ANALISIS SENTIMEN PADA MEDIA SOSIAL INSTAGRAM KLUB PERSIJA JAKARTA MENGGUNAKAN METODE NAIVE BAYES

Andres Mondaref Jon
Program Studi Informatika – Program Sarjana Universitas Islam Indonesia
Daerah Istimewa Yogyakarta, Indonesia
andres.jon@students.uii.ac.id

Irving Vitra Paputungan
Jurusan Informatika
Universitas Islam Indonesia
Daerah Istimewa Yogyakarta, Indonesia
irving@uii.ac.id

Abstrak — **Pemberitaan mengenai Persija tidak terlepas dari peranan media dari era perserikatan hingga sekarang. Pemberitaan Persija pertama kali di media cetidak pada tahun 1938 di surat kabar Pemandangan. Hingga ditahun 2000-an suporter Persija Jakarta mendirikan portal berita *online* dengan nama Jak*Online* di tahun 2001 sebagai wadah pemberitan khusus untuk pecinta Persija Jakarta dan The Jakmania. Menggunakan Metode Analisis Deskriptif, berbentuk *Word cloud*, yang digunakan dalam penelitian ini untuk mengidentifikasi dan membentuk pola kata yang dapat berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting. Metode *Machine Learning* yaitu *Naive Bayes Classifier*, yang dalam penelitian ini digunakan untuk mengklasifikasikan ulasan atau review yang dapat berbentuk kelas positif maupun kelas negatif. pembagian terstruktur mengenai ulasan holistik data sebanyak 3.000 data dihasilkan beberapa kata yang paling banyak terdapat yaitu istilah "terbaik" sebesar 408 kali, kata "edan" sebesar 396 kali, kata "semangat" sebesar 328 kali, istilah "persija" sebesar 292 kali, istilah "alhamdulillah" sebanyak 261 kali serta seterusnya. Semua Data yang diperoleh memiliki scope berasal asal komentar bahasa Indonesia saja pada postingan akun official Persija selama BRI Liga 1 berlangsung Postingan yang diambil merupakan unggahan tentang *Starting Eleven*, *Halftime* serta *Fulltime* pada postingan Persija Jakarta selama BRI liga 2022/2023 berlangsung.**

Keywords – Sentiment Analysis; Instagram; Naïve Bayes Classifier; Persija Jakarta; Wordcloud.

I. PENDAHULUAN

Info tentang Persija tidak lain serta tidak bukan karena media berasal era perserikatan sampai sekarang. gosip Persija Jakarta pertama muncul pada media cetak di tahun 1938 pada surat kabar Pemandangan. sampai ditahun 2000-an the Jak Jakarta mendirikan pusat berita *online* yang diberi nama Jak*Online* pada tahun 2001 menjadi pusat semua berita buat The Jakmania [1]. Sesuai dengan pengoptimalan dari perkembangan teknologi, tim Persija memberitahukan dan menginfokan seluruh kegiatannya melalui *website* www.persija.co.id. Hal tadi adalah cara yang tepat buat menjawab kebutuhan akan informasi serta mengambarkan keberadaan klub Persija Jakarta kepada *fans* nya. Selain itu Persija mempunyai akun media sosial instagram resmi yaitu @persija. sesuai data terakhir yang ditemukan pada 7 September 2022, akun @persija telah berhasil menjaring *followers* sebesar lebih dari 3,3 juta pengguna serta mengupload postingan sebesar 9.958 (https://www.instagram.com/persija/). Beberapa hal yang muncul pada komentar akun Instagram @persija berkaitan menggunakan sikap supporter yang mengomentari yang bersifat negatif maupun positif yang berpengaruh terhadap lingkungannya serta sikap suporter selanjutnya. Umumnya salah satu *problem* sepele bisa mengakibatkan adu argumen antar suporter ialah karena adanya tim kalah, nyanyian rasis, aksi saling ejek, keputusan wasit yang tidak adil merendahkan klub lawan.

Di sisi lain penggunaan aplikasi sosial media Instagram merupakan aplikasi yang sangat populer digunakan [2]. Berdasarkan permasalahan yang disinggung diatas, maka dibutuhkan suatu teknik atau cara-cara analisis sentimen untuk mencoba menentukan sentimen seseorang (label evaluasi positif atau label evaluasi negatif) dari suatu objek (unggahan sosial media). Pada umumnya ada beberapa metode yang biasa dipergunakan untuk analisis sentimen terutama pada klasifikasi, diantaranya *Naïve Bayes Classifier*. *Naïve Bayes Classifier* mempunyai taraf ketepatan tertinggi pada mengkategorikan teks bahasa. Bahwasanya *Naïve Bayes Classifier* menghasilkan teknik menemukan nilai probabilitas bersyarat terbesar berasal masing-masing kelas [3]. Karena penelitian ini dinilai asal kemungkinan komentar *negative* serta *positive*, maka metode yang paling sempurna yang digunakan yaitu *Naïve Bayes Classifier*.

Metode *Naïve Bayes Classifier* artinya penggolongan kemungkinan sederhana buat mencari beberapa kemungkinan dengan menambahkan kombinasi serta frekuensi nilai berasal set data yang digunakan. Definisi lain berkata Metode *Naive bayes Classifier* adalah penggolongan menggunakan kemungkinan metode dan statistik pada publish sang ilmuwan asal England Bernama Thomas Bayes, dia menyebutkan bahwa prediksi atau tebakan di masa yang akan tiba yaitu asal pengalaman di masa lampau [4]. Kelebihan dalam memakai penggunaan *Naive bayes Classifier* adalah metode yang dapat menggunakan pelatihan data minim untuk memilih asumsi paremeter apa saja dipergunakan dalam proses pengelompokan.

Penelitian ini akan membahas untuk mengetahui performa dari metode *Naïve Bayes Classifier* dalam melakukan klasifikasi sentimen postif dan negatif pada media sosial Instagram klub sepakbola Persija Jakarta. Dengan dilakukannya penelitian tersebut diharapkan dapat memberikan gambaran tentang bagaimana reaksi audiens pada setiap topik unggahan menegenai komentar postif dan negatif pada akun Instagram Persija Jakarta.

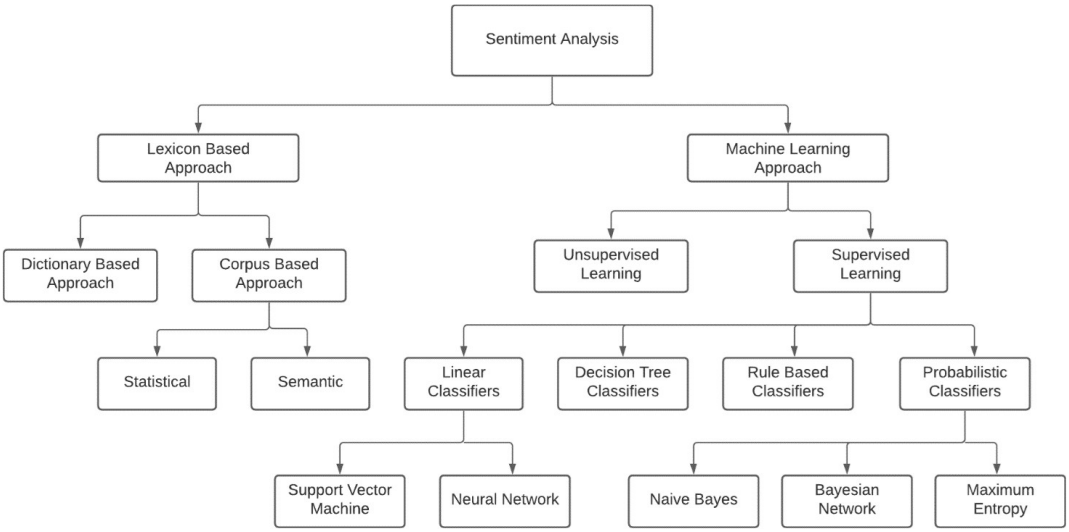
II. KAJIAN PUSTAKA

Analisis sentimen artinya metode analisis yang dipergunakan buat memilih sentimen audiens mengenai reaksi atau pandangannya terhadap suatu kenyataan yang terjadi. *Sentiment Analysis* ialah suatu langkah pada

melakukan serta tahu proses ekstraksi serta pula mengolah data yang berupa teks menggunakan automatis tujuan buat mendapatkan pengetahuan (*insight*) Sentiment yang terdapat pada dalam sebuah komentar. *Sentiment Analysis* dapat juga melihat pendapat atau kecenderungan pendapat terhadap suatu konflik dari seorang, pendapat tadi cendrung berpandangan ke arah opini negatif atau ke arah opini positif [5].

Sentiment Analysis digunakan buat pembagian terstruktur mengenai ulasan. Penjabaran ulasan yang dilakukan di penelitian tadi artinya penjabaran ulasan terhadap *business process* pada *University*. Opini publik tadi mayoritas mengarah pada ulasan teks (ulasan positif dan ulasan negatif) mereka mengenai proses usaha yang terdapat di perguruan tinggi. penekanan utama asal *opinion mining* di pada proses melakukan analisis sentimen yaitu melakukan analisis opini asal suatu dokumen yang berbentuk tekstual. Analisis sentimen pula dapat dilakukan dalam melakukan pelacakan opini di pada diskusi *online.*

Pada sisi lain, analisis sentimen dipergunakan buat memilih sentimen seseorang (mampu berupa reaksi serta kecendrungan) mereka apakah mengarah pada sentimen positif ataupun sentimen negatif, semua itu tergantung kepada reaksi dan pandangan mereka terhadap penekanan objek eksklusif [6]. Selain itu, analisis sentimen juga dapat melakukan kompendium ulasan, proses ekstraksi (pemisahan) antara sinonim (persamaan kata) serta antonim (lawan kata) dari suatu dokumen yang berbentuk tekstual. Berikut metode analisis sentimen dapat dilihat pada Gambar 1.



Gambar 1. Metode analisis sentimen

Metode dalam melakukan analisis sentimen banyak ragamnya. Metode analisis sentimen *berbasis machine learning* (pembelajaran mesin) dan berbasis *lexicon* (leksikon atau kosa istilah). Metode analisis sentimen berbasis pembelajaran mesin dikelompokkan oleh dua golongan yaitu *unsupervised learning* dan *supervised learning* sedangkan metode analisis sentimen berbasis leksikon dibagi menjadi dua yaitu *dictionary based* serta *corpus based*. Pada metode pembelajaran mesin yaitu *supervised learning*, pendekatan ini digunakan Jika ada data berlabel yang telah tersedia buat dilatih modelnya. Langkah-langkah yang dilakukan pada *supervised learning* adalah melatih contoh dan sisa lainnya ialah melakukan prediksi. pada *unsupervised learning*, pendekatan ini digunakan Bila realibilitas data berlabel diposisikan sulit. Kalimat yang dikoleksi akan dikategorikan sesuai kata kunci di setiap kategori, untuk menganalisis data yang bersifat dependen (bergantungan) maka pendekatan *unsupervised learning* lebih praktis untuk dipergunakan [7].

Selanjutnya, di metode leksikon yaitu *dictionary based*, pendekatan ini dipergunakan buat mencari sinonim kata menggunakan tujuan untuk memperbesar berukuran pada deretan kata yang sama. Selain itu opini kata-istilah opini yang unik akan diperluas sebagai fitur buat melakukan analisis sentimen. Di pendekatan *corpus based* ditemukan bahwa sebuah kata memliki orientasi konteks. Pendekatan berbasis korpus (struktur bahasa) dilakukan menggunakan

dua cara yaitu pendekatan berbasis statistik (data nomor) serta pendekatan semantik (makna kata). Pada penelitian wacana metode analisis sentimen yang dilakukan oleh [7] tersebut menyebutkan bahwa metode analisis sentimen yang akan dilakukan haruslah menyesuaikan dengan penekanan dan lingkup penelitian tentang analisis sentimen.

Berbicara mengenai analisis sentimen pada media umum, *sentiment analysis* pada media sosial mampu diinterpretasikan menggunakan teknik pengelompokan *positive sentiment* dan *negative sentiment* pada postingan di media sosial, merupakan unggahan yang mempunyai reaksi positif ditempatkan di sentimen positif sedangkan unggahan yang memiliki reaksi negatif ditempatkan pada sentimen negatif. pada proses pengklasifikasiannya, polaritas (hal yang bersifat berlawanan) di setiap kata juga diperhatikan. Mereka mencoba mengklasifikasikan sentimen menggunakan mempertimbangkan unggahan yang merepresentasikan hal negatif atau fokus terhadap objek eksklusif [8]. Selain itu, penelitian tadi menunjuk kepada evaluasi terhadap efektivitas media umum menjadi destinasi buat memberikan komunikasi tertentu.

Penelitian tadi menyebutkan bahwa konten yang terdapat di media sosial dapat dianalisis sentimennya yaitu mirip dapat dikategorikan unggahan sesuai menggunakan format atau tipe, mampu jua konten bersama daya tariknya [9]. Terkait analisis sentimen di media umum, analisis topik pada media sosial umum dapat menolong *decision making* institusi negara untuk menjalankan keputusan (mirip menjalankan penerapan keputusan yang berafiliasi dengan suatu larangan) [10]. Pada penelitian yang dilakukan oleh Spillane membuat rumusan dilema bagaimana hasil analisis konten unggahan di Instagram dan pengaruhnya terhadap keputusan kebijakan pemerintah pada memerangi perkara narkoba [10]. Ada beberapa metode yang menjadi perbandingan pada penelitian ini, berikut dapat dilihat pada Tabel 1.

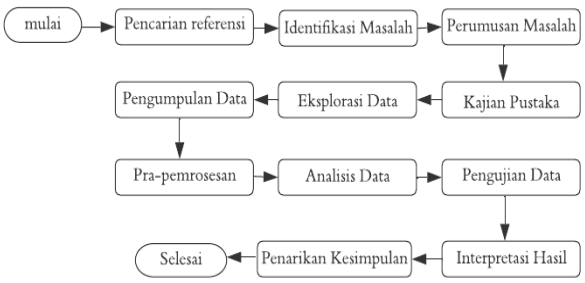
Tabel 1. Penelitian terkait

No	Metode	Kekurangan dan Kelebihan
1	*Support Vector Machine*	1. Perlu didesain kamus khusus unuk menangani masalah yang menggunakan data Twitterberbahasa Indonesia, yang cenderung memakai bahasa tidak baku, supaya hasilnya menjadi maksimal [11]. 2. Data didapatkan dari sistem opini, kemudian faktor penilaiannya dapat digunakan bagi pihak tertentu, agar mempertinggi hasil kerja pada pelayanan transportasi awam darat, terkhusus pada kota di Indonesia [11].
		1. Dapat diimplementasikan di *sentiment anlysis* pada kepuasan pengguna *provider* seluler menggunakan pembagian terstruktur mengenai berupa *positive sentiment* atau *negative sentiment*. Penyajian informasi riset ini diperoleh berasal Twitter berisi data teks yang adalah kumpulan cuitan berasal pengguna *provider* seluler paling popular [12]. 2. Menerima ketepatan bernilai 79%, presisi bernilai 65%, persenan *recall* bernilai 97%, dan *f-measure* bernilai 78%. *Recall value* yang tinggi didistraksi sang jumlah uji data yang harusnya bernilai *positive* serta dideteksi sebagai nilai *positive* oleh system. Hal ini memberikan warta bahwa sentiment analysis dengan menggunakan metode SVM dan LBF baik dipakai buat pengelompokan sentiment analysis [12].
		Menggali informasi perusahaan aviasi di cuitan Twitter menggunakan optimal. Mendapatkan hasil parameter *learning rate* (gamma) sebanyak 0,03 dan nilai C sebanyak 10 untuk hasil maksimal. Dihasilkan taraf ketepatan maksimal sebanyak 40%, *precision* sebanyak 40%, *recall* sebanyak 100%, dan *f-measure* sebanyak 57,14%. Hasil *accuracy* diperoleh sebanyak perulangan maksimal sebesar 50 *loop* menggunakan fitur *lexicon based*. Pada menghitung ketepatan dipergunakan pengukuran paling baik yang dipaparkan sebelumnya, yang akan terjadi klasifikasi dipengaruhi parameter yang paling baik atau maksimal serta objek yang dipakai [13].
2	*Naive Bayes Classifier*	Akurasi yang didapatkan mencapai nomor 97% maka asal itu *Naïve bayes classifier Method* berfungsi buat mengklasifikasi cuitan menggunakan *negative sentiment* dan *positive sentiment* menggunakan cara otomatis. [14].
		1. *Naïve Bayes Classifier* mampu mengkalkulasikan ulasan pada produk *online* menggunakan perangkat yang digunakan [15]. 2. S*entiment analysis* pada ulasan produk *online* menggunakan metode *NBC* membuat nilai *accuracy* terkecil di pengujian 5 kelas menggunakan *dataset* 80% latih serta 20% data uji sebanyak 52.66%, dan pada pengujian 3 kelas memakai *dataset* 90% data latih serta 10% data uji memperoleh *accuracy* tertinggi sebesar 77.78% [15].
		Sistem yang dikembangkan dapat dapat merubah sentimen ulasan teks di website prudential serta mampu memvisualisasikan info sentimen rakyat mengenai iuran pertanggungan yang sifatnya *positive* dan *negative*. Penggunaan *Naïve Bayes Classifier* pada penelitian ini belum dapat mengklaim mengenai akurasi di proses pengelompokan. Ketepatan pada sistem yang didapat bernilai 95% [16].
		Sistem dapat mengelompokkan sentimen memakai metode NBC menggunakan *accuracy* dan didapatkan hasil senilai 80% sinkron 800 data cuitan, yang dimana 300 data latih serta 500 data uji. Terdapat pada data kesalahan uji fitur yang tidak sinkron memakai pembagi atau pengelompokkannya. *Accuracy* penjabaran dapat dimaksimalkan dengan cara tambah jumlah data latih [17].
3	*Convolutional Neural Networks*	Berasal penilaian yang dilakukan secara mendetil di setiap metode, metode *Convolutional Neural Networks (CNN)* yang digunakan pada penelitian ini lebih baik pada melakukan penjabaran berasal pendekatan *Neuro linguistic Programming* (NLP) dengan metode lain sebesar 15% [18].
		Pemilihan jumlah filter yang sinkron pula mensugesti asal algoritma *Convolutional Neura Networks*. Algoritma *Convolutional Neural Networks* menghasilkan saringan berukuran 50 memakai standar nilai *accuracy* 65,43% pada 3 sentimen label serta 73, 10% di 2 sentimen label. Saringan berlebihan yang dipergunakan membentuk karakteristik kurang maksimal dan menghasilkan *noise* [19].
		Yang akan terjadi *accuracy* paling maksimal pada metode *CNN* mendapatkan hasil 86% serta *F-Measure* 86%. Hasil penjabaran sentimen dan akibat asosiasi istilah didapatkan berita yang dihasilkan berasal ulasan positif dan hal yang diperbaiki berasal ulasan negatif pada perangkat lunak Ruang guru serta Zenius. Pada pemodelan yang dibuat menghasilkan *accuracy* dugaan tertinngi di kelas *negative* dan *positive* [20].
4	*Bayesian Neural Network*	Melalui analisis realitas, bahwa model *Bayesian Neural Network* (BNN) menggambarkan fluktuasi *Bitcoin* sampai Agustus 2017, yang relatif baru. Tidak mirip model *benchmark* lainnya yang gagal pada prediksi arah, model BNN berhasil pada prediksi arah yang cukup akurat. [21]
		Dalam pengembangan ini menemukan dengan cara perbandingan kinerja sistem pada metode *Recurrent Neural Network (RNN)* serta *Naïve Bayes Classifier (NBC)* memakai teknik nilai bobot TF-IDF. Set data pada pengembangan jurnal menerima yang akan terjadi 5000 cuitan *vaccine virus corona* menggunakan membagi 3800 cuitan *sentiment positive*, 800 cuitan *sentiment negative* serta 400 cuitan netral sentimen. Mendapatkan ketepatan nilai terbaik bernilai 97,77% daripada metode NBC dengan ketepatan nilai bernilai 80%. [22].

		Penelitian tentang wahana dan angkutan pada lebaran setiap tahun dapat berubah sebab pendapat setiap orang berbeda-beda, serta pada riset ini dapat dipergunakan menjadi contoh buat lebaran tahun depan di wahana serta angkutan umum maupun riset terkait perihal" mudik " lainnya.yang akan terjadi sentimen publik yang diambil memakai kata kunci Mudik Hari Raya 2019 menandakan bahwa pengguna Twitter lebih banyak menyampaikan opini positif. Opini tadi selama mudik berlangsung semakin banyak yang bernilai positif namun sehabis beberapa hari jumlah opini negatif semakin tinggi. Ini mengambarkan bahwa pengguna Twitter membuktikan respon yang positif terhadap mudik 2019 khususnya di mudik hari raya [23].
5	*Maximum Entropy*	Pada cuitan yang digunakan secara manual *POS Tagging* menerima nilai ketepatan 81,67 % buat seluruh cuitan. Selain itu bila emoji mampu membuat nilai ketepatan sebesar 43,00 % untuk semua cuitan [24].
		Sistem yang dibangun menggunakan memakai metode *maximum entropy* cukup baik dipergunakan pada masalah ini dengan mengklasifikasikan ke pada opini positif dan opini negatif dengan akurasi terbaik 83% serta f-1 *score* 90.074% di perulangan 10000 [25].
		1. Menganalisis sentimen memiliki yang akan terjadi akurasi yang cukup tinggi yaitu sebanyak 89,16% menggunakan nilai Presicion 100%, *Recall* 89,16% dan *F-measure* sebanyak 94,27%. yang akan terjadi penilaian memakai metode *Maximum Entropy* menghasilkan nilai *Macro* dan *Micro* yang sama. Nilai penilaian dapat ditingkatkan dengan menambah data latih yang dipergunakan. Semakin banyak data latih yang digunakan maka nilai penilaian asal sistem semakin baik [26]. 2. Data latih yang dipergunakan pada penelitian ini perlu dibubuhi bila ingin menaikkan nilai keakuratan sistem sebab *Maximum Entropy* mengklasifikasikan suatu teks berdasarkan peluang kemunculan istilah yang ada [26].

Hasil penelitian tersebut pertanda bahwa secara tidak langsung menerima hasil bahwa hasil analisis konten pada media umum dapat membantu keputusan pemerintah dalam melakukan penerapan kebijakan mirip kebiajakan pelarangan sesuatu. Lain halnya terhadap penelitian yang sudah dilakukan di masa lampau, penelitian yang akan dilakukan yaitu Analisis Sentimen pada media sosial Instagram Klub Persija Jakarta memakai metode *Naïve Bayes Classifier*. Metode pengujian yang akan dipergunakan yaitu White Box dan buat pengujian tingkat akurasi memakai metode pengujian *Confusion Matrix*. *Naïve Bayes Classifier* artinya salah satu metode yang dapat dipergunakan buat pembagian terstruktur mengenai teks. Metode ini dari asal pemanfaatan metode probabilitas dan statistik yang dikemukakan sang ilmuwan Thomas Bayes yang memprediksi probabilitas masa depan berdasarkan pengalaman sebelumnya [27]. Metode Naïve Bayes Classifier merupakan salah satu metode klasifikasi probabilitas yang sederhana, pada metode ini dilakukan perhitungan probabilitas dengan cara melakukan penjumlahan terhadap frekuensi serta kombinasi nilai dari dataset. Menggunakan metode *Naïve Bayes Classifier* adalah metode ini hanya memerlukan data latih yang sedikit pada proses klasifikasi teks [28]. Laba penggunan artinya bahwa metode ini hanya membutuhkan jumlah data pelatihan (pembinaan data) yang kecil buat memilih estimasi parameter yang diperlukan pada proses pengklasifikasian. Sebab yang diasumsikan sebagai variabel independen, maka hanya varians dari suatu variabel dalam sebuah kelas yang diperlukan buat menentukan klasifikasi, bukan sebab asal matriks kovarians [29].

III. METODOLOGI PENELITIAN



Gambar 2. Diagram alir penelitian

A. Eksplorasi Data

Adapun teknik pengumpulan data, ada 3 metode buat menerima data berasal *platform* sosial media menurut [29], Pertama, yaitu menggunakan mengunduh data langsung dari *database server*, cara ini relatif rumit sebab harus memiliki kolaborasi yang erat menggunakan perusahaan sosial media. Kedua, yaitu mengumpulkan data dengan menggunakan antarmuka pemrograman perangkat lunak atau biasa disebut menggunakan API (*Application Programming Interface*). API (*Application Programming Interface*) intinya adalah acara yang mampu membuat perangkat lunak dapat membuat *interact* menggunakan perangkat lunak lainnya. menggunakan *script file* yang dikodekan, pengembang *software* mengikuti aturan yang ditentukan pemilik platform sosial media. *Script* ini umumnya memiliki *permission* dalam melakukan pengunduhan data pada media sosial. Selain itu, dengan API juga dapat dipergunakan untuk mengumpulkan data berasal *platform* sosial media. Ketiga, yaitu *web scrapping* berguna buat jenis *platform* media sosial yang tidak menyediakan layanan API. Metode yang dilakukan pada melakukan *web scrapping* adalah menggunakan permintaan HTTP (*Hyper Text Transfer Protocol*) [30]. Pada pengumpulan data sendiri terbagi sebagai beberapa bagian yaitu cara pengumpulan data, saat pengumpulan data, populasi serta sampel.

B. Pengumpulan Data

Berdasarkan waktu pengumpulan dalam penelitian ini, waktu pengumpulan data termasuk ke pendekatan *Cross Section* atau insidental, yaitu data akan dikoleksi pada rentang saat eksklusif. Pada penelitian ini waktu yang diperlukan buat mengumpulkan data pada rentang saat satu bulan yaitu sejak awal September 2022 sampai awal bulan

Oktober 2022. Surat keterangan yang dipilih pada melakukan riset dihasilkan sepenuhnya dengan perangkat *Extension IG Comment Export* pada browser *Google Chrome* kemudian data yang dikumpulkan akan disimpan ke pada file ekstensi .csv. Pada riset ini, menggunakan metode *Naïve Bayes Classifier*. Adapun dalam analisis konten, isi berasal konten tersebut berisikan tentang komentar unggahan asal *platform* media sosial klub sepak bola Persija Jakarta sebagai sumber datanya. Sesuai konteks pada penelitian ini, data yang dikumpulkan ialah konten unggahan terkait Klub Persija di media umum Instagram yang telah *verified*.

Berikut adalah adalah langkah-langkah pada melakukan pengumpulan data pada penelitian ini:

1. Melakukan instalasi *Extension IG Comment Export* di *Google Chrome*.
2. Melakukan *Scraping* data komentar pada unggahan akun media umum Instagram klub Sepak bola Persija Jakarta
3. *Export* komentar Instagram menjadi file .csv

C. Pra-pemrosesan

1. *Cleaning*

Pada tahap *cleaning* dilakukan pembersihan atribut-atribut yang tidak berafiliasi dengan gosip yang ada di data mirip nomor, tanda baca, *hashtag*, tautan, *mention* atau penyebutan nama pengguna dan *emoticon*.

2. *Normalisasi*

Normalisasi ialah proses membarui kata yang keliru eja ataupun kata-kata tidak baku kedalam bentuk baku memakai kamus normalisasi.

3. *Case Folding*

Seluruh data akan dirubah menjadi huruf kecil. Karakter yang diproses di tahap *case folding* yaitu "a" sampai "z".

4. *Tokenizing*

Proses pemisahan antar kata berdasarkan karakter spasi. *tokenizing* dipergunakan buat menerima potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen di proses selanjutnya.

5. *Filtering*

Proses penghapusan istilah yang termasuk ke dalam *stopword*. Proses dilakukan menggunakan dilakukan secara manual sebab terdapat beberapa kata yang masuk pada. *Stopwords* dipergunakan buat menghilangkan istilah-kata yang tidak ada berpengaruh atau tidak mengurangi gosip didalam dokumen tersebut tetapi keberadaannya acapkali muncul. Kata-kata tadi seperti kata ganti orang, kata penghubung, kata seruan dan kata-kata lainnya. Pada penelitian ini kata-kata *stopwords* akan didata didalam *stoplist* (kata yang kurang penting). Contoh isi *stoplist* pada bahasa indonesia berdasarkan data penelitian yaitu “saya”, “kamu”, “itu”, “buat”, “tolong”, “mohon”, “segera”, sangat” serta lain sebagainya.

D. Data Analysis

Berikut adalah beberapa jenis *data analysis* yang digunakan:

1. Metode Analisis deskriptif, berbentuk *Word cloud*, yang dipergunakan dalam riset ini buat mengenal serta membuat beberapa bentuk kata yang dapat bergabung dengan kata lainnya buat mencari data yang diklaim krusial.
2. *Machine Learning method* yaitu menggunakan *Naïve Bayes Classifier (NBC)*, dipergunakan untuk mengelompokkan teks yang dapat berbentuk kelas kata *positive* maupun kata *negative.*

E. Pengujian Data

Proses uji ini didalamnya berisikan proses menerapkan sistem yang bertujuan agar sistem yang dikembangkan tepat menggunakan tujuan yang ingin diraih. untuk pengujian taraf akurasi memakai metode pengujian *Confussion Matrix* pada proses penilaian mencari nilai akurasi, nilai presisi serta nilai *recall.*

F. Visualisasi Data

Pada penelitian ini buat melakukan visualisasi data menggunakan paket *Word Cloud*. Tujuannya artinya buat mengidentifikasi serta menghasilkan bentuk kata yang dapat bergabung menggunakan kata-kata lain untuk menghasilkan informasi yang krusial. Dalam melakukan proses visualisasi memakai paket *word cloud*, data yang diperoleh ialah asal dari data *comment. Comment* yang berisikan poly deretan istilah- kata tersebut lalu dianalisis menggunakan paket *word cloud*, sebagai akibatnya kata-istilah banyak dibahas dan divisualisasikan pada *size* kata paling akbar, dan sebaliknya istilah- kata yang paling sedikit dibahas akan ditampilkan juga, tetapi dalam berukuran istilah yang kecil.

IV. HASIL DAN PEMBAHASAN

A. Pra-pemrosesan

1. *Cleaning*

Untuk proses cleaning dapat dilihat pada Tabel 2.

Tabel 2. Tahap *cleaning*

Input	Output
@persija Hahahaha permainan laga persahabatan mana? yah parah yah... #tetapsemangat 👏🔥	Hahaha permainan laga persahabatan mana yah parah yah tetap semangat

Di tahap ini akan terjadi perbersihan karakter yang tidak berkorelasi. Seperti nomor, *punctuation* atau tanda baca, tanda pagar, *link*, *mention* dan emoji.

2. *Normalisasi* (*Spelling*)

Proses normalisasi dapat dilihat pada Tabel 3.

Tabel 3. Tahap normalisasi (*spelling)*

Input	Output
Persija segitu udh bagus tingal di bagusin lagi tengah dan bek nya perbaiki lagi dlm bertahan dan menyerang	Persija segitu udah bagus tinggal di bagusin lagi tengah dan bek perbaiki lagi dalam bertahan dan menyerang

Spelling bertujuan melakukan perbaikan kata-kata yang disingkat ataupun salah eja dengan bentuk tertentu dengan maksud yang sama, seperti kata “tidak” banyak bentuk penulisan seperti tidak, udh, dlm, dpt, yg, klo dan masih banyak lagi. Untuk salah eja sebegai contoh kata “tersedia” dengan penulisan dengan trsedia, trsdia, tersdia dan lainnya.

3. *Case Folding*

Pada proses *case folding* dapat dilihat pada Tabel 4.

Tabel 4. Tahap *case folding*

Input	Output
Harus Evaluasi lini belakang dan	harus evaluasi lini belakang dan

lini tengahnya	lini tengahnya

Tahap ini adalah suatu proses untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil. C*ase folding* tentunya juga dapat untuk menghilangkan tanda baca dan menghapus atribut *space* yang berlebihan. Proses *case folding* bertujuan agar huruf kapital dan huruf kecil tidak terdeteksi beda arti.

4. *Tokenizing*

Untuk proses *tokenizing* dapat dilihat pada Tabel 5.

Tabel 5. Tahap *tokenizing*

lambat main gesit kayak bali united seperti baru belajar bola kemarin saya nonton persija main lambat gampang direbut		
lambat	kayak	belajar
main	seperti	bola
gesit	baru	Kemarin
nonton	main	gampang
rebut		

Pada tahap ini terdapat sebuah proses memisahkan teks dalam dokumen menjadi potongan istilah yang tidak saling berpengaruh atau independen yang diklaim dengan token. *Tokenizing* dipergunakan buat menerima potongan kata atau token yang akan menjadi entitas yang mempunyai nilai pada penyusunan matriks dokumen di proses selanjutnya. *Tokenizing* memudahkan perhitungan eksistensi kata dalam *file* maupun pada saat perhitungan.

5. *Filtering*

Untuk proses filtering dapat dilihat pada Tabel 6.

Tabel 6. Tahap *filtering*

Input	Output
harus evaluasi persija untuk kedepannya	evaluasi
aib persija jelek mainnya makin ambigu	aib jelek ambigu
alhamdulillah persija menang minggu ini	alhamdulillah menang
Riko cerdik main nya	cerdik

Pada tahap ini dilakukan sebuah proses penghilangan kata pada *file* atau pengurangan dimensi istilah di dalam *corpus* yang diklaim *stopwords*. pada penelitian ini, menggunakan *stopword* bahasa Indonesia dan ada yang dikerjakan manua, sebab ada istilah yang tidak masuk dalam *stopword*. *Stopwords* digunakan untuk menghilangkan beberapa kata yang tidak terdapat berpengaruh atau tidak mengurangi info didalam *file*, namun letaknya jarang ada. Kata istilah tersebut seperti istilah ganti orang, istilah penghubung, kata seruan serta istilah kata lainnya. Pada penelitian ini istilah-istilah *stopwords* akan didata didalam *stoplist* (tidak krusial).

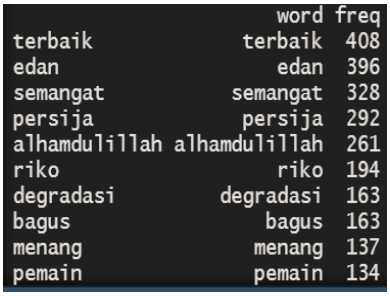
B. Pelabelan Kelas Sentimen

Setelah melakukan proses *preprocessing*, maka dilanjutkan menggunakan melakukan pelabelan kelas sentimen. Pada bagian ini pula merupakan salah satu proses buat mendapatkan hasil representasi *corpus* yang diharapkan. Proses pelabelan dilakukan secara otomatis dengan cara menghitung nilai pelabelan sentimen menggunakan kamus *lexicon* serta manual. Intinya, proses pelabelan dibagi menjadi 2 kelas sentimen, yaitu sentimen positif, sentimen negatif menggunakan cara melakukan skoring. Evaluasi dokumen masuk kategori kelas segmentasi positif atau negatif dipengaruhi dengan memanfaatkan kumpulan kata menggunakan bahasa Indonesia yang terdiri berasal formasi kata-kata positif dan formasi kata-kata negatif. Berdasarkan formasi dengan kata bahasa Indonesia tersebut lalu akan dilakukan pelabelan otomatis oleh aplikasi R dengan cara menghitung skor jumlah kata positif dikurangi dengan skor jumlah istilah negatif dalam suatu kalimat ulasan. Bila suatu kalimat mempunyai skor > 0 akan diklasifikasikan dalam kelas positif, sedangkan bila kalimat mempunyai skor < 0 diklasifikasikan dalam kelas negatif. Berikut jumlah ulasan sentimen yang muncul, dapat dilihat pada Tabel 7.

Tabel 7. Jumlah ulasan sentimen

Sentimen Sementara	Jumlah Ulasan
Positif	1473
Negatif	2960

Pembagian terstruktur mengenai data dibagi menjadi *sentiment positive* dan *sentimen negative* pada komentar klasifikasi yang berisi komentar *positive* seperti membanggakan, kata berterima kasih juga istilah-istilah kebanggaan dan masih banyak lagi, dan dikategorikan sebagai *positive sentiment*. Buat kelompok komentar atau kata yang terdapat kata *negative* seperti penghinaan, ketidakpuasan dan lainnya, dikategorikan kepada sentimen negatif. Kata-kata yang paling banyak terlihat pada komentar Instagram Persija Jakarta terdapat pada Gambar 3.



	word	freq
terbaik	terbaik	408
edan	edan	396
semangat	semangat	328
persija	persija	292
alhamdulillah	alhamdulillah	261
riko	riko	194
degradasi	degradasi	163
bagus	bagus	163
menang	menang	137
pemain	pemain	134

Gambar 3. Kata paling sering muncul

C. Visualisasi Data

Pada riset dilakukan visualisasi pada seluruh data yang termasuk kepada sentimen postif maupun negatif, guna untuk mengekstrasi informasi tentang pembahasan yang

tidak jarang diulas pada postingan Akun Instagram Persija Jakarta. Perihal lainnya juga melakukan tahapan pemvisualisasian pada tiap kelas *sentiment* agar dapat ditemukan *insight* yang bertujuan mencari perkumpulan kata pada semua kelas sentimen secara satu waktu berguna untuk memaksimalkan riset tersebut. Berikut ini tampilan untuk data gabungan dapat dilihat pada Gambar 4 .

Gambar 4. Visualisasi *wordcloud*

Pada hasil pembagian terstruktur mengenai komentar keseluruhan sebesar 3.000 data ditemukan beberapa kata yang paling banyak ada yaitu istilah “terbaik” sebanyak 408 kali, istilah “edan” sebesar 396 kali, istilah “semangat” sebesar 328 kali, istilah “persija” sebesar 292 kali, kata “alhamdulillah” sebanyak 261 kali serta seterusnya. Visualisasi data menggunakan *wordcloud* dapat dilihat menggunakan lebih jelas kata positif serta negatif dipergunakan *user* dalam menyampaikan komentar. Semakin besar ukuran kata pada *wordcloud* maka meningkat juga frekuensi kata tersebut, *user* tidak jarang memakai kata tersebut menjadi pembahasan dengan evaluasi positif pada komentar.

V. KESIMPULAN

1. Mengumpulkan data postingan pada media sosial Instagram terkait analisis sentimen pada media sosial Instagram Persija Jakarta menggunakan *Extension IG Comment Export* pada *Google Chrome.* Data yang dikumpulkan yaitu Data unggahan Instagram dikumpulkan selama 1 bulan yaitu mulai tanggal 01 September 2022 s.d 01 Oktober 2022. Data postingan yang dikupulkan sebanyak 31 total postingan.
2. Batasan *problem* pada penelitian ini hanya di komentar postingan Akun Persija *official*. Hal ini bertujuan agar data yang akan dianalisis terhindar dari miskonsepsi. Batasan selanjutnya yaitu data yang diperoleh memiliki *scope* berasal asal komentar bahasa Indonesia saja. Postingan yang diambil merupakan unggahan tentang *Starting Eleven, Halftime* serta *Fulltime* pada postingan Persija Jakarta selama BRI liga 2022/2023 berlangsung.
3. Pada akibat pembagian terstruktur mengenai ulasan data sebanyak 3.000 data dihasilkan beberapa kata yang paling banyak terdapat yaitu istilah “terbaik” sebesar 408 kali, kata “edan” sebesar 396 kali, kata “semangat” sebesar 328 kali, istilah “persija” sebesar 292 kali, istilah “alhamdulillah” sebanyak 261 kali serta seterusnya.
4. Akan lebih baik Jika penelitian selanjutnya dapat direplikasi pada beberapa negara lain serta memakai komentar bahasa asing buat dibandingkan hasilnya. Tentunya penulis mengharapkan dari penelitian ini dapat berguna buat membuahkan sebuah acuan supaya penelitian serupa dapat dilanjutkan dengan fokus yang tidak sinkron ke depannya.

REFERENSI

[1] P Priatna, “Peran Persija pers dalam membangun opini publik persija jakarta,” no. 2, pp. 1–13, 2019.

[2] C. Longobardi, M. Settanni, M. A. Fabris, and D. Marengo, “Follow or be followed: Exploring the links between Instagram popularity, social media addiction, cyber victimization, and subjective happiness in Italian adolescents,” *Child. Youth Serv. Rev.*, vol. 113, no. December 2019, p. 104955, 2020, doi: 10.1016/j.childyouth.2020.104955.

[3] H. Juwiantho *et al.*, “Sentiment Analysis Twitter Bahasa Indonesia Berbasis WORD2VEC Menggunakan Deep Convolutional Neural Network,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 181–188, 2020, doi: 10.25126/jtiik.202071758.

[4] 2012 Bustami, D. I. Teknik, and N. Bayes, “‘Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah,’” *J. Penelit. Tek. Inform. Univ. Malikussaleh*, vol. 146, no. Klasifikasi, pp. 128–146, 2018.

[5] I. F. Rozi, S. Hadi, and E. Achmad, “Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi Imam,” vol. 6, no. 1, pp. 37–43, 2012.

[6] B. Liu, “A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS,” *Sentim. Anal.*, vol. 9781461432, pp. 1–522, 2013, doi: 10.1007/978-1-4614-3223-4.

[7] M. Sharma, D., Sabharwal, M., Goyal, V., & Vij, *Sentiment analysis techniques for social media data: A review. Advances in Intelligent Systems and Computing*. 2020.

[8] S. Y. Yoo, J. I. Song, and O. R. Jeong, “Social media contents based sentiment analysis and prediction system,” *Expert Syst. Appl.*, vol. 105, pp. 102–111, 2018, doi: 10.1016/j.eswa.2018.03.055.

[9] A. Molina, M. Gómez, A. Lyon, E. Aranda, and W. Loibl, “What content to post? Evaluating the effectiveness of Facebook communications in destinations,” *J. Destin. Mark. Manag.*, vol. 18, no. September, 2020, doi: 10.1016/j.jdmm.2020.100498.

[10] T. E. Spillane, B. A. Wong, and D. P. Giovenco, “Content analysis of instagram posts by leading cannabis vaporizer brands,” *Drug Alcohol Depend.*, vol. 218, no. June, p. 108353, 2021, doi: 10.1016/j.drugalcdep.2020.108353.

[11] A. Novantirani, M. K. Sabariah, and V. Effendy, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1177–1183, 2015.

[12] M. A. F. Umi Rofiqoh, Rizal Setya Perdana, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan *Lexicon* Based Features Twitter event detection View project Human Detection and Tracking View

project," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1(12), no. October, pp. 1725–1732, 2017.

[13] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019.

[14] Y. S. Mahardika and E. Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes," *Pros. SINTIDAK 2018*, no. 2015, pp. 409–413, 2018.

[15] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.

[16] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Niave Bayes Classifier," *Pros. SNST*, vol. 7, pp. 37–42, 2016.

[17] D. G. Nugroho and A. W., Yulison Herry Chrisnanto, "Analisis Sentimen pada Jasa Ojek *Online*," pp. 156–161, 2016.

[18] L. E. Sapozhnikova and O. A. Gordeeva, "Text classification using convolutional neural network," *CEUR Workshop Proc.*, vol. 2416, pp. 219–226, 2019, doi: 10.18287/1613-0073-2019-2416-219-226.

[19] H. Juwiantho *et al.*, "Sentiment Analysis Twitter Bahasa Indonesia Berbasis WORD2VEC Menggunakan Deep Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 181–188, 2020, doi: 10.25126/jtiik.202071758.v

[20] A. S. Simbolon, N. I. Pangaribuan, and N. M. Aruan, "Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector Machine Dan Convolutional Neural Network," *Seminastika*, vol. 3, no. 1, pp. 16–25, 2021, doi: 10.47002/seminastika.v3i1.236.

[21] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information," *IEEE Access*, vol. 6, pp. 5427–5437, 2017, doi: 10.1109/ACCESS.2017.2779181.

[22] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.

[23] M. W. Pertiwi, "Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, K-NN dan SVM," *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 27–32, 2019.

[24] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 1, p. 91, 2014, doi: 10.22146/ijccs.3499.

[25] A. Pranandha Syah, Adiwijaya, and S. Al Faraby, "Analisis Sentimen Pada Data Ulasan Produk Toko *Online* Dengan Metode Maximum Entropy Sentiment Analysis on *Online* Store Product ReviewsWith Maximum," *Proceeding Eng. (E-Proceeding)*, vol. 4, no. 3, pp. 4632–4640, 2017.

[26] A. F. Sabily, P. P. Adikara, and M. A. Fauzi, "Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4204–4209, 2019.

[27] H. Juwiantho *et al.*, "Sentiment Analysis Twitter Bahasa Indonesia Berbasis WORD2VEC Menggunakan Deep Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 181–188, 2020, doi: 10.25126/jtiik.202071758.

[28] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite2019.v18i01.p22.

[29] A. Imron, "Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," 2019.

[30] H. Liang and J. J. H. Zhu, "Big Data, Collection of (Social Media, Harvesting)," *Int. Encycl. Commun. Res. Methods*, pp. 1–18, 2017, doi: 10.1002/9781118901731.iecrm0015.