Daftar kata kamus normalisasi yang digunakan, dibuat oleh penulis berdasarkan data yang digunakan dalam penelitian. Kamus normalisasi dapat dilihat dibawah ini:



season	musim
ngikutin	ikut
hero	karakter
staycation	liburan
jkt	jakarta
outfit	pakaian
bosen	bosan
bapuk	payah
game	permainan
ngakak	lucu
pake	pakai
kalo	kalau
rewatch	menonton ulang
yg	yang
gk	tidak
gak	tidak
harusterima	harus terima
samaaa	sama
drmn	dari mana
klo	kalau
jg	juga
tp	tapi
for	untuk

bs	bisa
kuuu	aku
it	itu
dapet	dapat
match	pertandingan
cm	Cuma
muat	mau
kontolllll	zakar
mnjdi	menjadi
announce	mengumumkan
prcy	percaya
rang	orang
kontol	zakar
bodo	bodoh
gajelas	tidak jelas
jaman	zaman
nontonin	menonton
jdi	jadi
sebakpukbakpuknya	payah
bang	abang
asalasalan	asal-asalan
ladies	perempuan
blunder	kesalahan
aje	aja
diem	diam
comeback	kembali
mengucapka	mengucapkan
bete	bosan
gratitude	terima kasih

senyumsenyum	senyum
livechat	obrolan langsung
haziim	anjing
fans	pendukung
pens	pendukung
keinget	ingat
nyemangatin	penyemangat
bg	abang
konspiresyen	konspirasi
emak	ibu
gabile	tidak boleh
bole	boleh
enjoy	santai
moment	momen
ez	mudah
tolol	bodoh
baperan	emosi
provok	provokasi
ngehujat	hujat
gameplay	cara bermain
bener	benar
scene	pemandangan
kalaah	kalah
king	raja
santuy	santai
team	tim
aing	saya
itulaah	itulah
kaya	seperti




difans	pendukung
statement	pernyataan
dewees	dewasa
males	malas
kemaren	kemarin
lebihbaik	lebih baik
pgn	pengen
temennyaatemannya	teman
gada	tidak ada
mystery	misteri
shop	toko
limit	membatasi
gatau	tidak tahu
playing	menjadi
victim	korban
respect	menghormati
ngebanter	keras
world	dunia
bullshit	bohong
gede	besar
sabi	bisa
join	gabung
chanel	jalur
sbnrny	sebenarnya
berseason	musim
ngecengin	mengejek
masalahin	masalah
banned	larang
dapetin	dapat



flare	marak
anggep	anggapan
adekku	adik
gg	keren
geter	getar
stage	panggung
enggk	tidak
aamiin	amin
coach	pelatih
player	pemain
diinterview	wawancara
dek	adik
ilangin	hilang
interview	wawancara
follow	mengikuti
cewk	cewek
cowk	cowok
ajh	saja
kelar	selesai
sempet	sempat
org	orang
kyknya	kayaknya
abis	habis
shooting	penembakan
weekly	mingguan
bkin	bikin
smakin	semakin
ama	sama
ilang	hilang

tir	tingkat
atasgila	atas gila
kocak	lucu
weekday	hari kerja
mudahtetapui	mudah tetapi
tepatdgn	tepat dengan
jumatminggu	jumat minggu
dapt	dapat
kembalikehilangan	kembali kehilangan
byk	banyak
cr	cara
ggltetapi	gagal tetapi
top	atas
dln	dalam
orangtdk	orang tidak
dgn	dengan
sbg	sebagai
warnya	perang
dah	udah
matiin	matiikan
malesin	malas
fair	adil
dibuatbua	dibuat
adlh	adalah
dr	dari
pro	profesional
persentase	presentasi
skillnya	keahlian
draft	penyusunan

demage	kerusakan
decision	keputusan
making	membuat
control	kontrol
map	peta
overallnya	keseluruhannya
that	itu
fact	fakta
taunting	mengejek
battlefield	pertempuran
anteng	tenang
gwh	saya
nub	cupu
jga	juga
ttp	tetap
dikatain	ejek
playoff	pertandingan ualng
bikin	bikin
ak	aku
mingdep	minggu depan
gantiin	ganti
gapernah	tidak pernah
no	tidak
counter	melawan
pantes	pantas
turunin	turun
respectt	menghormati
resign	berhenti
gacukup	tidak cukup



freezer	dingin
nongol	muncul
kambek	kembali
kasi	beri
mindset	berpikir
enjoyed	santai
cool	keren
gameplaynya	permainan
goodluck	semoga beruntung
mudahmudahan	mudah
general	umum
purpose	tujuan
berbudak	anak
dengerin	mendengarkan
casternya	komentator
lengser	menurunkan
ranking	peringkat
katakata	kata
uda	sudah
apaansi	kenapa
playin	bermain
skipp	melewati
ngoceh	berbicara
orangorang	orang
gibah	ngomong
dh	sudah
msh	masih
nyeseknya	sesak
level	tingkat



alhasil	hasil
yaang	yang
ttg	tentang
udahudah	sudah
update	memperbarui
cewe	perempuan
poto	foto
mas	abang
ngalahin	mengalahkan
atmosfer	suasana
confident	yakin
patch	tambalan
lol	bodoh
smw	semua
ma	sama
caster	komentator
english	inggris
indo	indonesia
malay	malasya
ph	filipina
id	indonesia
kgk	tidak
need	butuh
saimbang	seimbang
muter	mutar
mencooba	coba
jdinya	jadinya
terhornat	terhormat
something	sesuatu

nothing	tidak
standings	klasemen
gm	bagaimana
gilak	gila
settingan	pengaturan
anjggg	anjing
analyst	analisis
time	waktu
journey	perjalanan
reverse	balik
nnton	nonton
trnyata	ternyata
garagara	gara
nntn	nonton
anjincce	anjing
hr	hari
kl	kalaupun
dede	adik
kak	kakak
seseorg	orang
ketikk	ketik
ajaa	saja
keturon	ketiduran
ndelok	melihat
nd	di
omah	rumah
skali	sekali
maok	mau
tdk	tidak

skaligus	bersamaan
bkrj	berkerja
ngeprovok	provokasi
bgtupun	begitu
mnding	lumayan
ntn	nonton
stranger	asing
ngabarin	kabar
brengsek	kacau
sebrengsekbrengseknnya	kacau
ngegemesinnya	gemas
vs	versus
like	suka
would	akan
form	membentuk
banyakkk	banyak
fuck	persetan
pause	berhenti
only	hanya
downgrade	menurunkan
flexible	fleksibel
venue	lokasi
info	informasi
bestie	sahabat
ngitung	menghitung
brapa	berapa
ngonten	konten
ngmng	omong
nyobain	mencoba





remeh	berantakan
cwo	cowok
kebonhongan	bohong
bgusny	bagusnya
dmn	dimana
penyepong	nyepong
benerbener	benar
ngecast	komentator
backdoor	belakang
good	bagus
tpi	tapi
wajarin	wajar
stronger	kuat
goblok	bodoh
fav	favorit
nomer	nomor
lgsung	langsung
minlu	admin
nggak	tidak
scrim	latihan
nuke	tukar
bro	kawan
ntar	sebentar
nemenin	menemani
bgttt	sangat
iyaaaaa	iya
teamfight	pertarungan
rookie	calon
maybe	mungkin

playernya	pemain
mliat	melihat
jugaa	juga
kenapaa	kenapa
gnian	ginian
mang	emang
akuin	aku
bhas	bahas
membernarkan	membenarkan
ngatain	hina
mateng	matang
gen	teman
valid	benar
dhlah	sudah
greget	geram
ngga	tidak
winner	juara
ngepost	posting
syahduu	syahdu
kostan	rumah
nggk	tidak
nyebelin	sebel
onfire	semangat
chat	mengobrol
daily	hari
bsa	bisa
manasin	panas
defeat	mengalahkan
sego	nasi



iki	ini
sipaling	paling
jago	kuat
ketawain	tertawa
guys	teman
mrs	merasa
ajalaa	saja
mo	mau
gapake	tidak pakai
chaster	komentator
on	aktif
walo	walau
gamain	tidak main
lamalama	lama
excited	bersemangat
kosan	rumah
streamnya	siaran
clear	bersih
opening	pembukaan
tournament	turnamen
njir	anjing
mergo	karena
bufferbuffer	penahan
hayday	kebruntungan
mung	ingin
kids	anak
dadi	jadi
siji	satu
dino	sekarang



bareng	bersama
gezzz	teman
nyampe	sampai
nnt	nonton
moga	semoga
streak	garis
recall	pulang
jepret	potret
order	memesan
kuyy	ayo
gkk	tidak
anjir	anjing
ngelewat	lewat
gass	gas
futsalnonton	futsal
peha	filipina
defaultnya	bawaan
ngebadut	badut
ucapin	ucapkan
troubleshootnya	masalah
astagaaaa	astaga
ujansi	hujan
apaya	apa
ping	memeriksa
standartnya	standar
non	bukan
off	mati
kaget	terkejut
outdoornya	luar



ban	melarang
dulsss	dulu
mplharus	mpl harus
booster	pendorong
kek	seperti
ciptain	menciptakan
akeh	banyak
mi	mie
bli	beli
abistu	selanjutnya
idolku	idola
boboan	tidur
streamer	langsung
react	reaksi
bru	baru
ap	apa
gmnn	bagaimana
duo	dua
trio	tiga
duonya	dua
turney	turnamen
gueh	saya
idn	indonesia
citacitanya	cita
sayaaaa	saya
ngapain	mengapa
mikirin	mikir
ngelupain	melupakan
role	peran

lane	jalur
surprise	kejutan
ngerusak	merusak
musingin	pusing
ngegosip	gosip
ngerjain	mengerjakan
lemes	lemas
bg	abang
mam	makan
lose	kalah
multitasking	sekaligus
salute	takjub
fire	api
jago in	memilih
kalikali	kali
tulunk	tolong
harini	hari
intro	kenal
naikin	naik
bgstt	bangsat
dosadosaku	dosa
petcah	pecah
powernya	kuat
egk	tidak
freeze	diam
nggaaa	tidak
onoh	itu
ngejekin	ejek
nda	tidak

dipungkirin	pungkiri
nandingin	tanding
dirimujadi	diri jadi
galauin	galau
bgtttt	begitu
seruu	seru
tetep	tetap
samasama	sama
feeder	umpan
ketrigger	pemicu
ggwp	keren
yaudahlah	sudah
prnya	peran
improve	baik
turret	menara
improved	baik
most	paling
battery	baterai
satisfying	memuaskan
dibayangin	bayang
livestream	nonton
lgsg	langsung
krna	karena
phone	telepon
ancrit	anjing
ex	mantan
namunketika	namun ketika
memainkannyasemuanya	main semua
opo	apa



shadow	bayang
kidsz	anak
smpat	sempat
kegesrekanmu	bodoh
krn	karena
mw	mau
presdir	wakil
neror	teror
streamers	langsung
replay	ulang
tengok	lihat
cashier	kasir
apk	aplikasi
proplayer	profesional
diharamin	haram
kynya	seperti
yuk	ayo
capture	menangkap
want	ingin
buy	beli
impactnya	dampak
ingame	main
dibiarin	biar
twice	lipat
nerf	kurang
diheranin	heran
ngedoain	doa
jelekjelek	jelek
story	cerita



back	kembali
roster	daftar
nyari	cari
abangan	abang
anjay	anjing
ntahlah	entah
cayaangku	sayang
sayaaanggg	sayang
benarbenar	benar
kack	kakak
gaaa	tidak
roamer	kelana
rosak	rusak
apakah	apa
gipeway	gratis
development	kembang
kmaren	kemarin
promosiin	promosi
cepat	cepat
kerennnn	keren
puingpuing	puing
engga	tidak
diturunin	turun
sempurnakarna	sempurna karena
kmn	mana
diaaa	dia
worth	nilai
bbrp	berapa
ngerasain	rasa

pendektapi	pendek tapi
duniaa	dunia
adl	adalah
ni	ini
gmn	bagaimana
msi	masih
hrus	harus
pndamping	damping
abalabal	palsu
likubagaikan	rintang
ka	kakak
gamungkin	tidak mungkin
fake	palsu
dipraktekin	praktik
dpt	dapat
tuk	untuk
btp	betapa
amp	sampai
erti	arti
pessan	pesan
perawatanmurah	perawatan murah
puter	mutar
stucknya	jebak
solo	satu
jing	anjing
gakk	tidak
menjadikn	jadi
glory	jaya
moveon	pindah



remahremah	remah
jirr	anjing
pst	pasti
nyebut	sebut
buahbuahan	buah
aje	saja
utang	hutang
asa	rasa
ngejar	lari
samperin	jemput
one	satu
livein	langsung
studi	belajar
national	nasional
sleep	tidur
foundation	dasar
afahh	apa
iyahhh	iya
terkadang	kadang
toxican	buruk
and	dan
jauhdekatan	jauh dekat
chance	peluang
barengg	sama
saiki	sekarang
sik	yang
luwih	lebih
mba	kakak
maklumin	maklum

masalalu	masa lalu
km	kamu
prnah	pernah
jgo	ahli
dik	adik
ngadelin	andal
sikembar	kembar
kagak	tidak
masak	kenapa
cuman	cuma
nyakitin	sakit
ogak	tidak
ogah	tidak
indopride	indonesia
mutusin	putus
jadiin	jadi
membandingbandingkan	banding
gituuu	gitu
infoooo	informasi
infoooin	informasi
emote	emoji
ngeliat	liat
aplg	apalagi
epic	epik
trbaru	baru
pusher	dorong
pad	pada
bljr	belajar
darinyapencundang	dari pecundang

mengerjakannyapecundang	kerja pecundang
mengatakansaya	kata saya
andapecundang	pecundang
mengatakanitu	kata
hrs	harus
sesuatupecundang	sesuatu pecundang
mengatakanpasti	kata pasti
mengerjakannyapecundang	kerja pecundang
mengatakanmungkun	kata mungkin
mungkinpecundang	mungkin pecundang
penjelasanyg	jelas
emapti	empati
pemecahanpecundang	pecah pecundang
suhupecundang	suhu pecundang
perbedaantp	beda
kern	keren
gwejh	saya
sign	masuk
light	cahaya
text	tulis
camera	kamera
ketimbang	timbang
sihmakanny	maka
smwny	semua
playlistnya	daftar
bacod	mulut
syg	sayang
pokus	fokus
terampil	tampil



tmen	teman
gamau	tidak mau
affinitas	afinitas
gituu	gitu
spill	tau
lahhaku	aku
maennya	main
deleted	hapus
catalyst	katalisator
tengil	sebal
ngebucin	cinta
cooldown	tenang
equipment	alat
musicvideo	musik
music	musik
lv	tingkat
dibolehin	boleh
aktifin	aktif
event	peristiwa
callback	balik
drop	jatuh
draw	gambar
link	tautan
hardclone	keras
journalist	wartawan
meet	bertemu
contest	kontes
kepake	pakai
spin	putar



nice	bagus
plus	tambah
pristel	gaya
open	buka
reseller	ecer
recommended	rekomendasi
tri	tiga
hour	jam
latian	latih
reply	ulang
dibilangin	bilang
custom	biasa
dedek	adik
gaenak	tidak enak
nganti	ganti
ngasi	beri
temenku	teman
bacot	mulut
hape	telepon
ultimate	akhir
bapack	bapak
soft	lembut
gemoy	lucu
perundangan	undang
baaaanget	banget
homescapes	rumah
dikerjain	kerja
kemarin	kemarin
cooking	masak



apalg	apalagi
smp	sampai
lvl	tingkat
smurf	lancar
but	tetapi
somehow	bagaimanapun
stumble	tersandung
itumah	itu
diadain	diadakan
sapa	siapa
jom	ayo
noob	buruk
menimpapemain	menimpa pemain
maniac	gila
kire	kira
agik	lagi
ternyata	ternyata
sda	sudah
nie	ini
mmg	memang
tkleh	tidak
pkai	pakai
title	judul
pushrank	naik
plisss	tolong
budeg	tuli
hobby	hobi
all	semua
jst	hanya





feels	rasa
right	kanan
collector	pengumpul
not	bukan
my	saya
executive	eksekutif
ngajakin	ajak
postingannya	pindah
nyolong	curi
buff	positif
alesanku	alasan
tementemenku	teman
tauu	tau
deket	dekat
boy	laki
win	menang
apalgi	apalagi
claim	ambil
didapetin	dapat
tttapi	tapi
ngedrawnya	gambar
minim	minmal
dihapuss	hapus
hellp	tolong
nguras	usah
ksih	kasih
kesel	kesal
kroyokan	keroyok
hbis	habis



dri	dari
jngan	jangan
ngampang	gampang
bwgitu	begitu
orng	orang
bantuin	bantu
please	tolong
do	kerja
your	milik
magic	sihir
bright	terang
ituuu	itu
review	tinjau
dekk	adik
trusted	percaya
banyak	banyak
artist	artis
mayan	lumayan
nggaa	tidak
oge	oke
kepopulerannyaada	populer
gmnaa	bagaimana
sdia	sedia
tnyain	tanya
nunjukin	tunjuk
ngehack	retas
tungguin	tunggu
talk	bicara
mocked	ejek



beginning	awal
end	akhir
even	bahkan
sleeps	tidur
artiin	arti
bdanya	beda
mengahapus	hapus
idupnya	hidup
seruseru	seru
bet	kali
victory	menang
satnite	malam minggu
ratenya	tingkat
kbetulan	betul
disabler	aktif
early	awal
taik	tai
simple	simpel
afk	aktif
gtu	begitu
ngerokok	rokok
udahan	sudah
snowball	tambah
masukin	masuk
league	liga
via	lalu
community	masyarakat
alesannya	alasan
system	sistem




kirakira	kira
mananini	mana
tunggutunggu	tunggu
royal	raja
fredrinn	fredrin
flameshot	api
ajarin	ajar
haru	harus
pure	murni
land	tanah
dawn	fajar
free	gratis
req	minta
user	guna
ngeremehin	remeh
solooooo	diri
redeem	tukar
gin	begini
paok	bodoh
pastu	waktu
hyper	hiper
mak	ibu
ckp	bicara
testi	saksi
thread	benang
bae	saja
yu	ayo
esok	besok
herohero	karakter



skin	baju
skinbaju	baju
sengklek	bodoh
yaudah	sudah
fleeting	sejenak
fistmedia	baru
ngelagjaringan	jaring
guide	memandu
dtg	datang
bising	ribut
begini	gini
build	bangun
app	aplikasi
ngajak	ajak
stun	setrum
ooooom	paman
dibacotin	mulut
kae	itu
isoh	bisa
monteg	keren
dewe	diri
combo	kombo
blend	campur
designer	perancang
bilek	pikir
crowd	ramai
ability	mampu
ora	tidak
ratuanmilenial	milenial

genius	jenius
wand	tongkat
defense	tahan
damage	rusak
sadboy	sedih
kang	tukang
maling	curi
boss	bos
ndaaaa	tidak
feedback	masukan
highlights	ulang
sebelumsampai	sampai
interest	minat
autis	autisme
cba	coba
tsb	tersebut
trend	cenderung
rangked	peringkat
winstreak	menang
tutor	tutorial
belike	suka
kelazz	keren
sma	sama
mabar	main
sumpahin	sumpah
ngak	tidak
why	mengapa
ngerasa	rasa
berdamage	rusak



hp	ponsel
friendly	ramah
joki	jasa
pancet	tetap
ud	sudah
ngebagusin	bagus
wr	nilai
stuck	terjebak
in	di
tier	tingkat
tidack	tidak
presale	diskon
pala	kepala
we	saya
class	kelas
dateng	datang
kbanyakan	banyak
cyber	maya
ranger	pasukan
gaje	tidak jelas
krena	karena
stats	mulai
skuy	ayo
newbie	baru
ya	iya
dianggep	anggap
malemmalem	malam
ngomongnya	bicara
orgil	gila




mimin	admin
contohin	contoh
nerfff	lemah
asu	anjing
op	kuat
surrend	nyerah
kaks	kakak
ga bakal	tidak mungkin
jugando	juga
handphonesmartphonenya	ponsel
smartphone	ponsel
gaming	main
byarnya	bayar
baper	bawa rasa
barbar	kejam
professional	profesional
autisclass	autisme
ready	sedia
diamond	berlian
mumet	pusing
ngapelin	rindu
doi	pacar
sampaikewatan	lewat
parapemula	pemula
mwu	mau
ganteng	ganteng
setting	aturan
yaa	iya
mengethui	tahu



ground	tanah
balikin	balik
welcome	selamat datang
ngedit	edit
nabrakin	tabrak
sarisari	sari
bersamasama	sama
busway	jalan bis
fix	baik
taklukin	takluk
terbaique	baik
relate	kait
adikkk	adik
ngelamun	melamun
mens	laki
Injut	lanjut
deactive	nonaktif
kelimpahankesempatantetapi	limpah sempit
ngedraft	konsep
kecomeback	kembali
adore	cinta
unique	unik
line	garis
know	tahu
really	sebenarnya
vibesnya	atmosfer
saranin	saran
cynk	sayang
aq	aku

tafi	tapi
knffah	kenapa
anying	anjing
ikutin	ikut
tweetnya	komentar
keburukannya	jelek
akuuu	aku
gaess	teman
kutukan	kutuk
runner	lari
sirr	bapak
juaralolos	juara lolos
great	hebat
racestrategi	balap
ngadepin	hadap
segituan	baik
buffering	hambat
malem	malam
stop	henti
check	periksa
dimasing	masing
regiondan	wilayah
komennn	komentar
hujatlama	hujat lama
gausah	tidak perlu
tersakitiapa	sakit
provokpak	provokasi
wes	sudah
humble	rendah



brp	berapa
impianmu	impian
nton	nonton
dg	dengan
sebaikbaiknya	baik
rezekinya	rezeki
sikapsikap	sikap
menyianyiakan	manfaat
gabanyak	tidak banyak
norespect	hormat
diungkit	ungkit
terua	terus
reset	ulang
menangkalah	menang kalah
ayoo	ayo
ajakin	minta
utk	untuk
nt	bagus
nyalain	salah
kta	kata
turnament	turnamen
wakilin	wakil
weekend	akhir pekan
full	penuh
ngambek	marah
nikahin	nikah
next	berikut
draftpicknya	pilih
prioritasin	prioritas



dengerkan	dengar
mvp	baik
gt	gitu
bebudak	anak
kegedean	besar
digantiin	ganti
mantegin	lihat
ampe	sampai
endgame	selesai
hes	dia
just	hanya
moore	lagi
day	hari
kasian	kasih
keteken	tekan
bersihkan	bersih
diperhatiin	perbuatan
explore	jelajah
bully	gertak
sarkasin	sarkas
channel	saluran
spicy	pedas
kalahin	kalah
buruburu	terburu
heronya	karakter
bkl	bakal
down	turun
tulah	itu
fams	keluarga



prasaan	rasa
pdhal	padahal
sutsujinsebelum	sutsujin sebelum
ssekarang	sekarang
final	akhir
grandfinal	tanding akhir
kingdom	penggemar
thatâ€™s	itu
gamenya	main
kmrn	kemarin
football	sepak bola
manager	pengelola
dipungkiri	pungkir
ajg	anjing
lbh	lebih
tournamen	turnamen
matiikan	mematikan
kesuksesananda	kesuksesan anda
hsl	hasil
dkrnkan	karena
krs	keras
hdp	hidup
menunggutetapi	menunggu tetapi
bgs	bagus
dtng	datang
ggl	gagal
kpd	kepada
mrk	mereka
refresh	segar



amda	anda
gimang	bagaimana
nungguin	nunggu
iq	mampu
dibalikin	balik
skrg	sekarang
series	seri
real	nyata
sampe	sampai
cpt	cepat
war	perang
aj	saja
party	pesta
safety	aman
melangkah	langkah
tweet	komentar
gabisa	tidak bisa
maen	main
gua	saya
list	daftar
lu	kamu
najiss	najis
move	gerak
selau	selalu
retwet	terus
gaada	tidak ada
sm	sama
striming	memutar
molo	lagi

lag	hambat
anj	anjing
ngomongin	bicara
dihujaat	hujat
liatin	lihat
play	mulai
kece	keren
dukunin	dukun
cast	pemeran
evil	setan
kesenangn	senang
ngertiin	tahu
seneng	senang
gw	saya
flashback	kilas balik
live	langsung
lo	kamu
seller	penjual
actionan	tindakan
health	kesehatan
semangatin	semangat
horribles	mengerikan
upload	mengunggah
siapin	siap
death	kematian
note	buku
release	melepaskan
seingetku	ingat
sisen	musim



gasalah	tidak salah
inget	ingat
anjrot	anjing
ajggggg	anjing
anjirr	anjing
asw	anjing
anyinkkk	anjing
anjy	anjing
anjeeng	anjing
anjeng	anjing
anjinkkkk	anjing
anjng	anjing
ajggggg	anjing
anjgggggg	anjing
anjer	anjing
anjinggg	anjing
ajgg	anjing
anjg	anjing
jir	anjing
anyiing	anjing
room	ruang
bgt	begitu
support	mendukung
gasi	tidak
macem	macam
tourney	turnamen
tu	itu
anjaay	anjing
aja	saja





yuu	ayo
with	bersama
ayang	sayang
ra	tidak
pengen	mau
date	kencan
red	merah
ml	mobile legends
ga	tidak
struggle	berjuang
stay	tinggal
private	pribadi
belum	belum
superfans	penggemar
dlu	dulu
jelasin	jelas
btrsavage	gila
pertamanyasavage	pertama gila
udh	sudah
dipake	pakai
make	membuat
pick	memilih
tgl	tanggal
hotnews	berita baru
min	admin
knp	kenapa
gue	saya
sir	bapak
mnjd	menjadi


univ	universitas
bindo	bahasa indonesia
mengerikan	ngeri
core	inti
diunfol	berhenti mengikuti
invoice	faktur
pengecekan	cek
best	terbaik
lord	raja
seringkali	sering
rame	ramai
streaming	jalan
fansnya	penggemar
orangtua	orang
tv	televisi
funny	lucu
mlbb	mobile legends
trus	terus
sebenarnya	sebenarnya
permorma	performa
temennyyaa	teman
idol	idola
mengasihi	kasih
temen	teman
beneran	benar
nnti	nanti
sosmed	sosial media
belahan	belah
masnya	abang



leebih	banyak
pdhl	padahal
up	atas
cont	contoh
anakan	anak
thats	itu
msebelum	sebelum
mdlmpl	mdl mpl
makinn	semakin
respectnya	menghormati
menelan	telan
menghilangkan	hilang
auto	otomatis
sniper	penembak jitu
turet	menara
jungle	hutan
trial	uji coba
border	berbatasan
gapapa	tidak apa
boom	ledakan
sepakbola	sepak bola
setaun	tahun
diindo	indonesia
norespect	tidak menghormati
menggoa	gua
eval	evaluasi
perhatian	empati
ketigampl	mpl
belajarlh	belajar



masi	masih
gokill	keren
rblgeek	rbl geek
kelimpahanampkesempatantetapi	kelimpahan kesempatan tetapi
mas	abang
nyekip	lewat
menabg	menang
onlen	tersambung
online	tersambung
ngebolehkan	boleh
ngecomback	kembali
tower	menara
show	menunjukkan
anjgg	anjing
janco	setan
badplay	bermain buruk
berturuturut	turut
vibes	atmosfer
congrats	selamat
amatirrookietier	amatir baru
trs	terus
goldlaner	jalur emas
seruuuu	seru
gokilll	keren
bangett	banget
mengamukk	amuk
cosplay	kostum
trophy	piala



setup	mempersiapkan
nyaingin	saing
city	kota
gadapet	tidak dapat
yesss	iya
power	kekuatan
turtle	penyu
mechanic	montir
matahin	patah
try	coba
roaster	main
smua	semua
blm	belum
literally	secara
unpredictable	tidak bisa ditebak
dear	sayang
tdak	tidak
skinnya	baju
underated	bawah
roamernya	pengelana
bat	sahabat
bgst	bangsat
matchnya	tanding
grow	tumbuh
mood	suasana
trnyta	nyata
gaes	teman
gajadi	tidak jadi
dahlah	sudah



yaaallaahh	tuhan
offline	luring
ranked	peringkat
scroll	menggulir
number	nomor
limited	terbatas
soalnyaaaaa	soal
njeng	anjing
emaptijgn	empati
mrktetapi	mereka tetapi
break	jeda
catchy	menarik
anjrit	anjing
box	kotak
skip	melewati
cringe	jijik
lets	ayo
feeling	merasa
kgaa	tidak
into	dalam
black	hitam
viewersnya	penonton
lngsung	langsung
gais	teman
semalem	malam
ngaco	aneh
maap	maaf
dikomenin	komentar
ngebosenin	bosan



ngerank	peringkat
waiting	nunggu
bu	ibu
sekedar	hanya
funi	lucu
nyengirnya	nyegir
hoki	beruntung
gaselalu	tidak selalu
klah	kalah
oranglain	orang lain
cowo	cowok
sesibuksibuknya	sibuk
wangii	harum
gentle	lembut
malmingku	malam minggu
airmata	air mata
naek	naik
hatihati	waspada
temennya	teman
public	umum
perasaannya	rasa
eraterat	kuat
gilakkk	gila
mainin	main
elu	kamu
konteks	bahas
ngarti	arti
pemaen	pemain
mandiiiiii	mandi



kliatan	liat
pulak	pula
ngmg	bicara
kirain	kira
managernya	pengelola
blng	bilang
blh	boleh
iyaaa	iya
gils	gila
sounds	suara
director	direktur
pmplmpl	pmpl mpl
napa	kenapa
iyaa	iya
dikuasai	kuasa
almosphere	suasana
pride	bangga
gabermasut	tidak bermaksud
let	biar
flow	mengalir
take	mengambil
lupain	lupa
problem	masalah
junglernya	hutan
brpa	berapa
login	gabung
unexpected	tidak terduga
disini	sini
pede	percaya diri




spot	titik
skrng	sekarang
stress	stres
nyoba	coba
sedisepeleuhkan	abaikan
anakanak	anak
baguuus	bagus
makroplay	penguasaan
nyeimbangin	seimbang
maiiin	main
sy	saya
thn	tahun
naiss	keren
leg	tinggal
geng	teman
anjirt	anjing
games	permainan
gnteng	ganteng
	ganteng
gntengggg	ganteng
skwad	pasukan
new	baru
ngetroll	canda
pressure	tekanan
gapunya	tidak punya
unbeaten	tak terkalahkan
icon	ikon
kayagini	kayak begini
nga	tidak



standard	standar
balesin	balas
berantakan	berantak
solid	padat
psti	pasti
beratem	kelahi
ginian	begini
sender	pengirim
bufferingnya	penyangga
teruss	terus
serem	takut
malming	malam minggu
score	skor
kill	membunuh
evoooooss	evos
kelaazzz	keren
jgn	jangan
sneng	senang
thank	terima kasih
you	kamu
paraaah	parah
uga	juga
beres	selesai
demen	suka
tdi	tadi
plot	rencana
twist	memutar
gilaaaaa	gila
nyesel	menyesal



bner	benar
sebel	kesal
maenya	main
mls	malas
trouble	masalah
arenaa	arena
prlu	perlu
mnrutki	menurut
thingy	benda
gaseru	tidak seru
maaff	maaf
gess	teman
sorry	maaf
keasikan	seru
ayangku	sayang
ngelang	lambat
kelassss	keren
hayo	ayo
honey	sayang
body	tubuh
sweaty	berkeringat
wolf	serigala
ngarettt	terlambat
pasponsor	sponsor
kebruntungan	keberuntungan
ngaret	terlambat
denger	dengar
karo	sama
bales	balas



memesa	pesan
pulangg	pulang
duluuu	dulu
compensation	kompensasi
mlmnya	malam
malesan	malas
stream	memutar
anjirlah	anjing
ngudud	merokok
smbil	sambil
apaqah	apakah
ind	indonesia
commander	komandan
segoblog	bodoh
goblognya	bodoh
goblog	bodoh
goblognye	bodoh
ngeleg	lambat
diulangin	ulang
high	tinggi
dimainin	main
seini	ini
mage	penyihir
turn	turun
bahkn	bahkan
diban	larang
expect	menduga
push	dorong
gelut	kelahi



ngegame	main
malminganna	malam minggu
soalna	soal
start	mulai
anjim	anjing
thankyouu	terima kasih
sekalikali	sekali
lgi	lagi
first	pertama
mplrank	tingkat
gasuka	tidak suka
mendam	menyimpan
sebangku	kursi
termanis	manis
pengennnn	ingin
pinter	pintar
seasonnya	musim
mff	maaf
bnr	benar
skrang	sekarang
kluar	keluar
pacarlu	pacar
zoom	perbesar
smoga	semoga
roam	berkeliaran
sepower	kekuatan
sulitjimmy	sulit
coin	koin
straight	lurus



sombonkkk	sombong
avarice	ketamakan
gold	emas
dipick	pilih
ngebuff	monster
timnya	tim
xavierrrrr	xavier
tlg	tolong
sblm	sebelum
member	anggota
ane	saya
ngapa	mengapa
gangerti	tidak mengerti
ranknya	tingkat
mlnya	mobile legends
dideketin	dekat
imbangin	seimbang
anyare	baru
kayo	seperti
based	berdasarkan
pergerakan	gerak
creep	aneh
nontn	nonton
jauhh	jauh
yangg	yang
smpe	sampai
telepon	ponsel
dimulaiiii	mulai
dikk	adik



disuru	minta
klau	kalau
gaessaatnya	teman saat
sya	saya
diikutin	ikut
doain	doa
njirrr	anjing
ngeganjel	menghambat
seolaholah	olah
starter	mulai
cuan	uang
mid	tengah
bsk	besok
enteng	ringan
kudu	harus
mcm	macam
enemey	musuh
msih	masih
ato	atau
kecapai	capai
pamerin	pamer
savage	gila
kesono	sana
draftnya	konsep
ayok	ayo
farmingnya	tani
gilaa	gila
llite	ringan
luck	beruntung

selamanyaa	selamanya
sgala	segala
km	kamu
jjrr	anjing
newsscien	ilmuwan
opnya	kuat
nyaingi	saingan
kpkiran	pikir
pesen	pesan
looknya	tampil
iconicnya	ikon
darinyapecundang	dari pecundang
mengatakanmungkin	mengatakan mungkin
out	keluar
sold	habis
issue	masalah
kidz	anak
akuu	aku
see	melihat
seruuu	seru
official	resmi
theme	tema
song	lagu
didenger	dengar
sanz	santai
gausa	tidak perlu
makasii	terima kasih
kmrin	kemarin
kyaknya	kayak






akukamu	aku kamu
ingetin	ingat
skincare	perawatan kulit
say	mengatakan
goodbye	selamat tinggal
store	toko
happy	bahagia
gaa	tidak
mlmobile	mobile legends
dorang	mereka
friends	teman
bocah	anak
snack	camilan
gombrong	gombroh
gamers	main
ngopi	kopi
pegel	lelah
dolan	jalan
adventure	petualangan
lost	hilang
island	pulau
get	mendapatkan
rich	kaya
summer	musim panas
tips	tip
stressnya	stres
moontonnya	moonton
ricis	rincis
paylater	bayar nanti



said	dikatakan
sanjiiii	sanji
sebapuk	jelek
bapuknya	jelek
perform	melakukan
used	menggunakan
candy	permen
crush	menghancurkan
revamp	merubah
gwe	saya
yaallah	tuhan
jade	jadi
mengobrak	berantakan
bug	kesalahan
hype	kecenderungan
hollywings	sayap suci
terusss	terus
lagendyaaa	legenda
njirr	anjing
brawl	bertengkar
pundek	pundak
shampoo	sampo
capeek	capek
ndak	tidak
nurunin	turun
tergede	besar
diajarin	ajar
pasifiiiiik	pasifik
luculucu	lucu



supproam	mendukung
impact	dampak
mlbbid	mobile legends
winlosewinlose	menang kalah
love	cinta
trimakasii	terima kasih
pngen	mau
acc	menerima
karukaruan	berantakan
kaaak	kakak
nambahin	nambah
starlight	cahaya bintang
championship	kejuaraan
ongly	hanya
master	menguasai
mytic	mitis
trol	terserah
mkanya	maka
lbih	lebih
tidal	tidak
temenan	teman
rocket	roket
ngelampiasin	melampiasikan
idiot	idiot
keunikan	unik
sentry	penjaga
big	besar
smg	semoga
ditawarinnya	menawarkan



didownload	unduh
maaaaaap	maaf
ngeselin	kesal
mythic	mitis
chess	catur
subs	langganan
pingin	mau
sek	bentar
duwe	memiliki
asahlah	asah
saudiyaa	saudiya
counternya	menangkal
paid	bayar
promote	memajukan
yaallahmain	main
terus lupa	terus lupa
maksudnta	maksud
gelud	kelahi
bucin	pacar
durabilitynya	daya tahan
zodiac	zodiak
coment	komentar
ngekill	bunuh
elo	kamu
goooouuubbbllloookk	goblok
purchase	pembelian
rolenya	peran
lengkaap	lengkap
ngimpi	mimpi



players	main
springtime	musim semi
blush	tersipu
chessnya	catur
meeting	pertemuan
lejaat	lezat
deleting	hapus
kamuu	kamu
murmer	murah
winstreakjd	berturut
name	nama
maintenance	pemeliharaan
ofdown	fajar
sungkem	hormat
fast	cepat
cape	capek
premier	perdana
kyak	kayak
fanepage	penggemar
asuwww	anjing
palaku	kepala
fury	marah
mlz	malas
motion	gerakan
sickness	penyakit
soalnya	soal
belii	beli
slotmu	slot
plaza	alun



mouse	tikut
hello	halo
kitty	kucing
fighter	pejuang
iudp	hidup
spend	menghabiskan
cheat	menipu
jalantikus	jalan tikus
damagenya	kekuatan
topup	isi ulang
dodge	menghindari
systemnya	sistem
btul	betul
ngilangin	hilang
palagi	apalagi
ckup	cukup
nyampah	saampah
ngadain	ada
disiapiin	siap
rundown	kebiasaan
ngagetin	kaget
art	seni
legendsku	legenda
membeberkan	bocor
feel	merasa
skill	keahlian
diblok	blokir
regular	reguler
komen	komentar



jungler	hutan
sport	olahraga
assasin	pembunuh
nila	nilai
mainya	main
pemeran	peran
rasis	rasialisme
booming	naik
meghilangkan	hilang
rbl	rebellion
btr	bigetron
rrq	rex regum qeon
gpx	geng kapak
zion	rebellion
albertkairi	albert kairi
ae	alter ego
aex	alter ego
mplmain	mpl main
alberttt	albert
gaikutin	ikut
ngebantai	bantai
perhatiin	lihat
kerenn	keren
fomo	tinggal
dikalahin	kalah
mall	tempat belanja
semangatt	semangat
gakuat	tidak kuat
bykampmenggunakan	guna



ngurusin	mengurus
highground	dataran tinggi
explanernya	jalur dekat
langusng	langsung
nikmatin	nikmat
baheula	dulu
wangiii	wangi
ngebuat	buat
netesin	tetes
bar	setelah
nyepong	sedot
auraonicrrq	aura onic
nangis	tangis
nemeni	teman
disin	sini
kdg	kadang
nntin	nanti
rereki	rex regum qeon
ngelag	lambat
memesan	memesankan
kerinduanchristopher	rindu
kenyataanwilliam	nyata
percumamichaelangelo	Cuma
pingnya	sinyal
cliff	jurang
tag	menandai
ngabisin	habis
bernilai	nilai
apain	apa



malikk	malik
cup	cangkir
branzclover	branz clover
tazzcriteclover	tazz crite clover
bkn	bukan
lite	ringan
gapercaya	tidak percaya
disisiku	sisi
ryuu	naga
ryu	naga
by	oleh
ajalah	saja
sad	sedih
filipin	filipina
dell	lembah
slalu	selalu
cash	tunai
diperlukanbukan	bukan
androidnya	android
recomended	rekomendasi
dolanan	jalan
lagendnyaaa	legenda
chef	koki
ngajakinnya	ajak
serious	serius
ulti	terakhir
ultinya	terakhir
avail	tersedia
website	situs



classic	klasik
ewallet	dompet
handphone	ponsel
dynasty	dinasti
warriors	pejuang
meneh	terus
error	kesalahan
legand	legenda
realitanya	nyata
banyak	banyak
saampah	sampah
mnghibur	hibur
hack	retas
tikut	takut
rollercoaster	gulung
troops	pasukan
install	memasangkan
monmaap	mohon maaf
change	ubah
estesnya	estes
coachzyes	pelatih zyes
gaakan	tidak akan
download	unduh
youtubenya	youtube
drill	bor
mdlgtmpl	mdl mpl
respon	tanggapan
bayangin	bayang
mmbuat	buat

Daftar kata kamus Stopword Removal, diperoleh dari GitHub kemudian mendapatkan tambahan dari penulis. Daftar kata dalam kamus yang digunakan dalam penelitian ini dapat dilihat dibawah ini:

"yak", "mah", "ehh", "gk", "wjj", "kalo", "cmiiw", "pp", "wkwkwk", "sih", "nih", "deh", "ba", "ama", "bibub", "gak", "pake", "yg", "ya", "nya", "jg", "klo", "gpp", "bs", "kuuu", "it", "ala", "hallo", "btw", "cm", "bs", "prcy", "guys", "kayak", "gajelas", "ss", "haha", "dih", "wkwkw", "pas", "aje", "wkwk", "eh", "he", "tp", "dah", "ko", "ez", "atu", "bujed", "maah", "aing", "in", "si", "pgn", "doi", "dihh", "sok", "wts", "ihhh", "gg", "ni", "ajh", "drted", "yt", "tdk", "cr", "byk", "org", "dgn", "sbg", "dpt", "weh", "coyyy", "fyp", "ap", "drpd", "dr", "set", "oh", "ea", "oi", "lh", "gwh", "lak", "jga", "ttp", "kwkwkw", "wokwokwok", "yeayyy", "chuakzzzz", "ha", "nohh", "ls", "nder", "yaa", "sssttt", "grgr", "dh", "ygy", "ggwp", "srr", "ny", "krn", "ekh", "ekhh", "oiya", "asskskskskjddjs", "wkakakakaka", "kyk", "cuy", "sihhh", "smw", "ma", "hahaha", "coy", "hmmmmmmmm", "haha", "et", "cok", "gwwwwww", "wggw", "nahhh", "oalah", "hahahah", "whhwhhkwjwk", "gan", "kahh", "ssi", "wakakak", "wkwkwkwk", "weyyyyyy", "aaaaaaa", "lt", "vid", "dmn", "wei", "mu", "nii", "wh", "ih", "xixixi", "ws", "yah", "bro", "hihi", "siii", "bgtttt", "ajdhajaa", "wkakwkwk", "wkwkwkw", "wkwkwkwkw", "gweh", "kek", "hehe", "wkwkw", "nggk", "ekhh", "bedehh", "kn", "yukk", "gkk", "coo", "gaje", "awkekek", "hh", "dulssss", "woakwokw", "lho", "gmnn", "duo", "solo", "trio", "wehh", "ahshdj", "kpn", "wekewkwkk", "gd", "mlu", "km", "bye", "nyaaa", "ph", "id", "wow", "hi", "nihh", "egk", "nggaaa", "yaaa", "laa", "co", "hehehehe", "hmm", "dll", "cuk", "yagesya",

"iiiiii", "ak", "lgsg", "jd", "krna", "yh", "kk", "hbd", "pmpl", "coyyyyyy", "niii",  
"loh", "gakk", "dongg", "adlh", "nanya", "long", "gm",

"ktana", "bgmn", "amp", "pst", "kfc", "afahh", "iyahhh", "", "dik", "km", "oot",  
"mw", "wk", "bg", "hayukkk", "dmn", "apkh", "ki", "bbrp", "emote", "awowkwok",  
"kern", "gwejh", "smwny", "haahaa", "yeayyyyy", "hahahaha", "yuk", "wkek",  
"hahahha", "ye", "hooh", "euy", "nggak", "ngga", "wkwkk", "ndaaaa",  
"asksksksksk", "awoaowk", "yukkk", "kl", "mna", "xixi", "owh", "yo", "oot",  
"hahahha", "ehehehehe", "cong", "wqwq", "jung", "kah", "its", "rep", "pinoy",  
"panteq", "awokwkwkwk", "awokwkoawk", "rereqi", "gb", "krip",  
"wakakaa", "kbo", "ziggy", "niiy", "yakali", "gapapa", "ahshdhshs", "gp", "jip",  
"wih", "of", "akmj", "phi", "ina", "wuuih", "kochhh", "dehig", "etc", "mol", "up",  
"nyaa", "pre", "ig", "sc", "lj", "is", "tuh", "halhal", "cengan", "satsetsatset", "imo",  
"ah", "duh", "lg", "be", "the", "moba", "wkw", "hai", "kapabayaan", "endangpams",  
"din", "kailangan", "ang", "bata", "mga", "ng", "kaga", "walang", "panindingan",  
"dahil", "lang", "na", "awareness", "sa", "diunfol", "pepekkkkkk", "dkk", "br",  
"les", "cernettes", "https", "pls", "dm", "cam", "naspad", "offflane", "invmpl",  
"yena", "gini", "doang", "ceo", "ceonya", "kb", "koed", "intl", "dao", "ucl",  
"satusatunya", "ketir", "flis", "idnya", "bos", "gar", "mcl", "ges", "sea", "panteg",  
"limax", "gffah", "mic", "yoi", "pl", "grasa", "grusu", "phpin", "aaaand", "ter",  
"eptipi", "dl", "meme", "retripal", "gipewey", "cia", "lstmb", "sreg", "fpl",

"mm", "emyu", "huru", "hara", "grusa", "ttapi", "gapapaa", "ayeuna", "timna",  
"less", "sksk", "koarkoar", "cengo", "emyunya", "inv", "wal", "jembut", "sin", "esl",  
"ibl", "tudei", "wiii", "ci", "shaniemang", "waktuapalagi", "wkwkwkaku", "haduh",  
"lit", "sn", "gs", "lek", "slr", "ngayon", "wsl", "huhuuu", "diremix", "jaw", "bm",

"orok", "dahal", "so", "ku", "dahtu", "omg", "hii", "im", "emg", "daddies", "pleier",  
 "hahha", "yauda", "peka", "buset", "ltd", "wihh", "lahh", "nyablak", "siak", "besh",  
 "yung", "noh", "bolak", "msp", "basreng", "rylia", "monton", "oia", "bib", "drg",  
 "dp", "brg", "poqe", "pool", "tnc", "wsnya",

"ft", "tingin", "nahh", "cookk", "an", "segabut", "kek", "pd", "btk", "siiii", "niii",  
 "sia", "ngeh", "jingle", "sara", "najokn", "bk", "gh", "au", "oke", "soton", "klasiko",  
 "lalakas", "nasa", "viva", "lemong", "naman", "dias", "psk", "woi", "epl", "cook",  
 "mbek", "neh", "loose", "drachin", "wa", "kt", "as", "jog", "ooo", "ling",  
 "yoweslah", "mlh", "gbs", "muntun", "hiper", "liliput", "ahaha", "pss", "zippie",  
 "imah", "bede", "psby", "degdegan", "bt", "boleg", "rsg", "ups", "jt", "usd",  
 "sustain", "rtonya", "plis", "gi", "ngab", "yok", "yutub", "hoodiejaket", "hanin",  
 "wibu", "dap", "amy", "pantek", "gagi", "ampota", "fps", "rt", "mio", "omi", "atuh",  
 "chuu", "ive", "jan", "caper", "ntah", "la", "didu", "aya", "reno", "mla", "pg", "pwi",  
 "pk", "hime", "doom", "njun", "pengbes", "bill", "ngepong", "hm", "bia", "azizom",  
 "elaaang", "ente", "bolo", "meh", "takziah", "allan", "wahh", "san", "abrik", "bngt",  
 "aka", "hanam", "ksl", "cieee", "bokep", "abg", "ngewe", "bio", "mlers", "rp", "joo",  
 "wombo", "irl", "coa", "aplot", "gyu",

"betmut", "taeyong", "nct", "bl", "fu", "gasii", "mgl", "kai", "nich", "hilkanaeru",  
 "krq", "nick", "ngedrakor", "khn", "hei", "geer", "sii", "dejag", "jedug", "cung",  
 "kpop", "put", "debuz", "kage", "playstore", "kobo", "hololive", "bengbeng", "pns",  
 "avial", "gabut", "neosef", "tara", "go", "hut", "de", "sunggas", "sungdone", "mao",  
 "bebel", "da", "ahlan", "fred", "fil", "nfc", "codm", "pi", "tetus", "marn",  
 "monteng", "ijek", "kesampluk", "teh", "kon", "mengkis", "duwa", "sireplays",  
 "lohb", "subrek", "kbi", "zeejkt", "idk", "lgbt", "potbar", "agora", "gmf", "tgif",

"xd", "wha", "penangan", "sagatroops", "apanih", "begiji", "dns", "us", "yajjj",  
 "wib", "gabakal", "slave", "miss", "gdm", "poll", "kwkwk", "ahh", "pms", "mobi",  
 "chuakz", "mleb", "apex", "elah", "pyonfuwa", "fuwa", "waku", "fb", "xl", "tbk",  
 "minnie", "yes", "tea", "gowes", "tbtb", "smk", "wawa", "yokk", "tsuriz", "wc",  
 "mmclassic", "kominfo", "heeeehhh", "liquipedia", "mariokm", "umumumu",  
 "katalisator", "aaaaaaaaa", "huwoooo", "uuwoooo", "skater", "bangbang",  
 "selulermobile", "donk", "mamba", "awowkwk", "rankclassic", "kwkwkwkw",  
 "ngangongngangong", "boro", "dumai", "ellajkt", "seifukunya", "nyutnyutan",  
 "tweeps", "akaihaato", "hachama", "keunn", "choukai", "mfzeromfzero", "lod",  
 "hyperblend", "dmchat", "gtumzk", "taniyyaw", "mfzeroo", "daha", "kotu",  
 "legendskurtulamadim", "polkrab", "tin", "subathon", "esp", "azkun", "rm", "pape",  
 "tah", "wae", "zackyyart", "nek", "classicranked", "sie", "mbah", "week", "gasie",  
 "ff", "biw", "txt", "monya", "apayak", "jikwana", "besendilala", "kakade", "sihh",  
 "wkwkwwkwkwkwkwkwkwkw", "patayin", "ingayyy", "bgtt", "nge", "yailah",  
 "gabermksut", "waduueuchhh", "emyumplpmpl", "man", "ngangngong",  
 "aoruraborealis", "yaye", "kemendikbud", "huuufft", "juday", "spes", "smcon",  
 "asn", "malasya", "hahahahahah", "bade", "jo", "wtb", "hadah", "wota", "okwe",  
 "spec", "knc", "bwc", "woy", "wkkwk", "bileg", "mencapainyatheodore",  
 "roosevelt", "ky", "membung", "ms", "hehehe", "hw", "ai", "ser", "tod", "tt", "teu",  
 "dc", "gpm", "po", "at", "manonood", "ok", "je", "or", "pubg", "cls", "to", "tix",  
 "wale", "jagah", "hota", "tha", "lekin", "unhu", "ne", "arm", "pe", "likha",  
 "kab", "lage", "bihar", "me", "sing", "nepikeun", "sapuluh", "ptdrrrrr", "hein",  
 "hayang", "modar", "arisia", "beuli", "duitna", "euweuh", "meuleug",  
 "meuleeeuuugg", "nihada", "nak", "kda", "mint", "hihi", "woah", "mon", "hdpdlm",

"haldilalui", "wuidi", "kung", "hindi", "manggugukat", "nalang", "dripda", "kuy", "ta", "jak", "dom", "koh", "ad", "yss", "ano", "hlo", "kya", "aap", "bta", "saktea", "kisay", "zit", "sakta", "hu", "lakilacknat", "yap", "lali", "rya", "mfzero", "immfhrzl", "gmepiclegend", "season", "owkaiii", "sbng", "haasa",

"gendha", "therey", "dhanee", "ada", "adalah", "adanya", "adapun", "agak", "agaknya", "agar", "akan", "akankah", "akhir", "akhiri", "akhirnya", "aku", "akulah", "amat", "amatlah", "anda", "andalah", "antar", "antara", "antaranya", "apa", "apaan", "apabila", "apakah", "apalagi", "apatah", "artinya", "asal", "asalkan", "atas", "atau", "ataukah", "ataupun", "awal", "awalnya", "bagai", "bagaikan", "bagaimana", "bagaimanakah", "bagaimanapun", "bagi", "bagian", "bahkan", "bahwa", "bahwasanya", "baik", "bakal", "bakalan", "balik", "banyak", "bapak", "baru", "bawah", "beberapa", "begini", "beginian", "beginikah", "beginilah", "begitu", "begitukah", "begitulah", "begitupun", "bekerja", "belakang", "belakangan", "belum", "belumlah", "benar", "benarkah", "benarlah", "berada", "berakhir", "berakhirilah", "berakhirnya", "berapa", "berapakah", "berapalah", "berapapun", "berarti", "berawal", "berbagai", "berdatangan", "beri", "berikan", "berikut", "berikutnya", "berjumlah", "berkali-kali", "berkata", "berkehendak", "berkeinginan", "berkenaan", "berlainan", "berlalu", "berlangsung", "berlebihan", "bermacam", "bermacam-macam", "bermaksud", "bermula", "bersama", "bersama-sama", "bersiap", "bersiap-siap", "bertanya", "bertanya-tanya", "berturut", "berturut-turut", "bertutur", "berujar", "berupa", "besar", "betul", "betulkah", "biasa", "biasanya", "bila", "bilakah", "bisa", "bisakah", "boleh", "bolehkah", "bolehlah", "buat", "bukan", "bukankah", "bukanlah", "bukannya", "bulan", "bung", "cara", "caranya", "cukup", "cukupkah", "cukuplah", "cuma", "dahulu",

"dalam", "dan", "dapat", "dari", "daripada", "datang", "dekat", "demi", "demikian", "demikianlah", "dengan", "depan", "di", "dia", "diakhiri", "diakhirinya", "dialah", "diantara", "diantaranya", "diberi", "diberikan", "diberikannya", "dibuat", "dibuatnya", "didapat", "didatangkan", "digunakan", "diibaratkan", "diibaratkannya", "diingat", "diingatkan", "diinginkan", "dijawab", "dijelaskan", "dijelaskannya", "dikarenakan", "dikatakan", "dikatakannya", "dikerjakan", "diketahui", "diketahuinya", "dikira", "dilakukan", "dilalui", "dilihat", "dimaksud", "dimaksudkan", "dimaksudkannya", "dimaksudnya", "diminta", "dimintai", "dimisalkan", "dimulai", "dimulailah", "dimulainya", "dimungkinkan", "dini", "dipastikan", "diperbuat", "diperbuatnya", "dipergunakan", "diperkirakan", "diperlihatkan", "diperlukan", "diperlukannya", "dipersoalkan", "dipertanyakan", "dipunyai", "diri", "dirinya", "disampaikan", "disebut", "disebutkan", "disebutkannya", "disini", "disinilah", "ditambahkan", "ditandaskan", "ditanya", "ditanyai", "ditanyakan", "ditegaskan", "ditujukan", "ditunjuk", "ditunjuki", "ditunjukkan", "ditunjukkannya", "ditunjuknya", "dituturkan", "dituturkannya", "diucapkan", "diucapkannya", "diungkapkan", "dong", "dua", "dulu", "empat", "enggak", "enggaknya", "entah", "entahlah", "guna", "gunakan", "hal", "hampir", "hanya", "hanyalah", "hari", "harus", "haruslah", "harusnya", "hendak", "hendaklah", "hendaknya", "hingga", "ia", "ialah", "ibarat", "ibaratkan", "ibaratnya", "ibu", "ikut", "ingat", "ingat-ingat", "ingin", "inginkah", "inginkan", "ini", "inikah", "inilah", "itu", "itukah", "itulah", "jadi", "jadilah", "jadinya", "jangan", "jangan", "janganlah", "jauh", "jawab", "jawaban", "jawabnya", "jelas", "jelaskan", "jelaslah", "jelasnya", "jika", "jikalau", "juga", "jumlah", "jumlahnya", "justru", "kala", "kalau", "kalaulah", "kalaupun", "kalian", "kami",



"kamilah", "kamu", "kamulah", "kan", "kapan", "kapankah", "kapanpun", "karena", "karenanya", "kasus", "kata", "katakan", "katakanlah", "katanya", "ke", "keadaan", "kebetulan", "kecil", "kedua", "keduanya", "keinginan", "kelamaan", "kelihatan", "kelihatannya", "kelima", "keluar", "kembali", "kemudian", "kemungkinan", "kemungkinannya", "kenapa", "kepada", "kepadanya", "kesampaian", "keseluruhan", "keseluruhannya", "keterlaluhan", "ketika", "khususnya", "kini", "kinilah", "kira", "kira-kira", "kiranya", "kita", "kitalah", "kok", "kurang", "lagi", "lagian", "lah", "lain", "lainnya", "lalu", "lama", "lamanya", "lanjut", "lanjutnya", "lebih", "lewat", "lima", "luar", "macam", "maka", "makanya", "makin", "malah", "malahan", "mampu", "mampukah", "mana", "manakala", "manalagi", "masa", "masalah", "masalahnya", "masih", "masihkah", "masing", "masing-masing", "mau", "maupun", "melainkan", "melakukan", "melalui", "melihat", "melihatnya", "memang", "memastikan", "memberi", "memberikan", "membuat", "memerlukan", "memihak", "meminta", "memintakan", "memisalkan", "memperbuat", "mempergunakan", "memperkirakan", "memperlihatkan", "mempersiapkan", "mempersoalkan", "mempertanyakan", "mempunyai", "memulai", "memungkinkan", "menaiki", "menambahkan", "menandaskan", "menanti", "menanti-nanti", "menantikan", "menanya", "menanyai", "menanyakan", "mendapat", "mendapatkan", "mendatang", "mendatangi", "mendatangkan", "menegaskan", "mengakhiri", "mengapa", "mengatakan", "mengatakannya", "mengenai", "mengerjakan", "mengetahui", "menggunakan", "menghendaki", "mengibaratkan", "mengibaratkannya", "mengingat", "mengingatkan", "menginginkan", "mengira", "mengucapkan", "mengucapkannya", "mengungkapkan", "menjadi", "menjawab", "menjelaskan", "menuju",

"menunjuk", "menunjuki", "menunjukkan", "menunjuknya", "menurut", "menuturkan", "menyampaikan", "menyangkut", "menyatakan", "menyebutkan", "menyeluruh", "menyiapkan", "merasa", "mereka", "merekalah", "merupakan", "meski", "meskipun", "meyakini", "meyakinkan", "minta", "mirip", "misal", "misalkan", "misalnya", "mula", "mulai", "mulailah", "mulanya", "mungkin", "mungkinkah", "nah", "naik", "namun", "nanti", "nantinya", "nyaris", "nyatanya", "oleh", "olehnya", "pada", "padahal", "padanya", "pak", "paling", "panjang", "pantas", "para", "pasti", "pastilah", "penting", "pentingnya", "per", "percuma", "perlu", "perlukah", "perlunya", "pernah", "persoalan", "pertama", "pertama-tama", "pertanyaan", "pertanyakan", "pihak", "pihaknya", "pukul", "pula", "pun", "punya", "rasa", "rasanya", "rata", "rupanya", "saat", "saatnya", "saja", "sajalah", "saling", "sama", "sama-sama", "sambil", "sampai", "sampai-sampai", "sampaikan", "sana", "sangat", "saya", "sayalah", "se", "sebab", "sebabnya", "sebagai", "sebagaimana", "sebagainya", "sebagian", "sebaik", "sebaik-baiknya", "sebaiknya", "sebaliknya", "sebanyak", "sebegini", "sebegitu", "sebelum", "sebelumnya", "sebenarnya", "seberapa", "sebesar", "sebetulnya", "sebisanya", "sebuah", "sebut", "sebutlah", "sebutnya", "secara", "secukupnya", "sedang", "sedangkan", "sedemikian", "sedikit", "sedikitnya", "seenaknya", "segala", "segalanya", "segera", "seharusnya", "sehingga", "seingat", "sejak", "sejauh", "sejenak", "sejumlah", "sekadar", "sekadarnya", "sekali", "sekali-kali", "sekalian", "sekaligus", "sekalipun", "sekarang", "sekarang", "sekecil", "seketika", "sekiranya", "sekitar", "sekitarnya", "sekurang-kurangnya", "sekurangnya", "sela", "selain", "selaku", "selalu", "selama", "selama-lamanya", "selamanya", "selanjutnya", "seluruh", "seluruhnya", "semacam", "semakin", "semampu", "semampunya", "semasa", "semasih",

"semata", "semata-mata", "semaunya", "sementara", "semisal", "semisalnya",  
 "sempat", "semua", "semuanya", "semula", "sendiri", "sendirian", "sendirinya",  
 "seolah", "seolah-olah", "seorang", "sepanjang", "sepantasnya", "sepantasnyalah",  
 "seperlunya", "seperti", "sepertinya", "sepihak", "sering", "seringnya", "serta",  
 "serupa", "sesaat", "sesama", "sesampai", "sesegera", "sesekali", "seseorang",  
 "sesuatu", "sesuatunya", "sesudah", "sesudahnya", "setelah", "setempat",  
 "setengah", "seterusnya", "setiap", "setiba", "setibanya", "setidak-tidaknya",  
 "setidaknya", "setinggi", "seusai", "sewaktu", "siap", "siapa", "siapakah",  
 "siapapun", "sini", "sinilah", "soal", "soalnya", "suatu", "sudah", "sudahkah",  
 "sudahlah", "supaya", "tadi", "tadinya", "tahu", "tahun", "tak", "tambah",  
 "tambahnya", "tampak", "tampaknya", "tandas", "tandasnya", "tanpa", "tanya",  
 "tanyakan", "tanyanya", "tapi", "tegas", "tegasnya", "telah", "tempat", "tengah",  
 "tentang", "tentu", "tentulah", "tentunya", "tepat", "terakhir", "terasa", "terbanyak",  
 "terdahulu", "terdapat", "terdiri", "terhadap", "terhadapnya", "teringat", "teringat-  
 ingat", "terjadi", "terjadilah", "terjadinya", "terkira", "terlalu", "terlebih", "terlihat",  
 "termasuk", "ternyata", "tersampaikan", "tersebut", "tersebutlah", "tertentu",  
 "tertuju", "terus", "terutama", "tetap", "tetapi", "tiap", "tiba", "tiba-tiba", "tidak",  
 "tidakkah", "tidaklah", "tiga", "tinggi", "toh", "tunjuk", "turut", "tutur", "tuturnya",  
 "ucap", "ucapnya", "ujar", "ujarnya", "umum", "umumnya", "ungkap",  
 "ungkapnya", "untuk", "usah", "usai", "waduh", "wah", "wahai", "waktu",  
 "waktunya", "walau", "walaupun", "wong", "yaitu", "yakin", "yakni", "yang"