

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



Disusun Oleh:

Nama : Andres Mondaref Jon

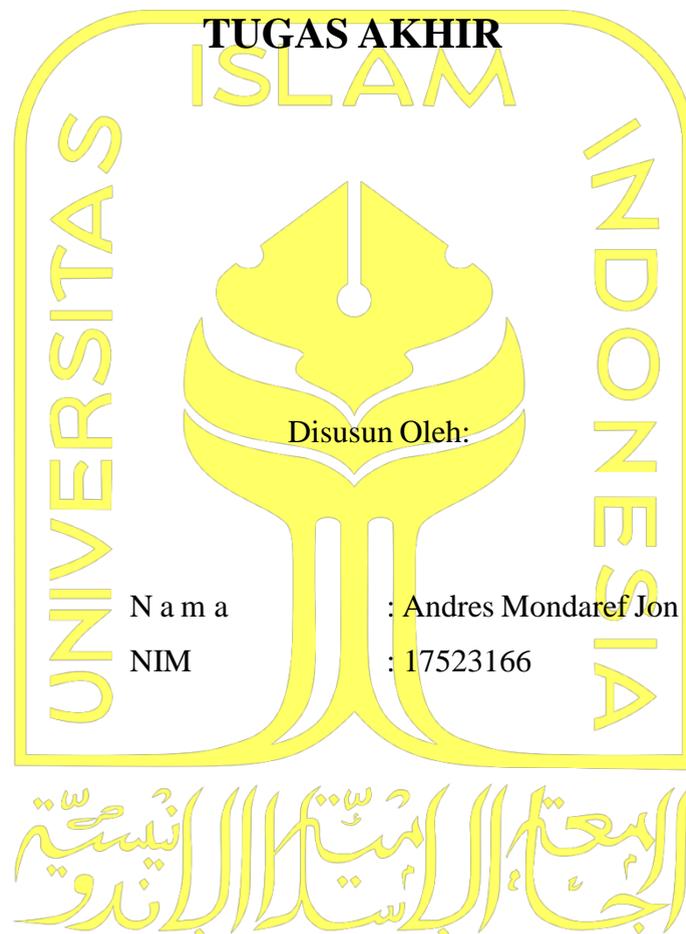
NIM : 17523166

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2023

HALAMAN PENGESAHAN DOSEN PEMBIMBING

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



Yogyakarta, 30 Oktober 2023

Pembimbing,

(Irving Vitra Papatungan, S.T., M.Sc., Ph.D.)

HALAMAN PENGESAHAN DOSEN PENGUJI

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

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia
 Yogyakarta, 30 Oktober 2023

Tim Penguji

Irving Vitra Papatungan, S.T., M.Sc.,
 Ph.D.

Anggota 1

DR. Syarif Hidayat, S.Kom., M.I.T.

Anggota 2

Sheila Nurul Huda, S.Kom., M.Cs.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dhomas Hatta Fudholi S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Andres Mondaref Jon

NIM : 17523166

Tugas akhir dengan judul:

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

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 30 Oktober 2023

A handwritten signature in black ink is written over a pink and white electronic stamp. The stamp features a QR code on the left, the Garuda Pancasila emblem in the center, and the text 'METERAI ELEKTRONIK 10000' and 'Rp10.000,00'.

(Andres Mondaref Jon)

HALAMAN PERSEMBAHAN

Alhamdulillah segala puji bagi Allah SWT Yang Maha Pengasih dan Maha Penyayang. Sholawat serta salam tidak lupa mari kita junjung tinggi kepada Nabi Muhammad SAW. Laporan tugas akhir ini di-dedikasikan sepenuhnya untuk kedua orang tua saya, ayah dan ibu tercinta atas segala dukungan dan doa yang selalu dipanjatkan setiap ibadah yang dijalankan. Kepada seluruh keluarga dan sahabat yang telah mendukung penulis dalam menyusun laporan tugas akhir baik secara langsung maupun tidak langsung. Terimakasih sudah mau belajar dan berkarya hingga saat ini.

Terimakasih untuk Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D. selaku dosen pembimbing, dan Seluruh Dosen Informatika Universitas Islam Indonesia atas bimbingan dan arahnya untuk selalu ber-inovasi tanpa batas serta yang telah mengajarkan ilmu-ilmu pengetahuan bermanfaat sebagai bekal kehidupan di masa yang akan datang.

Terimakasih untuk seluruh keluarga besar terutama Ayah dan Omak atas do'a, dorongan, dan motivasi yang diberikan kepada penulis selama ini. Terima kasih sudah memberi semangat terus menerus tiada henti hingga laporan tugas akhir ini selesai.

Semoga Allah SWT mengganti kebaikan-kebaikan kalian semua dengan kebahagiaan, keberkahan dan ladang pahala yang seluas-luasnya. Aamiin.

HALAMAN MOTO

“Balas dendam terbaik adalah menjadikan dirimu lebih baik.”

(Ali bin Abi Thalib)

"Raihlah ilmu dan untuk meraih ilmu, belajarlh untuk tenang dan sabar."

(Umar bin Khattab)

KATA PENGANTAR

Segala puji dan syukur bagi Allah SWT yang telah memberikan rahmat serta karunia- Nya, sehingga laporan tugas akhir ini dapat diselesaikan dengan judul “Analisis Sentimen Pada Media Sosial Instagram Klub Persija Jakarta Menggunakan Metode *Naive Bayes Classifier*. Shalawat dan salam tidak lupa tercurah kepada Rasulullah SAW.

Penyusunan laporan tugas akhir ini dimaksudkan untuk memenuhi sebagian syarat untuk menempuh gelar Sarjana pada Program Studi Informatika – Program Sarjana Fakultas Teknologi Industri di Universitas Islam Indonesia. Tanpa dukungan dari berbagai pihak yang telah membantu saya dalam menyelesaikan laporan tugas akhir yang sangat berarti ini dan tidak dapat terselesaikan apabila tanpa dukungan dari berbagai pihak tersebut. Oleh karena itu, saya ingin menyampaikan ucapan terima kasih kepada semua pihak yang berperan dalam penyusunan laporan tugas akhir ini terutama kepada:

1. Allah SWT yang telah memberikan kesehatan, rezeki, kesempatan dan karunia Nya sehingga penulis dapat menyusun dan menyelesaikan laporan tugas akhir.
2. Kedua orang tua yaitu ayah dan ibu saya yang telah membimbing dan membesarkan saya sehingga saya bisa menempuh pendidikan sampai sarjana tidak lupa berkat dukungan, motivasi, dan pengalaman hidup serta doa yang tiada henti-hentinya kepada saya.
3. Prof., Dr., Ir., Hari Purnomo M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
4. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
5. Bapak DThomas Hatta Fudholi S.T., M.Eng., Ph.D. selaku Kaprodi Informatika Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia.
6. Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D. selaku Dosen Pembimbing Tugas Akhir. Terima kasih atas bimbingan, saran, dukungan, ilmu selama penulisan laporan tugas akhir ini.
7. Bapak Hanson Prihantoro Putro, S.T., M.T. selaku Dosen Pembimbing Akademik Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
8. Seluruh Bapak/Ibu dosen Jurusan Informatika Fakultas Industri di Universitas Islam Indonesia yang telah memberikan pengetahuan dan ilmu yang sangat bermanfaat selamasaya menjalani masa perkuliahan.

9. Sahabat dan teman hidup di Himpunan Mahasiswa Teknik Informatika (HMTF) khususnya bidang Bakat dan Kreativitas (BATAS NORMAL), Alhamdulillah terimakasih atas waktu, pelajaran, pengalaman dan tentunya rasa kebersamaan yang melekat selama di Jogja.
10. Teman-teman satu rumput futsal (Cadangan Minared) yang tak bisa penulis sebutkan satu persatu, terimakasih untuk kebersamaan sekedar menyalurkan hobi positif di waktu luang yang membuat Yogyakarta dirindukan. Semoga futsal kita bisa diagendakan terus.
11. Seluruh sahabat saya selama di Yogyakarta yang selalu memberi dukungan dan semangat serta saran terhadap laporan tugas akhir yang saya kerjakan.
12. Seluruh teman-teman seangkatan yaitu Pixel Informatika 17 yang sudah banyak membantu saya dalam melewati berbagai kesulitan selama masa perkuliahan.
13. Semua pihak yang tidak bisa dituliskan satu persatu namun senantiasa mendukung baik secara langsung atau tidak langsung.

Bisa dikatakan laporan tugas akhir ini masih jauh dari kata sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang saya miliki. Oleh karena itu, saya menerima segala bentuk saran dan kritik yang membangun dari berbagai pihak untuk kebaikan bersama. Semoga laporan tugas akhir ini dapat bermanfaat bagi semua khalayak, terutama dalam bidang Informatika.

Yogyakarta, 30 Oktober 2023



(Andres Mondaref Jon)

SARI

Persija merupakan salah satu tim terbesar dan tim tersukses di Indonesia. Penggunaan media sosial pada klub sepakbola terjadi karena pergeseran budaya, perkembangan teknologi serta kebiasaan para suporter dalam mengakses informasi tentang klub mereka. Beberapa hal yang muncul pada komentar akun Instagram @persija berkaitan dengan perilaku suporter baik itu perilaku yang bersifat negatif maupun positif tentunya yang berpengaruh terhadap lingkungannya dan perilaku suporter selanjutnya. Biasanya salah satu persoalan sepele yang menyebabkan adu argumen antar suporter adalah karena adanya tim kalah, nyanyian rasis, aksi saling ejek, keputusan wasit yang tidak adil dan spanduk yang berisikan kata-kata mengumpat yang bersifat merendahkan klub lawan.

Naïve Bayes Classifier merupakan salah satu metode yang dapat digunakan untuk klasifikasi teks. Metode ini berasal dari pemanfaatan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Thomas Bayes yang memprediksi probabilitas masa depan berdasarkan pengalaman sebelumnya. Metode *Naïve Bayes Classifier* merupakan salah satu metode klasifikasi probabilitas yang sederhana, pada metode ini dilakukan perhitungan probabilitas dengan cara melakukan penjumlahan terhadap frekuensi dan kombinasi nilai dari dataset. Keuntungan lain menggunakan metode *Naïve Bayes Classifier* adalah metode ini hanya memerlukan data latih yang sedikit dalam proses klasifikasi teks.

Dalam rentang waktu 22 Maret 2021 – 17 September 2023 dalam 60 postingan media sosial Instagram Persija Jakarta ketika kalah dan menang didapatkan hasil analisis data komentar sebanyak 6.000 data ulasan. 3.000 data ulasan untuk postingan ketika Persija menang dan 3.000 data untuk kondisi Persija kalah. Hasil pengujian *confussion matrix* pada *data train* dan *data test* dengan perbandingan 90:10 menghasilkan tingkat akurasi paling optimal sebesar 82,51% ketika kondisi Persija Jakarta menang.

Kata kunci: Analisis Sentimen; Instagram; *Naïve Bayes Classifier*; *Confussion Matrix*; Persija;

GLOSARIUM

API	Antarmuka pemrograman aplikasi
Caption	Penjelasan atau keterangan yang di sisipkan pada sebuah unggahan
CSV	(<i>Comma Separated Values</i>) merupakan format data yang berisikan kumpulan basis data
Database	Kumpulan data yang dikelola secara sistematis
Generate	Proses menghasilkan instruksi yang berasal dari kode program
Insight	Suatu wawasan yang didapatkan dari topik tertentu
Klasifikasi	Metode untuk mengelompokkan data secara sistematis ke dalam kategori telah ditetapkan
Labelling	Proses pemberian label pada data berdasarkan kelas yang telah ditentukan
Platform	Media yang digunakan untuk menjalankan perangkat lunak
Preprocessing	Teknik awal dalam pengolahan data agar informasi data lebih bersih
Python	Bahasa pemrograman interpretatif yang bisa digunakan pada berbagai <i>platform</i> , khususnya <i>platform</i> yang berfokus pada keterbacaan kode
Lexicon	Istilah lain dalam penggunaan kata kunci
Verified	Informasi mengenai keaslian akun yang ditandai dengan centang biru
Word Cloud	Representasi visual dari frekuensi kata

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI.....	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTO.....	vi
KATA PENGANTAR.....	vii
SARI.....	ix
GLOSARIUM.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	15
1.1 Latar Belakang.....	15
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penulisan.....	6
BAB II KAJIAN PUSTAKA.....	8
2.1 Analisis Sentimen.....	8
2.1.1 Metode Analisis Sentimen.....	9
2.1.2 Analisis Sentimen Pada Media Sosial.....	10
2.2 Langkah Analisis Sentimen.....	11
2.2.1 <i>Text Pre-Processing</i>	11
2.2.2 <i>Word Cloud</i>	12
2.3 Penelitian Terkait.....	13
BAB III METODOLOGI PENELITIAN.....	18
3.1 Pengumpulan Data.....	19
3.1.1 Waktu Pengumpulan Data.....	20
3.1.2 Teknik Pengumpulan Data.....	20

	xii
3.1.3 Populasi.....	20
3.1.4 Sampel.....	21
3.2 Pra-pemrosesan	23
3.3 Pembobotan TF-IDF (<i>Term-Frequency Times Inverse Document Frequency</i>)	24
3.4 Analisis Data	25
3.5 Pengujian Data	25
3.6 Visualisasi Data	26
3.7 Perancangan Pengujian	27
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Pengumpulan Data.....	28
4.2 Implementasi <i>Pre-Processing</i>	29
4.3 Pelabelan Kelas Sentimen.....	36
4.4 Klasifikasi dengan Metode Naïve Bayes Classifier.....	38
4.5 Visualisasi Data	41
4.5.1 Word Cloud Positif	42
4.5.2 Word Cloud Negatif.....	44
4.6 Pengujian Analisis Sentimen	45
BAB V KESIMPULAN DAN SARAN	47
5.1 Kesimpulan	47
5.2 Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN.....	52

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terkait	13
Tabel 4.1 Hasil <i>Cleaning</i> pada <i>comment text</i>	30
Tabel 4.2 Hasil <i>Case Folding</i> pada <i>comment text</i>	31
Tabel 4.3 Hasil <i>Tokenizing</i> pada <i>comment text</i>	32
Tabel 4.4 Hasil <i>Filtering</i> pada <i>comment text</i>	33
Tabel 4.5 Contoh kata <i>slang</i> dan perbaikannya	34
Tabel 4.6 Contoh kata singkatan dan perbaikannya.....	34
Tabel 4.7 Contoh kata <i>typo</i> dan perbaikannya	35
Tabel 4.8 Contoh kata asing dan perbaikannya	35
Tabel 4.9 Contoh karakter berulang dan perbaikannya	35
Tabel 4.10 Hasil Pelabelan kelas data	37
Tabel 4.11 Jumlah ulasan setiap kelas sentimen pada postingan Persija menang.....	38
Tabel 4.12 Jumlah ulasan setiap kelas sentimen pada postingan Persija Kalah.....	38
Tabel 4.13 Perbandingan data latih dan data uji ketika Persija Menang.....	39
Tabel 4.14 Perbandingan data latih dan data uji ketika Persija Kalah	39
Tabel 4.15 Hasil Akurasi data uji.....	40

DAFTAR GAMBAR

Gambar 2.1 Metode Analisis Sentimen	9
Gambar 3.1 Diagram Alir Analisis Sentimen	18
Gambar 3.2 Instalasi <i>Extension IG Comment Export</i>	21
Gambar 3.3 Cara Salin Tautan Pada Unggahan Persija	22
Gambar 3.4 <i>Copy link</i> Unggahan ke <i>Extension IG Comment Export</i>	22
Gambar 3.5 <i>Download</i> data komentar menjadi file.csv	22
Gambar 3.6 File csv yang berhasil di <i>download</i>	23
Gambar 4.1 Menginstall <i>Library</i>	28
Gambar 4.2 Mengimport <i>Library</i>	28
Gambar 4.3 <i>Source Code</i> koneksikan <i>Google Drive</i> ke <i>Google Colab</i>	29
Gambar 4.4 Contoh data ulasan hasil <i>Scraping</i>	29
Gambar 4.5 Kode Program Tahap <i>Cleaning</i>	30
Gambar 4.6 Kode program tahap <i>case folding</i>	31
Gambar 4.7 Kode program tahap <i>Tokenizing</i>	31
Gambar 4.8 Kode Program Tahap <i>Filtering</i>	32
Gambar 4.9 Kode program proses <i>Stemming</i>	34
Gambar 4.10 Kode Program proses pelabelan data	37
Gambar 4.11 Kode Program jumlah ulasan setiap sentimen	38
Gambar 4.12 Proses membagi data latih dan data uji.....	39
Gambar 4.13 Kode Program Proses <i>Confussion Matrix</i> pada <i>Data Train dan Data Test</i>	40
Gambar 4.14 Visualisasi <i>Word Cloud</i> Ketika Persija Menang.....	41
Gambar 4.15 Visualisasi <i>Word Cloud</i> Ketika Persija Kalah	42
Gambar 4.16 Visualisasi <i>Word Cloud</i> Positif ketika persija Menang	43
Gambar 4.17 Visualisasi <i>Word Cloud</i> Positif ketika persija Kalah	43
Gambar 4.18 Visualisasi <i>Word Cloud</i> Negatif ketika Persija Menang.....	44
Gambar 4.19 Visualisasi <i>Word Cloud</i> Negatif ketika Persija Kalah	44
Gambar 4.20 Pengujian Analisis sentimen	45
Gambar 4.21 Uji coba sampel data ketika Persija Menang	46
Gambar 4.22 Uji coba sampel data Negatif pada Persija Kalah	46

BAB I

PENDAHULUAN

1.1 Latar Belakang

Olahraga merupakan salah satu unsur yang berpengaruh dalam kehidupan manusia yang ikut berperan dalam mengharumkan nama daerah dan bangsa baik melalui kompetisi regional, nasional maupun internasional. Setiap bangsa dari seluruh dunia berlomba-lomba menciptakan prestasi dalam kegiatan olahraga karena prestasi olahraga yang baik akan meningkatkan citra daerah, bangsa di dunia internasional (Riefky & Pramesti, 2020). Ada banyak jenis cabang olahraga yang muncul atau tercipta dari setiap negara yang salah satunya adalah Sepak bola. Saat ini Sepak bola menjadi salah satu olahraga yang diminati dan digemari oleh hampir semua orang di bumi. Sepak bola juga bisa dinikmati oleh semua kalangan usia dan lapisan masyarakat. Hingga kini, tidak ada olahraga yang mendapatkan sambutan paling meriah dan gegap gempita dari masyarakat di berbagai negara dunia selain sepak bola.

Di Indonesia, permainan sepak bola diperkenalkan oleh bangsa Hindia Belanda yang datang dan bekerja di instansi-instansi pemerintahan Hindia Belanda sebagai pegawai dalam perkebunan, kantor-kantor perdagangan, perkapalan, dan pertambangan. Mereka memilih permainan yang tengah populer di Eropa sebagai sarana rekreasi dan menjaga kebugaran. Sepak bola adalah olahraga yang terkenal dan paling luas penyebarannya, oleh karena itu jumlah klub sepak bola semakin meningkat. Puncaknya pada tanggal 19 April 1930 di Yogyakarta, terbentuk panitia organisasi sepak bola yang bernama Persatuan Sepak Raga Seluruh Indonesia (PSSI) dengan ketua Ir. Soeratin Sosrosoegondo. Pembentukan PSSI pada dasarnya merupakan bentuk perlawanan terhadap *Nederland Indische Voetball Bond* (NIVB) yang merupakan bentukan dari Belanda. Pada awalnya olahraga dalam hal ini sepak bola hanya diperbolehkan untuk anggota NIVB menggunakan fasilitas lapangan, namun tidak untuk pemain bumiputera yang hanya ditempat seadanya (Wicaksono, 2019).

Pada tanggal 19 April tahun 1930, terdapat klub sepak bola diantaranya Persatuan Sepakraga Mataram (PSM)(sekarang PSIM Yogyakarta) bersama VIJ Jakarta (sekarang Persija Jakarta), *Bandoeng Indonesische Voetbal Bond* (BIVB) (sekarang Persib Bandung), *Vortenlandsche Voetbal Bond* (VVB) (sekarang Persis Solo), *Madionsche Voetbal Bond* (MVB) (sekarang Madiun Putera Fc), *Indonesia Voetbal Bond Magelang* (IVBM) (sekarang

PPSM Magelang), *Soerabajashe Indonesische Voetbal Bond* (SIVB) (sekarang Persebaya Surabaya) turut membidani kelahiran PSSI dalam pertemuan yang diadakan di Sosiet Handeprojo Yogyakarta (Sekarang Gedung Batik). Sejak tahun itu pula kompetisi tahun antar kota/perserikatan diselenggarakan (Wicaksono, 2019). Dari pemaparan diatas dapat disimpulkan topik yang diangkat pada penelitian ini adalah salah satu klub sepak bola yaitu *Voetbalbond Indonesische Jacatra* (VIJ) atau sekarang lebih dikenal sebagai Persija Jakarta.

Voetbalbond Indonesische Jacatra (VIJ) lahir pada tanggal 28 November 1928 yang pada akhirnya nama VIJ diganti Persija (persatuan sepak bola Indonesia Jakarta) ditahun 1950. Sebelum mengganti nama ke Persija, VIJ berhasil merebut empat piala liga perserikatan Indonesia pada tahun 1931, 1933, 1934, dan 1938. Hingga VIJ mengganti nama ke Persija Jakarta pada tahun 1950. Enam gelar piala Liga Perserikatan dan liga Indonesia berhasil dimasukkan kelemari trofi Persija (P Priatna, 2019). Keberadaan Persija Jakarta dalam kancah sepakbola Indonesia yaitu bermain dalam Liga 1 Indonesia memberikan warna tersendiri, bukan hanya oleh permainannya yang menawan tetapi juga pada kelompok suporternya yaitu The Jakmania.

Persija Jakarta memiliki supporter fanatik yaitu The Jakmania, merupakan kelompok pendukung atau supporter kesebelasan sepak bola Persija Jakarta yang merupakan salah satu supporter terbaik di Indonesia yang berdiri sejak Ligin IV, tepatnya pada tanggal 19 Desember 1997. Markas dan Sekretariat Jakmania berada di Jl. H.R. Rasuna Said. Fanatisme ini menciptakan atmosfer dukungan yang kuat untuk Persija di stadion maupun di media sosial, ini dibuktikan dengan dinobatkan The Jakmania menjadi kehadiran penonton terbanyak Liga 1 Indonesia selama 5 tahun terakhir. Pengertian kelompok dari segi persepsi berdasarkan asumsi yaitu bahwa anggota kelompok sadar dan mempunyai persepsi bersama akan hubungan mereka dengan anggota lain. The Jakmania sendiri mengalami pasang surut dengan jumlah massa yang besar, masalah-masalah sering kali bermunculan. Perseteruan dengan kelompok supporter klub lain pun muncul.

Pemberitaan mengenai Persija tidak terlepas dari peranan media dari era perserikatan hingga sekarang. Pemberitaan Persija pertama kali di media cetak pada tahun 1938 di surat kabar Pemandangan. Hingga ditahun 2000-an supporter Persija Jakarta mendirikan portal beritaonline dengan nama JakOnline di tahun 2001 sebagai wadah pemberitan khusus untuk pecinta Persija Jakarta dan The Jakmania (P Priatna, 2019). Sesuai dengan pengoptimalan dari perkembangan teknologi, klub sepakbola Persija Jakarta menginformasikan dan

memublikasikan seluruh kegiatannya melalui *website* www.persija.co.id. Hal tersebut merupakan cara yang tepat untuk menjawab kebutuhan akan informasi serta menunjukkan eksistensi klub Persija Jakarta kepada penggemarnya.

Di sisi lain, penggunaan aplikasi sosial media Instagram merupakan aplikasi yang sangat populer digunakan (Longobardi et al., 2020). Dengan pengguna yang beragam dari berbagai latar belakang, ini menjadikannya subjek yang menarik untuk memahami dinamika komunikasi digital, interaksi antarindividu, dan pengaruh sosial media pada masyarakat. Instagram memiliki potensi untuk memengaruhi perilaku dan persepsi penggunanya. Oleh karena itu, Instagram dapat digunakan dalam penelitian untuk memahami dampak sosial dan psikologis dari media sosial pada individu dan kelompok. Persija Jakarta memiliki akun media sosial instagram resmi yaitu @persija. Berdasarkan data terakhir yang peneliti temukan pada 30 November 2023, akun Persija Jakarta telah berhasil menjaring *followers* sebanyak lebih dari 3,9 juta *followers* dan menempati sebagai klub sepakbola nomor 2 di Indonesia dengan *followers* Instagram terbanyak, dan telah mengupload postingan sebanyak 13,2 ribu postingan (<https://www.instagram.com/persija/>).

Dalam pemanfaatan media sosial sebagai sarana *digital branding*, hal yang harus di perhatikan adalah mengenai *engagement*. Secara sederhana *engagement* yaitu sebuah tolak ukur guna mengetahui tingkat keterlibatan pelanggan yang berupa interaksi sosial terhadap konten media sosial yang kita punya dengan tujuan meningkatkan popularitas dan pendapatan. Untuk meraih interaksi sosial yang tinggi, Persija memanfaatkan pengelolaan konten Instagram dengan maksimal agar para pengikut Persija dapat memberikan *engagement* berupa *like*, *comment*, *repost*, dan *share*. Dibandingkan sosial media Persija Jakarta yang lain seperti Facebook dan Twitter contohnya. Sosial media Facebook hanya menjaring *followers* sebanyak 1,7 juta pengikut dan media sosial Twitter sebanyak 3,2 juta pengikut dan untuk media sosial Instagram menjaring *followers* sebanyak 3,9 juta pengikut lebih banyak dibandingkan yang lainnya. Tidak hanya dari segi *followers*, Instagram Persija Jakarta juga mengungguli dari segi *Insight* seperti total *like* dan komentar pada setiap postingan.

Dari statistik yang dihimpun oleh Gery Maulana Thiar sebagai pengamat sepakbola, mengatakan bahwa Persija Jakarta menempati nomor 1 dari 18 klub Liga 1 Indonesia dengan 3 interaksi dan *engagement* tertinggi. Meskipun lewat jumlah *followers* Persija Jakarta hanya

menduduki peringkat ke-dua, namun Persija Jakarta unggul dalam statistik jumlah total interaksi, dengan jumlah 123,58 juta (Persija.id, 2020). Persija Jakarta juga menjadi nomor satu dalam kategori klub paling produktif di Instagram dalam sisi konten pada tahun 2020 dengan 1.694 postingan (Persija.id 2020). Raihan *engagement* Persija Jakarta juga berlanjut di awal tahun 2022. Lewat survey yang dilakukan oleh Deportez & Finanzas selama bulan Januari 2022, Persija Jakarta menempati peringkat ke-2 sebagai tim sepakbola di Asia dengan jumlah interaksi sosial tertinggi pada *platform* Instagram dengan total 17,8 Juta interaksi (Roy et al., 2022).

Dari segi Sejarah dan tradisi, prestasi, supporter fanatik dan media sosial sudah dijelaskan bahwa Persija Jakarta adalah salah satu klub terbesar yang ada di Indonesia. Secara garis besar sebagai klub dari ibu kota negara, Persija Jakarta mungkin memiliki akses yang lebih besar ke sumber daya finansial, fasilitas pelatihan, dan infrastruktur yang mendukung pengembangan klub. Kehadiran geografis ini dapat memberikan keuntungan dalam hal daya tarik pemain, dukungan sponsor, dan cakupan media. Persija Jakarta sebagai klub yang besar di Indonesia, memiliki kemampuan finansial untuk mendatangkan pemain-pemain berkualitas, termasuk pemain asing. Sebagai klub yang berbasis di ibu kota tentunya, Persija Jakarta mendapatkan cakupan media yang lebih luas daripada klub lain dan memiliki pengaruh yang signifikan dalam komunitas sepakbola Indonesia, ini juga yang membuat Persija lebih dilirik daripada klub lain di Liga Indonesia. Dalam penelitian ini, peneliti memilih Persija Jakarta sebagai target penelitian karena Persija merupakan salah satu tim terbesar dan tim tersukses di Indonesia. Persija juga memiliki musuh bebuyutan yang selalu disebut *derbi elclasico* yaitu Persib Bandung. 2 klub ini tidak hanya di Indonesia selalu punya daya tarik hingga dunia ketika mereka bertemu di Liga 1 Indonesia. Daya tarik ini lah mengapa alasan peneliti untuk memilih Persija Jakarta sebagai objek penelitian.

Penggunaan media sosial pada klub sepakbola terjadi karena pergeseran budaya, perkembangan teknologi serta kebiasaan para supporter dalam mengakses informasi tentang klub mereka. Beberapa hal yang muncul pada komentar akun Instagram @persija berkaitan dengan perilaku supporter baik itu perilaku yang bersifat negatif maupun positif tentunya yang berpengaruh terhadap lingkungannya dan perilaku supporter selanjutnya. Ketika Persija Jakarta sedang mengalami kekalahan, komentar di Instagram bisa mencerminkan berbagai macam respon baik maupun buruk dari supporter.

Respon atau komentar pada Instagram Persija Jakarta ini bisa bervariasi dari kekecewaan kepada klub, kritik kinerja pelatih ataupun pemain, dukungan dan optimisme untuk pertandingan selanjutnya, humor atau sarkasme ringan, komentar yang mengarah positif dan negatif dan bisa juga karena salah satu persoalan sepele yang menyebabkan adu argumen antar supporter adalah karena adanya tim kalah, nyanyian rasis, aksi saling ejek, keputusan wasit yang tidak adil dan spanduk yang berisikan kata-kata mengumpat yang bersifat merendahkan klub lawan. Dari berbagai komentar penggemar inilah yang akan ditentukan sentimen termasuk kedalam sentimen positif, sentimen negatif, maupun sentimen netral.

Pada umumnya terdapat beberapa metode yang biasa digunakan untuk analisis sentimen terutama dalam klasifikasi, antara lain *Support Vector Machine* dan *Naive Bayes*. Penelitian dengan algoritma *Support Vector Machine* dilakukan (Naradhipa & Purwarianti, 2012) dalam klasifikasi pesan berbahasa Indonesia pada sosial media menghasilkan kesimpulan bahwa metode *Support Vector Machine* memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks. Sementara metode *Naive Bayes* memiliki tingkat akurasi tertinggi dibandingkan metode *K-Nearest Neighbor* dan metode gabungan *K-means* dan *Learning Vector Quantization* dalam mengkategorikan buku komputer berbahasa Indonesia. Pada dasarnya *Support Vector Machine* bekerja dengan cara mencari *hyperplane* terbaik yang memisahkan dua buah kelas, sedangkan *Naive Bayes* bekerja dengan cara mencari nilai probabilitas bersyarat terbesar dari masing-masing kelas.

Definisi lain mengatakan *Naive Bayes Classifier* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Bustami et al., 2018). Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes Classifier* juga cepat dalam mengklasifikasikan teks, tepat digunakan untuk analisis sentimen pada media sosial atau analisis umpan berita, selain itu juga belum ada penelitian analisis sentimen yang dilakukan tentang olahraga khususnya sepak bola di Indonesia menggunakan metode *Naive Bayes Classifier*. Masih banyak metode analisis sentimen selain *Naive Bayes Classifier*, metode ini dipilih karena memiliki akurasi tinggi diantara metode lainnya pada analisis sentimen melalui media sosial.

Berdasarkan berbagai penjelasan latar belakang diatas, penelitian yang akan diangkat untuk tugas akhir adalah “Analisis Sentimen Pada Media Sosial Instagram Klub Persija Jakarta Menggunakan Metode *Naive Bayes* ”.

1.2 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang, dapat diambil suatu rumusan masalah, yaitu bagaimana menerapkan metode *Naïve Bayes Classifier* dalam analisis sentimen pada media sosial Instagram klub sepakbola Persija Jakarta.

1.3 Batasan Masalah

Penetapan batasan-batasan masalah dalam penelitian dilakukan dengan tujuan agar penelitian yang dilakukan lebih terarah. Adapun batasan masalah dalam penelitian ini sebagai berikut:

- a. Data yang digunakan merupakan data komentar yang diambil dari postingan media sosial *Instagram* klub Persija Jakarta dalam rentang waktu 22 Maret 2021 – 17 September 2023.
- b. Penelitian ini hanya berfokus pada komentar atau *text* pada 60 *match* (pertandingan) dengan jumlah 6.000 data komentar ketika Persija Jakarta bertanding menang dan kalah, untuk postingan diambil 30 postingan (3.000 data) ketika Persija Jakarta menang dan 30 postingan (3.000 data) ketika Persija Jakarta kalah.
- c. Metode yang digunakan untuk analisis sentimen adalah *Naïve Bayes Classifier*.

1.4 Tujuan Penelitian

Hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data komentar pada postingan akun Instagram Persija Jakarta menjadi sentimen positif, netral dan sentimen negatif.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah menjadi referensi bagi penelitian berikutnya yang relevan mengenai media sosial *Instagram* dengan tema yang terkait. Penelitian ini juga diharapkan dapat memberikan beberapa gambaran tentang bagaimana cara menganalisis sentimen menggunakan metode *Naïve Bayes Classifier* dari media sosial *Instagram* khususnya tentang sepakbola dan pada topik-topik lainnya.

1.6 Metodologi Penelitian

Terdapat beberapa tahapan yang akan dilakukan dalam proses metodologi penelitian pada komentar akun Instagram Persija Jakarta. Tahapan yang dilakukan sebagai berikut:

a. *Data Selection*

Tahapan awal dalam proses *data mining* ini adalah proses penentuan atau pemilihan data yang akan digunakan sesuai atau berkaitan dalam proses klasifikasi.

b. *Data Cleaning (Pre-processing)*

Pada tahap ini merupakan tahap *pre-processing*, dimana data pada penelitian ini dilakukan pembersihan agar data lebih rapi dan terstruktur. Pembersihan data dilakukan adalah menghapus atribut yang tidak diperlukan dalam proses klasifikasi dan prediksi.

c. *Data Labelling (Pelabelan Data)*

Pada tahap ini setelah melalui proses *pre-processing*, data akan dilabeli kedalam sentimen positif, sentimen negatif, atau sentimen netral.

d. *Klasifikasi Naïve Bayes Classifier*

Tahapan ini adalah proses menerapkan metode dari *data mining* untuk mengolah data yang ada, metode yang digunakan adalah *Naïve Bayes Classifier* dan untuk *output* nya adalah nilai *accuracy*, *precision* dan *recall*.

e. *Data Testing (Pengujian Data)*

Pada tahap ini adalah proses prediksi atau pengujian data. *Data testing* dilakukan bertujuan untuk memvalidasi atau menganalisis data ulasan, apakah ulasan tersebut benar termasuk ke dalam sentimen positif, sentimen negatif, maupun sentimen netral.

1.7 Sistematika Penulisan

Sistematika penulisan pada laporan tugas akhir penelitian ini dimuat dalam 5 bab, masing-masing bab menjelaskan informasi terstruktur yang berfungsi untuk memudahkan pembaca untuk mendapatkan informasi yang disampaikan oleh penulis, sistematika pembahasan terdiri dari:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan dalam penelitian analisis sentimen pada media sosial klub sepak bola Persija Jakarta menggunakan metode *Naïve Bayes Classifier*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi bagan alur penelitian beserta penjelasannya, teknik yang digunakan atau metode yang digunakan dalam penelitian serta data yang akan digunakan sebagai data kajian dalam analisis sentimen pada media sosial Instagram klub Persija Jakarta.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan pembahasan mengenai hasil pengolahan data serta batasannya yang telah dilakukan dengan mengacu pada teori dan alur penelitian yang telah dijelaskan sebelumnya sehingga dapat memberikan rekomendasi maupun usulan dalam menyelesaikan masalah.

BAB V KESIMPULAN DAN SARAN

Berisi tentang kesimpulan akhir dari analisis sentimen pada media sosial Instagram klub sepakbola Persija Jakarta serta saran atau perbaikan pada penelitian berikutnya.

BAB II

KAJIAN PUSTAKA

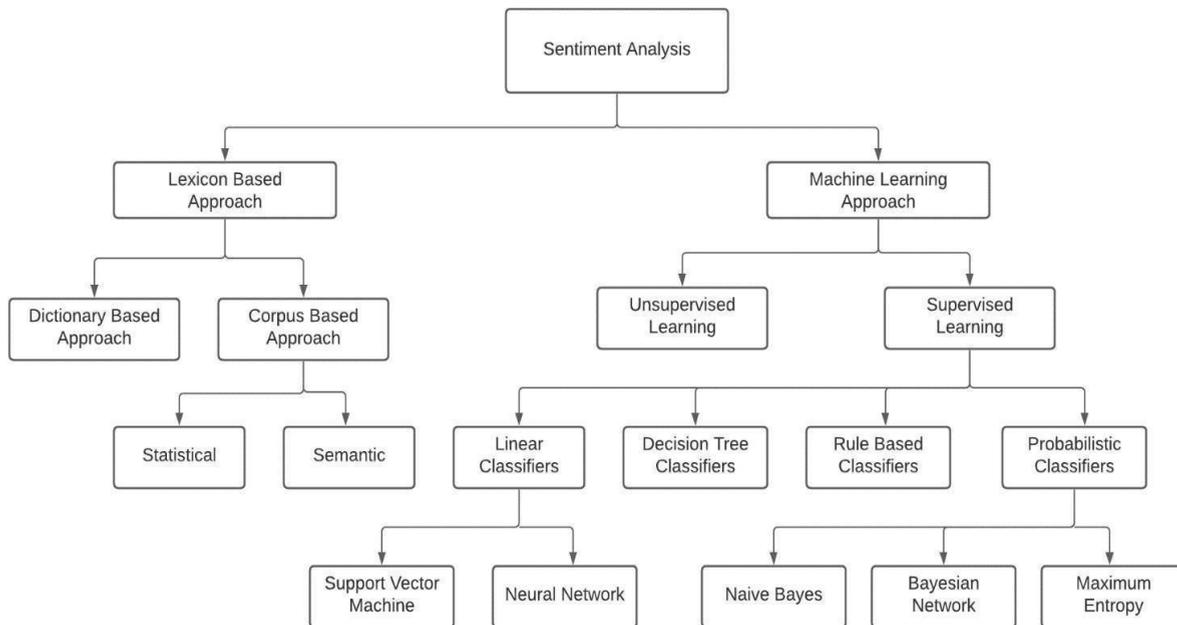
2.1 Analisis Sentimen

Analisis sentimen adalah metode analisis yang digunakan untuk menentukan sentimen audiens mengenai reaksi atau pandangannya terhadap suatu fenomena yang terjadi. Analisis sentimen atau *opinion mining* adalah proses dalam memahami, melakukan proses ekstraksi dan juga mengolah data tekstual (data berupa teks) secara otomatis dengan tujuan untuk mendapatkan informasi sentimen yang terdapat dalam suatu kalimat opini. Analisis sentimen juga dilakukan untuk melihat pendapat atau tendensi (kecenderungan) opini terhadap suatu permasalahan atau objek dari seseorang, apakah pendapat tersebut cenderung berpandangan ke arah opini negatif atau ke arah opini positif (Rozi et al., 2012).

Analisis sentimen digunakan untuk klasifikasi ulasan. Klasifikasi ulasan yang dilakukan pada penelitian tersebut adalah klasifikasi ulasan terhadap proses bisnis di perguruan tinggi. Opini publik tersebut dominan mengarah kepada ulasan campuran (ulasan positif dan ulasan negatif) mereka mengenai proses bisnis yang ada di perguruan tinggi. Fokus utama dari *opinion mining* di dalam proses melakukan analisis sentimen yaitu melakukan analisis opini dari suatu dokumen yang berbentuk tekstual. Analisis sentimen juga dapat dilakukan dalam melakukan pelacakan opini di dalam diskusi *online*.

Di sisi lain, analisis sentimen digunakan untuk menentukan sentimen seseorang (bisa berupa reaksi dan kecenderungan) mereka apakah mengarah kepada sentimen positif ataupun sentimen negatif, semua itu tergantung kepada reaksi dan pandangan mereka terhadap fokus objek tertentu (Liu, 2013). Selain itu, analisis sentimen juga dapat melakukan ringkasan ulasan, proses ekstraksi (pemisahan) antara sinonim (persamaan kata) dan antonim (lawan kata) dari suatu dokumen yang berbentuk tekstual.

2.1.1 Metode Analisis Sentimen



Gambar 2.1 Metode Analisis Sentimen

Sumber: (Sharma, D., Sabharwal, M., Goyal, V., & Vij, 2020)

Metode dalam melakukan analisis sentimen banyak ragamnya. Metode analisis sentimen berbasis *machine learning* (pembelajaran mesin) dan berbasis *lexicon* (leksikon atau kosa kata). Metode analisis sentimen berbasis pembelajaran mesin dibagi menjadi dua yaitu *supervised learning* dan *unsupervised learning* sedangkan metode analisis sentimen berbasis leksikon dibagi menjadi dua yaitu *dictionary based* dan *corpus based*. Pada metode pembelajaran mesin yaitu *supervised learning*, pendekatan ini digunakan jika ada data berlabel yang telah tersedia untuk dilatih modelnya. Langkah-langkah yang dilakukan dalam *supervised learning* adalah melatih model dan sisa lainnya adalah melakukan prediksi. Pada *unsupervised learning*, pendekatan ini digunakan jika realibilitas data berlabel diposisikan sulit. Kalimat yang dikoleksi akan dikategorikan berdasarkan kata kunci pada setiap kategori. Untuk menganalisis data yang bersifat dependen (bergantungan) maka pendekatan *unsupervised learning* lebih mudah untuk digunakan.

Selanjutnya, pada metode leksikon yaitu *dictionary based*, pendekatan ini digunakan untuk mencari sinonim kata dengan tujuan untuk memperbesar ukuran pada kumpulan kata yang sama. Selain itu opini kata-kata opini yang unik akan diperluas sebagai fitur untuk

melakukan analisis sentimen. Pada pendekatan *corpus based* ditemukan bahwa sebuah kata memiliki orientasi konteks. Pendekatan berbasis korpus (struktur bahasa) dilakukan menggunakan dua cara yaitu pendekatan berbasis statistik (data angka) dan pendekatan semantik (makna kata). Pada penelitian tentang metode analisis sentimen yang dilakukan oleh (Sharma, D., Sabharwal, M., Goyal, V., & Vij, 2020) tersebut menyebutkan bahwa metode analisis sentimen yang akan dilakukan oleh peneliti haruslah menyesuaikan dengan fokus dan lingkup penelitian tentang analisis sentimen.

2.1.2 Analisis Sentimen Pada Media Sosial

Berbicara mengenai analisis sentimen pada media sosial, analisis sentimen di media sosial dapat dilakukan dengan cara mengklasifikasikan sentimen positif dan sentimen negatif berdasarkan unggahan yang ada di media sosial, artinya unggahan yang memiliki reaksi positif ditempatkan pada sentimen positif sedangkan unggahan yang memiliki reaksi negatif ditempatkan pada sentimen negatif. Pada proses pengklasifikasiannya, polaritas (hal yang bersifat berlawanan) pada setiap kata juga diperhatikan. Mereka mencoba mengklasifikasikan sentimen dengan mempertimbangkan unggahan yang merepresentasikan hal negatif atau penekanan terhadap objek tertentu (Yoo et al., 2018). Selain itu, penelitian tersebut mengarah kepada evaluasi terhadap efektifitas media sosial sebagai destinasi untuk menyampaikan komunikasi tertentu. Penelitian tersebut menyebutkan bahwa konten yang ada di media sosial dapat dianalisis sentimennya yaitu seperti dapat mengkategorikan unggahan sesuai dengan format atau tipe, bisa juga konten beserta daya tariknya (Molina et al., 2020).

Terkait analisis sentimen pada media sosial, analisis konten di media sosial dapat membantu keputusan pemerintah dalam menerapkan kebijakan (seperti melakukan penerapan kebijakan yang berhubungan dengan pelarangan sesuatu)(Spillane et al., 2021). Pada penelitian tersebut, rumusan masalah bagaimana hasil analisis konten unggahan di Instagram serta pengaruhnya terhadap keputusan kebijakan pemerintah dalam memerangi kasus narkoba. Hasil dari penelitian tersebut membuktikan bahwa secara tidak langsung mendapatkan hasil bahwa hasil analisis konten di media sosial dapat membantu keputusan pemerintah dalam melakukan penerapan kebijakan seperti kebijakan pelarangan sesuatu.

2.2 Langkah Analisis Sentimen

2.2.1 Text Pre-Processing

Text pre-processing merupakan suatu proses perubahan bentuk data yang tidak terstruktur menjadi data yang terstruktur. Proses ini bertujuan untuk memperkecil dimensi data sehingga proses komputasi lebih efisien dan menjadi lebih presisi. Proses *text pre-processing* merupakan tahap awal persiapan teks menjadi data yang selanjutnya akan diolah pada proses selanjutnya (Pramudita et al., 2018). Artinya bahwa proses pembersihan teks memiliki tujuan untuk membersihkan data dari komponen-komponen yang tidak memiliki hubungan dengan informasi yang ada pada dokumen.

Berikut merupakan beberapa tahapan dalam melakukan *text pre-processing* menurut (S Kurniawan, W Gata, 2017) dan (Pramudita et al., 2018):

a. *Cleaning*

Cleaning adalah tahap membersihkan dokumen dari kata yang tidak diperlukan, tanda baca, angka, *hashtag*, *link* dan *username*.

b. *Case Folding*

Case folding merupakan proses mengubah huruf ke dalam bentuk yang sama, yaitu kapital semua atau kecil semua. Dalam tahap ini, semua huruf yang ada dalam kalimat diseragamkan menjadi huruf kecil.

c. *Filtering*

Filtering merupakan proses pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan proses untuk menghilangkan kata yang dianggap tidak memiliki makna. Kata tersebut meliputi kata penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak memiliki arti dalam penentuan kelas suatu dokumen.

d. *Tokenizing*

Tokenizing merupakan proses pemisahan antar kata berdasarkan karakter spasi. *Tokenizing* digunakan untuk mendapatkan potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya.

e. *Stemming*

Stemming adalah proses pengambilan kata yang memiliki makna dan mengubah kata ke bentuk dasarnya sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Pada tahap ini digunakan *library Python Sastrawi*. *Stemming* bertujuan untuk mentransformasikan kata menjadi kata dasarnya dengan menghilangkan kata imbuhan awalan, sisipan, akhiran serta gabungan awalan dan akhiran.

2.2.2 *Word Cloud*

Word cloud adalah representasi visual dari frekuensi kata. Dalam penggunaannya *word cloud* semakin banyak digunakan sebagai alat sederhana untuk mengidentifikasi fokus materi secara tertulis. Semakin umum *file* atau berkas istilah yang muncul dalam teks yang sedang dianalisis, maka semakin besar ukuran kata yang muncul di gambar (Atenstaedt, 2012). *Word cloud* semakin banyak digunakan sebagai alat yang sederhana di banyak bidang seperti politik, bisnis dan pendidikan. Salah satu contoh fungsi dari *word cloud* adalah untuk memvisualisasikan konten pidato.

Word cloud merupakan alat visualisasi data yang jelas dan tegas sehingga dapat menyajikan data teks dalam format yang sederhana namun tidak mengurangi instrumen kejelasan dan ketegasannya (Van Den Rul, 2019). *Word cloud* merupakan alat yang informatif, artinya sangat berguna bagi peneliti untuk mengkomunikasikan tentang kejadian yang sedang terjadi di media sosial berdasarkan data dari kumpulan kata-kata seperti *caption* di media sosial. Selanjutnya, dalam melakukan analisis data dan memvisualisasikannya menggunakan *word cloud* ternyata pengguna dapat menarik beberapa wawasan terkait konten yang sedang dibahas dengan proses yang lumayan cepat. Format visualnya juga secara langsung merangsang khalayak untuk berpikir dan menarik wawasan terbaik terhadap konten yang sedang dibahas (Van Den Rul, 2019). Selanjutnya, dalam melakukan analisis data dan memvisualisasikannya menggunakan *word cloud* ternyata pengguna dapat menarik beberapa wawasan terkait konten yang sedang dibahas dengan proses yang lumayan cepat. Format visualnya juga secara langsung merangsang khalayak untuk berpikir dan menarik wawasan terbaik terhadap konten yang sedang dibahas.

2.3 Penelitian Terkait

Beberapa penelitian tentang analisis sentimen terdahulu sudah dilakukan. Penelitian tersebut sangatlah penting dalam penelitian ini sebagai kajian untuk mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan, untuk menghindari terjadinya tindakan duplikasi yang dilakukan oleh penulis. Tujuan dari tinjauan pustaka ini adalah untuk mengkaji metode yang akan dipakai dalam penelitian ini. Dalam penelitian analisis sentimen terdapat 2 metode yang paling populer digunakan yaitu *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Masing-masing metode memiliki nilai akurasi yang berbeda. Pada penelitian terdahulu yang penulis kaji, penelitian dengan metode *Naive Bayes Classifier* memiliki nilai akurasi yang terbilang tinggi.

Tabel 2.1 Tabel Kelebihan Naïve Bayes pada penelitian sebelumnya

Penulis	Fokus Penelitian	Metode	Hasil Penelitian
Xhemali (2009)	Perbandingan metode klasifikasi pada halaman <i>website</i>	<i>Naive Bayes Classifier</i> vs. <i>Decision Trees</i> vs. <i>Neural Networks</i>	Akurasi paling tinggi dengan metode <i>Naive Bayes</i> sebesar 97,26%.
Ramadhani (2015)	Klasifikasi Opini Publik (<i>Review Movie</i>)	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Akurasi paling tinggi dengan metode <i>Naïve Bayes</i> sebesar 62,6%.
Arifin (2016)	Klasifikasi prioritas data pelanggan PT. Time Excelindo	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Akurasi paling tinggi dengan metode <i>Naïve Bayes</i> sebesar 90%.
Himawan & Eliyanti (2021)	Perbandingan Akurasi Analisis Sentimen <i>Tweet</i> terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi	<i>Random Forest</i> , <i>Support Vector Machine</i> (SVM) dan <i>Naïve Bayes Classifier</i> (NBC)	Penelitian yang dilakukan untuk mengetahui sentimen pengguna twitter terhadap pemerintah provinsi DKI Jakarta di masa pandemi. Didapatkan nilai akurasi pada algoritma <i>Random Forest</i> sebesar 75.81%, pada algoritma <i>Naïve Bayes</i> sebesar 75.22%, dan untuk algoritma <i>Support Vector Machine</i> akurasi yang dihasilkan sebesar 77.58%.

Yonathan Sari Mahardhika dan Eri Zuliarso (2018)	Analisis Sentimen terhadap Pemerintahan Joko Widodo pada Media Sosial Twitter Menggunakan Algoritma <i>Naive Bayes Classifier</i>	<i>Naive Bayes Classifier</i>	Dengan jumlah data sebanyak 400 data <i>tweet</i> terdiri dari 300 data latih dan 100 data ujimemberikan prediksi kelas sentimen data berdasarkan <i>confussion matrix</i> dengan rincian 49 <i>tweet</i> sentimen negatif dan 51 <i>tweet</i> sentimen positif. Dan didapatkan hasil akurasi untuk analisis klasifikasi dengan metode <i>Naive Bayes Classifier</i> sebesar 97%
Bella Nurfadhila (2019)	Analisis Sentimen untuk Mengukur Tingkat Indikasi Depresi pada Twitter Menggunakan <i>Text Mining</i>	<i>Naive Bayes Classifier</i>	Didapatkan data <i>tweet</i> sebanyak 1498 yang kemudian dilakukan analisis dengan bahasa pemrograman <i>Python</i> dengan <i>data train</i> sebanyak 1000 data, dan <i>data test</i> sebanyak 498 data. Pada penelitian ini, tingkat akurasi yang dihasilkan sebesar 73,86%.
Salah (2015)	Implementasi Metode Klasifikasi <i>Naive Bayes</i> Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga	<i>Naive Bayes Classifier</i>	Metode <i>Naive Bayes</i> menghasilkan keakuratan atau akurasi sebesar 78.33% dalam memprediksi besarnya pengguna listrik rumah tangga
Syafitri (2017)	Penerapan Teknik <i>Web Scraping</i> dan Klasifikasi Sentimen Menggunakan Metode <i>Naive Bayes Classifier</i> dan Asosiasi (Studi Kasus: Data Ulasan <i>The House Of Raminten</i> Pada Situs <i>TripAdvisor</i>)	<i>Naive Bayes Classifier</i>	Metode <i>Naive Bayes Classifier</i> menghasilkan tingkat akurasi 84% pada ulasan bahasa Indonesia dan pada ulasan bahasa Inggris sebesar 77%.

Berbeda dengan penelitian sebelumnya, penelitian yang akan dilakukan yaitu Analisis Sentimen pada Media Sosial Instagram Klub Persija Jakarta menggunakan metode *Naive Bayes Classifier*.

Penelitian menggunakan metode *Naive Bayes Classifier* pernah dilakukan oleh Yonnathan Sari Mahardhika dan Eri Zuliarso pada tahun 2018 dengan judul "Analisis Sentimen Terhadap Pemerintahan Joko Widodo pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes Classifier*" yang membahas tentang berbagai macam komentar publik dalam media sosial Twitter dimana objek pada penelitian ini adalah *tweet* tentang Pemerintahan Presiden Joko Widodo dengan jumlah data sebanyak 400 data *tweet* terdiri dari 300 data latih dan 100 data uji. Dan didapatkan hasil akurasi untuk analisis klasifikasi dengan metode *Naive Bayes Classifier* sebesar 97%.

Dalam Tugas Akhir yang dilakukan oleh Mulajati (2017) dengan judul Implementasi Teknik *Web Scraping* dan Klasifikasi Sentimen Menggunakan Metode *Naïve Bayes Classifier* dan Asosiasi Teks (Studi Kasus : Data Ulasan Penumpang Maskapai Penerbangan Garuda Indonesia Pada Situs *TripAdvisor*) dengan menggunakan metode *Naïve Bayes Classifier* (NBC) diperoleh tingkat akurasi sebesar 82,02%.

Patel & Passi (2020) dalam penelitiannya membandingkan metode *Naïve Bayes*, *Support Vector Machine*, *Random Forest*, dan *K-Nearest Neighbour* untuk mengklasifikasikan ulasan mengenai Piala Dunia 2014 ke dalam tiga kelas, yaitu positif, netral, dan negatif. Data diambil dari Twitter pada Juni-Juli 2014 dan didapatkan sekitar 2 juta *tweet*. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* memiliki performansi yang paling baik, karena memiliki nilai akurasi dan AUC tertinggi dibandingkan metode lain (Patel & Passi, 2020)

Syarifuddin (2020) dalam penelitiannya menggunakan metode *Naïve Bayes* dan *Convolutional Neural Network* (CNN) untuk analisis sentimen opini masyarakat guna menyelaraskan dan memberi pandangan baru mengenai suatu isu tentang Covid-19, serta untuk mengetahui kecenderungan opini masyarakat di Twitter. Pengambilan data dilakukan dengan menggunakan API Twitter dan didapatkan 1098 opini dengan kata kunci “Covid-19”. Hasil penelitian ini menunjukkan bahwa metode *Naïve Bayes* memiliki akurasi dan presisi nilai positif lebih tinggi dibandingkan metode CNN, yaitu sebesar 63,21% dan 66,40%.

Penelitian menggunakan metode *Naive Bayes Classifier* pernah dilakukan oleh Silvia Aprilia, Muhammad Tanzil Furqon, dan Mochammad Ali Fauzi pada tahun 2018 dengan judul ”Analisis Sentimen untuk Mengukur tingkat Indikasi Depresi pada Twitter Menggunakan *Text Mining*” dimana objek pada penelitian ini adalah data rekam medik tentang penyakit skizofrenia dan episode depresi pada pasien gangguan kejiwaan dengan jumlah data sebanyak 200 data dengan ratio perbandingan *data train* dan *data test* 80% : 20% menghasilkan tingkat akurasi sebesar 79%. Keuntungan lain dalam penggunaan *naïve bayes classifier* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Metode *Naïve Bayes Classifier* merupakan salah satu metode klasifikasi probabilitas yang sederhana. Keuntungan dalam menggunakan metode *Naïve Bayes Classifier* adalah metode ini hanya memerlukan data latih yang sedikit dalam proses klasifikasi teks (Suryani et al., 2019).

Karena yang diasumsikan sebagai variabel *independent* maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari *matriks kovarians* (Imron, 2019). Rumus teorema bayes secara umum sebagai berikut:

$$P(H|X) = \frac{P(H|X) P(H)}{P(X)} \quad (1)$$

dengan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (*posteriori prob.*)

P(H) = Probabilitas hipotesis H (*prior prob.*)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

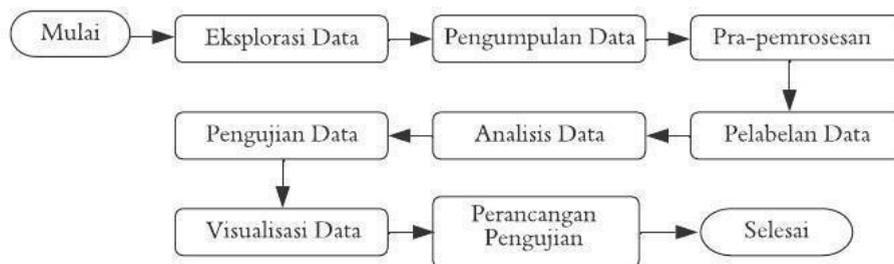
P(X) = Probabilitas dari X

Adapun aturan Bayes adalah sebagai berikut :

Jika $P(h_1|x) < P(h_2|x)$, maka x diklasifikasikan sebagai h2. Pernyataan $P(h_1|x)$ mengindikasikan probabilitas hipotesis h1 berdasarkan kondisi x terjadi, begitu pula dengan h2. Sehingga didapat klasifikasi dari x sesuai dengan probabilitas terbesar diantara probabilitas x terhadap semua kelas.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian adalah tahapan yang akan dilakukan selama pelaksanaan penelitian dan disusun dengan baik serta sistematis. Metodologi penelitian dijadikan acuan atau pedoman selama penelitian agar tercapainya tujuan yang diharapkan. Berikut merupakan tahapan metodologi penelitian atau tahapan penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Diagram Alir Analisis Sentimen

Dalam melakukan penelitian ini, ada beberapa tahapan yang dilakukan mulai dari proses Eksplorasi data dari akun Instagram Persija Jakarta. Kemudian tahap Pengumpulan Data dengan melakukan *scrapping* data komentar pada postingan Instagram Persija Jakarta menggunakan *Extension IG Comment Export* pada *browser Google Chrome*. Selanjutnya data hasil *scrapping* akan melalui tahapan Pra-pemrosesan yaitu meliputi proses *cleaning*, *case folding*, *filtering*, *tokenizing*, *stemming* dan *lemmatization*. Kemudian dilakukan proses Pelabelan data yaitu data ulasan yang akan di kategorikan menjadi sentimen positif, sentimen negatif dan sentimen netral. Selanjutnya proses Analisis data yang dimana sebelum pengujian dibagi menjadialisis deskriptif berbentuk *Word cloud* dan metode *machine learning* yaitu *Naïve Bayes Classifier*. Kemudian dilakukan pengujian data menggunakan *Confussion matrix* untuk menentukan nilai presisi, *recall*, dan nilai akurasi pada model *Naïve Bayes Classifier*. Selanjutnya proses Visualisasi data menggunakan *Word Cloud* yang akan menampilkan pola kata atau kata yang paling sering digunakan dalam komentar postingan akun Instagram Persija Jakarta.

3.1 Pengumpulan Data

Dalam segi pengumpulan data, waktu pengumpulan data memiliki dua pendekatan yaitu metode *Longitudinal* (biasa disebut pendekatan bujur) dan metode *Cross Section* (biasa disebut insidental). Pendekatan *Longitudinal* atau (pendekatan bujur) adalah penelitian yang meneliti perkembangan sesuatu aspek pada suatu hal dalam seluruh periode waktu, atau sebuah tahapan perkembangan dengan waktu yang cukup panjang (Siyoto, S., & Sodik, 2015). Selanjutnya, menurut penelitian yang dilakukan oleh Sugiyono (2015), populasi yaitu subjek ataupun objek pada wilayah yang memiliki kualitas serta karakteristik tertentu yang akan dipelajari oleh peneliti dalam rangka untuk menarik kesimpulan (Sugiyono, 2015). Sedangkan untuk sampel, *purposive sampling* adalah teknik pengambilan sampel yang sesuai dengan tujuan dan kriteria tertentu (Bhardwaj, 2019). Dalam melakukan *purposive sampling*, sampel yang sudah ditentukan seperti misalnya objek studi sudah ditentukan dan dipilih berdasarkan kriteria yang telah ditetapkan. Teknik pengambilan sampel *purposive sampling* biasa disebut juga dengan *deliberate sampling* atau *judgmental sampling* yang artinya sampel yang disengaja (Bhardwaj, 2019).

Adapun teknik pengumpulan data, terdapat tiga metode untuk mendapatkan data dari *platform* sosial media . Pertama, yaitu dengan mengunduh data langsung dari *database server*, cara ini cukup rumit karena peneliti harus memiliki kolaborasi yang erat dengan perusahaan sosial media. Kedua, yaitu mengumpulkan data dengan menggunakan antarmuka pemrograman aplikasi atau biasa disebut dengan API (*Application Programming Interface*). API pada dasarnya adalah program yang memungkinkan perangkat lunak untuk berinteraksi dengan perangkat lunak lain. Dengan *script file* yang dikodekan, pengembang aplikasi mengikuti aturan yang ditentukan pemilik *platform* sosial media. *Script* ini biasanya memiliki *permission* dalam melakukan pengunduhan data di media sosial. Selain itu, dengan API juga dapat digunakan peneliti untuk mengumpulkan data dari *platform* sosial media. Ketiga, yaitu *web scrapping* berguna untuk jenis *platform* media sosial yang tidak menyediakan layanan API. Metode yang dilakukan dalam melakukan *web scrapping* adalah dengan permintaan HTTP (*Hyper Text Transfer Protocol*) (Liang & Zhu, 2017). Di dalam pengumpulan data sendiri terbagi menjadi beberapa bagian yaitu cara pengumpulan data, waktu pengumpulan data, populasi dan sampel.

Dalam melakukan sebuah penelitian, perlu diperhatikan terlebih dahulu sebuah analisis dari berbagai aspek-aspek seperti dalam pengumpulan data. Di dalam pengumpulan data

sendiri terbagi menjadi beberapa bagian yaitu cara pengumpulan data, waktu pengumpulan data, populasi dan sampel.

3.1.1 Waktu Pengumpulan Data

Berdasarkan waktu pengumpulan dalam penelitian ini, waktu pengumpulan data termasuk ke pendekatan *Cross Section* atau insidental, yaitu data akan dikoleksi dalam rentang waktu tertentu. Dalam penelitian ini waktu yang diperlukan untuk mengumpulkan data adalah dalam rentang waktu dari 22 Maret 2021 – 17 September 2023.

3.1.2 Teknik Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini didapatkan sepenuhnya menggunakan perangkat *Extension IG Comment Export* pada browser *Google Chrome* lalu data yang dikumpulkan akan disimpan ke dalam file ekstensi *.csv*. Dalam penelitian ini, menggunakan metode *Naïve Bayes Classifier*. Adapun dalam analisis konten, isi dari konten tersebut berisikan tentang komentar pada 60 postingan Instagram Persija Jakarta ketika Menang dan Kalah dengan total data 6.000 data komentar. Untuk postingan kondisi Persija Jakarta menang terdapat 30 Postingan (3.000 data) dan postingan ketika Persija Jakarta kalah sebanyak 30 postingan (3.000 data) dari *platform* media sosial Instagram klub sepak bola Persija Jakarta sebagai sumber datanya. Melakukan analisis sentimen yang sederhana dan hanya ingin mendapatkan pemahaman umum tentang apakah suatu teks bersifat positif, negatif, atau netral, maka 3.000 data saja sudah cukup. Sesuai konteks dalam penelitian ini, data yang dikumpulkan adalah konten unggahan terkait akun media sosial Instagram klub Persija Jakarta yang telah *verified*.

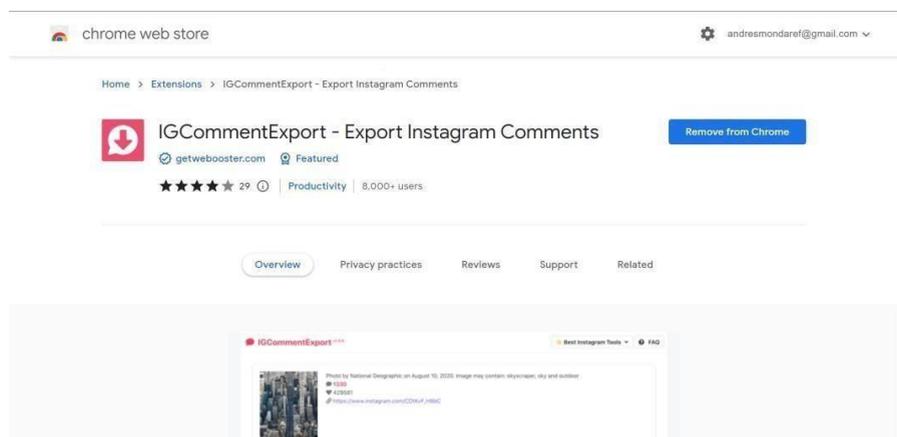
3.1.3 Populasi

Dalam melakukan penelitian ini populasinya adalah unggahan yang berkaitan dengan komentar pada akun media sosial Instagram klub sepak bola Persija Jakarta ketika Persija dalam kondisi Menang dan Kalah.

3.1.4 Sampel

Sampel yang digunakan dalam pengumpulan data dalam penelitian ini yaitu akun Instagram resmi dari klub sepak bola Persija Jakarta yang sudah *verified* dengan unggahan yang berkaitan dengan komentar positif, negatif dan netral. Berikut ini adalah langkah-langkah dalam melakukan pengumpulan data di penelitian ini:

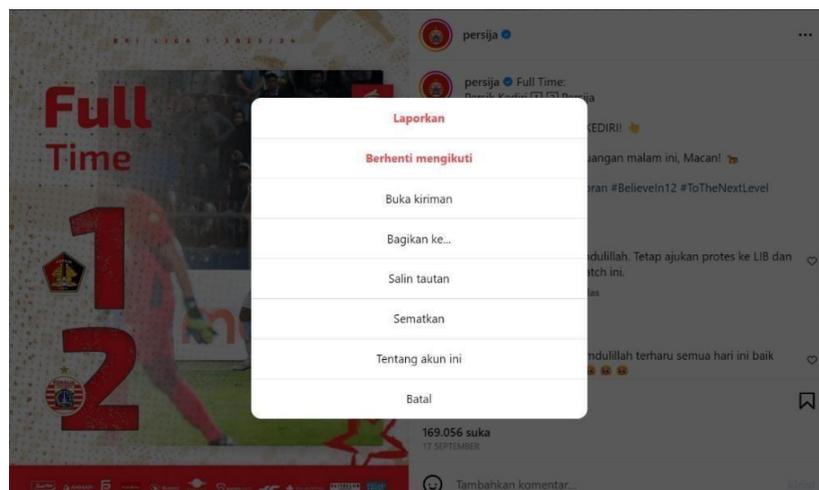
- a. Melakukan instalasi *Extension IG Comment Export* pada *Google Chrome*.



Gambar 3.2 Instalasi *Extension IG Comment Export*

- b. Melakukan *Scrapping* data komentar pada unggahan akun Media sosial Instagram klub Sepak bola Persija Jakarta

Langkah 1: Salin tautan unggahan pada salah satu postingan Instagram Persija Jakarta



Gambar 3.3 Salin Tautan Pada Postingan Persija Jakarta

Langkah 2: Copy link tersebut ke Extension IG Comment Export

BRILIGA 1 2023/24

Full Time

1-2

©2023 WeBooster v1.0.14

1 minggu

dikysoemarno Alhamdulillah. Tetap ajukan protes ke LIB dan PSSI mengenai wasit match ini.
1 minggu 3,082 suka Balas

Lihat balasan (270)

putrasiregarr17 Alhamdulillah terharu semua hari ini baik @persia & @fcbekasi 🙏🙏🙏

169.056 suka
17 SEPTEMBER

Tambahkan komentar... Kirim

Gambar 3.4 Copy link Unggahan ke Extension IG Comment Export

Langkah 3: Export komentar Instagram menjadi file .csv

Full Time: Persik Kediri 1-2 Persija 3 POIN PENTING DARI KEDIRI! 🙏 Terima kasih untuk perjuangan malam ini, Macan! 🙏 #Persija #MacanKemayoran #BelieveIn12 #ToTheNextLevel #PersijaDay

🗨️ 9707 ❤️ 169056

🔗 <https://www.instagram.com/p/CxS6jGkvdVT>

100%

DOWNLOAD 100 COMMENTS

Gambar 3.5 Download data komentar menjadi file.csv

Berikut contoh *file* .csv yang berhasil didownload:

User Id	Username	Comment	Comment Text	Profile UR	Profile Pic URL	Date
30790091	nisuk_wht	179401081	δΥ°cδΥ°c	https://www.instagram.com/nisuk_wht/	https://www.instagram.com/nisuk_wht/profile_pic/	10/3/2022, 8:59:08 AM
50068515	alex_slam	17949190	δΥ°cδΥ°c	https://www.instagram.com/alex_slam/	https://www.instagram.com/alex_slam/profile_pic/	10/3/2022, 9:19:22 AM
23154508	htw_28	17980816	Sebagai pembelajaran bagi kita semua.. jadi supporter far	https://www.instagram.com/htw_28/	https://www.instagram.com/htw_28/profile_pic/	10/3/2022, 9:39:19 AM
44998631	alaymisyif	18041871	Nangis bgtttttttδΥ°c	https://www.instagram.com/alaymisyif/	https://www.instagram.com/alaymisyif/profile_pic/	10/3/2022, 9:45:47 AM
45939574	meyliani.n	17957847	@andikahusni93 ini yg namanya provokator yg bkln sepak	https://www.instagram.com/meyliani.n/	https://www.instagram.com/meyliani.n/profile_pic/	10/3/2022, 9:46:12 AM
42391670	mhmdanv	17926755	Rip dunia sepak bolaδΥ°cEklas sembah indonesiak dri di	https://www.instagram.com/mhmdanv/	https://www.instagram.com/mhmdanv/profile_pic/	10/3/2022, 9:50:00 AM
40076692	ebithebetl	17937180	cukup saling tidak menyukai, gak usah saling melukai. dan	https://www.instagram.com/ebithebetl/	https://www.instagram.com/ebithebetl/profile_pic/	10/3/2022, 9:57:52 AM
53065078	baisgan	17999745	Main tanggal brp?	https://www.instagram.com/baisgan/	https://www.instagram.com/baisgan/profile_pic/	10/3/2022, 10:08:10 AM
17621131	richardjoe	17966498	afri, atri,	https://www.instagram.com/richardjoe/	https://www.instagram.com/richardjoe/profile_pic/	10/3/2022, 10:15:30 AM
56211069	afriilosury	17990861	@aziefgaye TIDAK AKAN ADA ASAP SEBELUM ADA API.	https://www.instagram.com/afriilosury/	https://www.instagram.com/afriilosury/profile_pic/	10/3/2022, 10:28:11 AM
21851545	ivanitimoti	18025661	#AREMAOUT	https://www.instagram.com/ivanitimoti/	https://www.instagram.com/ivanitimoti/profile_pic/	10/3/2022, 10:28:56 AM
11923137	ryoryuha.	18254683	"Kejadian di Kanjuruhan jadi pelajaran buat kita, di sana	https://www.instagram.com/ryoryuha./	https://www.instagram.com/ryoryuha./profile_pic/	10/3/2022, 10:38:20 AM
26282311	gakjebrott	17915044	@faatimahutz si cantik komen wkwk	https://www.instagram.com/gakjebrott/	https://www.instagram.com/gakjebrott/profile_pic/	10/3/2022, 10:51:18 AM
46370651	damdand	18326360	@ikhwanriyadi betul itu harus nya main di belakang Meg	https://www.instagram.com/damdand/	https://www.instagram.com/damdand/profile_pic/	10/3/2022, 11:00:26 AM
82455599	rizkifilmal	17958245	@ilhmwhydii tapi gak seru :v	https://www.instagram.com/rizkifilmal/	https://www.instagram.com/rizkifilmal/profile_pic/	10/3/2022, 11:13:36 AM
17993817	alysopina	18007567	@salsanjani pea bgt si Ibnu ma saδΥ°c	https://www.instagram.com/alysopina/	https://www.instagram.com/alysopina/profile_pic/	10/3/2022, 11:19:18 AM
60983473	mr_axc98	17988257	Semoga jangan terjadi lagi #respect	https://www.instagram.com/mr_axc98/	https://www.instagram.com/mr_axc98/profile_pic/	10/3/2022, 11:20:54 AM
53556322	ghalib_a	17941735	Padahal mau lihat persija mengaungδΥ°c	https://www.instagram.com/ghalib_a_/	https://www.instagram.com/ghalib_a_/profile_pic/	10/3/2022, 11:32:01 AM
43069186	vdjokto	18327122	@yudap8651 bukanya 5thn ya	https://www.instagram.com/vdjokto/	https://www.instagram.com/vdjokto/profile_pic/	10/3/2022, 11:33:46 AM
17605589	muhamad	17980009	Jangan ada lagi kebencian di diri kita Jak salam BobotohδΥ°c	https://www.instagram.com/muhamad17980009/	https://www.instagram.com/muhamad17980009/profile_pic/	10/3/2022, 11:48:45 AM
52204816	euforia.pr	17978114	δΥ°cE	https://www.instagram.com/euforia.pr/	https://www.instagram.com/euforia.pr/profile_pic/	10/3/2022, 11:53:52 AM
58421060	muhamm	18041672	@sandershendro12 ok met	https://www.instagram.com/muhamm18041672/	https://www.instagram.com/muhamm18041672/profile_pic/	10/3/2022, 12:00:45 PM

Gambar 3.6 Contoh hasil *File* .csv yang berhasil di *download*

3.2 Pra-pemrosesan

Pada tahap ini merupakan langkah awal setelah melakukan proses pembagian data. Pada *text pre-processing* ini akan dijelaskan proses dari pengolahan data untuk diolah pada proses selanjutnya. Berikut merupakan langkah-langkah pada proses *text pre-processing*:

a. *Cleaning*

Cleaning adalah tahap membersihkan dokumen dari kata yang tidak diperlukan, tanda baca, angka, *hashtag*, *link* dan *username*.

b. *Case Folding*

Case folding merupakan proses mengubah huruf ke dalam bentuk yang sama, yaitu kapital semua atau kecil semua. Dalam tahap ini, semua huruf yang ada dalam kalimat diseragamkan menjadi huruf kecil.

c. *Filtering*

Filtering merupakan proses pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan proses untuk menghilangkan kata yang dianggap tidak memiliki makna. Kata tersebut meliputi kata

penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak memiliki arti dalam penentuan kelas suatu dokumen.

d. *Tokenizing*

Tokenizing merupakan proses pemisahan antar kata berdasarkan karakter spasi. *Tokenizing* digunakan untuk mendapatkan potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya.

penentuan kelas suatu dokumen.

e. *Stemming*

Stemming adalah proses pengambilan kata yang memiliki makna dan mengubah kata ke bentuk dasarnya sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Pada tahap ini digunakan *library Python Sastrawi*.

f. *Lemmatization*

Lemmatization adalah proses untuk mengubah kata-kata ke bentuk dasar. Tujuan utama dari *lemmatization* adalah untuk mengurangi kata-kata ke bentuk dasar dan sehingga lebih dapat diidentifikasi atau dihitung lebih mudah dalam analisis teks.

3.3 Pembobotan TF-IDF (*Term-Frequency Times Inverse Document Frequency*)

Data yang telah melewati tahap *pre-processing* harus berbentuk angka, agar data yang masih berupa kata-kata dapat diolah dan dihitung. Data tersebut diubah menjadi vektor kemudian diberi nilai dan bobot untuk setiap kata kemudian diolah menggunakan algoritma prediksi. Pada tahap pembobotan kata, menggunakan algoritma *Term-Frequency Times Inverse Document Frequency* (TF-IDF).

Metode ini dipilih karena efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Invers Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen. Secara sederhana, metode ini digunakan untuk menghitung sering munculnya suatu kata atau *term* dalam dokumen. *Term Frequency* (TF) merupakan frekuensi kemunculan sebuah kata atau *term* dalam dokumen. Semakin besar jumlah *term* yang muncul (TF tinggi) maka semakin besar bobot dokumen atau memberikan nilai kemiripan yang semakin besar.

3.4 Analisis Data

Berikut ini adalah beberapa metode analisis data yang digunakan dalam penelitian ini:

- a. 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
- b. Metode *Machine Learning* yaitu *Naive Bayes Classifier*, yang dalam penelitian ini digunakan untuk mengklasifikasikan ulasan atau *review* yang dapat berbentuk kelas positif, negatif atau netral.

3.5 Pengujian Data

Tahapan pengujian atau klasifikasi menggunakan metode *Naive Bayes Classifier* ini berisikan tentang implementasi sistem dengan tujuan sistem yang dibangun sesuai dengan tujuan yang ingin dicapai. Pengujian ini dilakukan menggunakan sebuah himpunan pembagian data latih dan data uji. Untuk pengujian tingkat akurasi menggunakan metode pengujian *Confusion Matrix* dalam proses evaluasi mencari nilai akurasi, nilai presisi dan nilai *recall*.

Model klasifikasi yang dibuat merupakan pemetaan dari baris data dengan sebuah hasil prediksi kelas dari data tersebut. Klasifikasi yang memiliki dua *output* label disebut dengan klasifikasi biner yang direpresentasikan dalam $\{0,1\}$, $\{\text{ya, tidak}\}$, $\{+1,-1\}$ atau $\{\text{positive; negative}\}$ untuk setiap data input yang diberikan. Pada proses evaluasi klasifikasi ada empat kemungkinan yang terjadi dari hasil klasifikasi suatu data. Jika data positif dan diprediksi positif maka akan dihitung sebagai *true positive* dan jika data positif diprediksi negatif maka akan dihitung sebagai *false negative*. Pada data negatif jika diprediksi negatif maka akan dihitung *true negative* dan jika diprediksi positif maka akan dihitung *false positive* (Fawcett, 2006).

Aktual

		Aktual	
		<i>Positive</i>	<i>Negative</i>
Prediksi	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

a. Accuracy

Accuracy adalah tingkat ketepatan prediksi yang benar. Rumus *accuracy* dapat dilihat pada persamaan sebagai berikut:

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN} \quad (2)$$

b. Precision

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban sistem. Dalam analisis sentimen presisi memberikan gambaran tentang seberapa akurat model dalam mengklasifikasikan instansi sebagai positif. Presisi juga berguna untuk mengidentifikasi sejauh mana model dapat diandalkan ketika memberi label positif. Dengan meningkatkan presisi, model cenderung lebih hati-hati dalam memberikan label positif, mengurangi kemungkinan kesalahan positif palsu di mana ulasan yang seharusnya netral atau negatif salah diklasifikasikan sebagai positif. Jika kelas positif (misalnya, ulasan positif) sangat tidak seimbang dengan kelas negatif atau netral, maka presisi menjadi lebih penting. Presisi yang tinggi dapat memberikan keyakinan bahwa label positif yang diberikan oleh model benar-benar relevan. Rumus *precision* dapat dihitung melalui persamaan sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

c. Recall

Recall adalah perhitungan ketepatan prediksi yang digunakan sebagai ukuran tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Recall* memberikan informasi tentang kemampuan model untuk menemukan semua sentimen yang sebenarnya positif. *Recall* juga berguna untuk memastikan bahwa model tidak melewatkan terlalu banyak sentimen positif yang sebenarnya ada. Dengan meningkatkan *recall*, model lebih cenderung menangkap ulasan yang sebenarnya positif, mengurangi resiko kesalahan negatif palsu di mana ulasan yang seharusnya positif dianggap sebagai negatif atau netral. Rumus *recall* dapat dihitung melalui persamaan sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

3.6 Visualisasi Data

Pada penelitian ini dilakukan visualisasi data menggunakan paket *Word Cloud*. Tujuannya adalah untuk mengidentifikasi dan membentuk pola kata yang dapat berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting. Proses melakukan visualisasi menggunakan paket *word cloud*, data yang diperoleh berasal dari data *comment* dari Instagram Persija Jakarta. *Comment* atau komentar yang berisikan banyak kumpulan kata-kata tersebut kemudian dianalisis menggunakan paket *word cloud*, sehingga kata-kata yang paling sering dibahas akan ditampilkan dalam ukuran kata yang besar, dan sebaliknya kata-kata yang paling sedikit dibahas akan ditampilkan juga, namun dalam ukuran kata yang kecil.

3.7 Perancangan Pengujian

Tahapan ini adalah proses untuk merancang pengujian berdasarkan hasil analisis yang telah dilakukan. Perancangan bertujuan untuk mempermudah dalam proses pengujian data. Pengujian yang digunakan dalam penelitian ini berupa *Google Colab*. *Google Colab* di implementasikan menggunakan bahasa pemrograman *Python*. Perancangan pengujian diisi dengan visualisasi tentang perencanaan umum tahapan ulasan yang akan dianalisis. Masukan teks ulasan akan melalui tahap *Pre-processing* yang dimana melalui proses *cleaning*, proses *case folding*, *filtering*, *tokenizing*, *stemming* dan *lemmatization*.

Setelah itu melakukan tahap TF-IDF, yaitu proses pembobotan kata yang sering muncul pada komentar Instagram Persija Jakarta. Kemudian dilakukan proses pelabelan data, untuk menentukan termasuk kedalam sentimen positif, sentimen negatif dan sentimen netral. Selanjutnya tahap klasifikasi menggunakan algoritma *Naïve Bayes Classifier* untuk mengetahui model klasifikasi, diantaranya adalah *accuracy*, *precision*, dan *recall*. Kemudian untuk memvalidasi data ulasan, dilakukan proses prediksi supaya mengetahui data ulasan tersebut termasuk ke sentimen positif, negatif, atau netral.

BAB IV

HASIL DAN PEMBAHASAN

4.1 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 waktu sejak 22 Maret 2021 – 17 September 2023. Data yang digunakan sebanyak 6.000 data komentar pada postingan Instagram Persija Jakarta. Perangkat 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*. Tahapan ini melibatkan *library python* yaitu *Sastrawi* dan *Wordcloud*. Sebelum menggunakan *library* tersebut terlebih dahulu memberikan perintah *pip* pada terminal *google colaboratory*.

```
!pip install Sastrawi
!pip install wordcloud
```

Gambar 4.1 Menginstall *Library*

Penelitian ini menggunakan *library Python Sastrawi* dan *wordcloud* karena mudah untuk digunakan. Dataset yang diambil merupakan ulasan terhadap performa Persija Jakarta berdasarkan pada *platform* Instagram Persija Jakarta. Kemudian mengimport *library* yang sudah di *install*.

```
import re, json, os, pandas as pd
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

Gambar 4.2 Mengimport *Library*

Setelah beberapa *library* di *import*, koneksikan *google drive* ke *google colaboratory*.

```
from google.colab import drive
drive.mount('/content/drive')
```

Gambar 4.3 *Source Code* Koneksikan *Google Drive* ke *Google Colab*

Bertujuan untuk menghubungkan semua akses *Google Drive*, termasuk data hasil *scrapping* seperti data komentar berbentuk file.csv yang telah di download, kamus positif dan negatif, *stopwords*, dan kamus alay. Contoh hasil data komentar setelah di *scrapping* teks masih mengandung beberapa yaitu seperti *User Id*, *User Name*, *Comment Id*, *Comment Text*, *ProfileURL*, *Avatar URL*, dan *Publish Date*. Berikut contoh ulasan hasil dari *scrapping*:

	User Id	User Name	Comment Id	Comment Text	Profile URL	Avatar URL	Publ i
0	"=54122545945"	"=marfeladitya06"	"=18010711306753367"	"=Simic ga main menang kwkw"	"=https://www.instagram.com/marfeladitya06"	"=https://instagram.fmq3-1.fna.fbcdn.net/v/t5...	=9/18/2 6:24:20
1	"=44594660412"	"=jahemerahbarokah_"	"=18280683988150771"	"=Dapet penalti = disayang PSSI,, Tidak dapat..."	"=https://www.instagram.com/jahemerahbarokah_"	"=https://instagram.fjog1-1.fna.fbcdn.net/v/t5...	=9/18/2 7:02:05
2	"=4884172665"	"=ntrizno20"	"=17973898748544222"	"=@fajar.panjul cocok ni jd komentator 🤔"	"=https://www.instagram.com/ntrizno20"	"=https://instagram.fjog1-1.fna.fbcdn.net/v/t5...	=9/18/2 7:20:19
3	"=59564139880"	"=rodeg12_"	"=18208102435253061"	"=@carlslos_ sama kek fans Ajisrep pas Persib ..."	"=https://www.instagram.com/rodeg12_"	"=https://instagram.fmq3-1.fna.fbcdn.net/v/t5...	=9/18/2 7:38:33
4	"=58161293193"	"=tpnasir4"	"=17963617067496286"	"=Ok 🤔 🤔"	"=https://www.instagram.com/tpnasir4"	"=https://instagram.fjog1-1.fna.fbcdn.net/v/t5...	=9/18/2 7:51:19
...
995	"=3304199502"	"=ohlosono_"	"=17975588891118058"	"=Oke itu"	"=https://www.instagram.com/ohlosono_"	"=https://instagram.fjog1-1.fna.fbcdn.net/v/t5...	=3/16/2 9:33:44
996	"=7261053880"	"=sditoo_"	"=18168348127284997"	"=masih loyoh"	"=https://www.instagram.com/sditoo_"	"=https://instagram.focu19-1.fna.fbcdn.net/v/t5...	=3/16/2 9:36:32

Gambar 4.4 Contoh Data Ulasan hasil *Scraping*

Dapat dilihat data ulasan masih banyak memiliki *noise*, jadi hasil *scrapping* masih membutuhkan tahap *pre-processing* agar mendapatkan teks ulasan yang diinginkan sebelum ketahap selanjutnya.

4.2 Implementasi *Pre-Processing*

Data ulasan pada postingan Instagram Persija Jakarta yang didapatkan dari proses *scrapping* pada media sosial Instagram belum dapat digunakan karena memiliki bentuk kalimat atau teks tidak berstruktur dimana memiliki banyak *noise*, terdapat ulasan berbahasa asing (Inggris) yang belum diterjemahkan ke dalam bahasa Indonesia dimana informasi yang didapatkan akan sangat susah diterjemahkan, maka dari itu data tersebut perlu melakukan proses pembersihan terlebih dahulu. Adapun langkah-langkah pra-pemrosesan yang akan digunakan adalah :

a. *Cleaning*

Cleaning bertujuan untuk membersihkan data dengan cara menghapus *username* (@), *hashtag* (#), tanda baca, angka, *single character*, dan mengganti *multiple space* menjadi *single space*. Data tersebut dibersihkan bertujuan untuk mengurangi beban pemrosesan dan dianggap tidak relevan untuk dianalisa. Berikut merupakan proses dari *cleaning*:

```
data['Comment Text'] = data['Comment Text'].str.replace(r'@[^\s]+', '',
regex=True)

data['Comment          Text']          =          data['Comment
Text'].str.replace(r'((www\S+)|(http\S+))', '', regex=True)

def remove_non_ascii(text):
    text = text.encode('ascii', 'ignore').decode('ascii')
    return text
def remove_special_characters(text):
    # Menghapus angka
    text = re.sub(r'\d+', '', text)
    # Menghapus tab
    text = re.sub(r'\t', '', text)
    # Menghapus spasi berlebih
    text = re.sub(r'\s+', ' ', text)
    # Menghapus karakter khusus
    text = re.sub(r'^a-zA-Z\s', '', text)
    return text.strip()

data['Comment Text'] = data['Comment Text'].apply(remove_non_ascii)

data['Comment          Text']          =          data['Comment
Text'].apply(remove_special_characters)

data = data[data['Comment Text'].apply(lambda x: len(x) >= 2)]

data
```

Gambar 4.5 Source Code tahap *Cleaning*

Pada proses *cleaning*, terjadi penghapusan karakter selain huruf, yang dihapus seperti spesial karakter, *link*, emoji dan komentar kurang atau sama dari 2 huruf itu akan dihapus. Berikut pada Tabel 4.1 menjelaskan hasil sebelum dan setelah dilakukan proses implementasi *cleaning*:

Tabel 4.1 Hasil *Cleaning* pada *comment text*

Ulasan	Setelah Proses Cleaning
semangat trs persijaku	semangat trs persijaku
semangat	semangat
semangat jangan menyerah jaacckkk	semangat jangan menyerah jaacckkk

b. Case Folding

Kemudian *case folding* adalah suatu proses untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil. Proses *case folding* bertujuan agar huruf kapital dan huruf kecil tidak terdeteksi beda arti. Implementasi proses *case folding* menggunakan *method Python* yaitu *lower()*, berikut merupakan proses dari *case folding*:

```
data['Comment Text'] = data['Comment Text'].str.lower()
data
```

Gambar 4.6 *Source Code* tahap *Case Folding*

Pada proses *Case folding* terjadi proses menyeragamkan huruf menjadi huruf kecil, berikut dijelaskan pada Tabel 4.2 merupakan hasil sebelum dan sesudah dari proses *case folding*:

Tabel 4.2 Hasil *Case Folding* pada *comment text*

Ulasan	Setelah Proses Case Folding
Di puncak dingin boss	di puncak dingin boss
Semangat jangan menyerah JAACCKKK	semangat jangan menyerah jaacckkk
Semangat	semangat
JAKARTA FULL SENYUM	jakarta full senyum
Semangat macan	semangat macan

Hasil *case folding* bisa dilihat pada Tabel 4.2, Dimana huruf kapital akan diubah atau diseragamkan menjadi huruf kecil semua. Seperti kata Di diubah menjadi di, kata JAKARTA diubah menjadi jakarta, dan lain-lain.

c. Tokenizing

Tokenizing bertujuan untuk memisahkan kata pada kalimat menjadi satuan kata atau biasa disebut “token”. Implementasi *tokenizing* menggunakan dari *library Python* yang dinamakan *split()*. Berikut merupakan *source code* dari proses *tokenizing*:

```
data['Comment Text'] = data['Comment Text'].str.split()
data
```

Gambar 4.7 Kode Program tahap *Tokenizing*

Berikut merupakan hasil sebelum dan sesudah proses *tokenizing* bertujuan untuk memudahkan pada proses pelabelan nanti.

Tabel 4.3 Hasil *Tokenizing* pada *comment text*

Ulasan	Setelah Proses <i>Tokenizing</i>
puncak dingin bos	[puncak, dingin, bos]
menang kalah tetap persija semangat macan	[menang, kalah, tetap, persija, semangat, macan]
semangat persijaku	[semangat, persijaku]
jakarta senyum	[jakarta, senyum]
semangat	[semangat]

d. Filtering

Filtering merupakan proses pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan proses untuk menghilangkan kata yang dianggap tidak memiliki makna. Kata tersebut meliputi kata penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak memiliki arti dalam penentuan kelas suatu dokumen. Berikut merupakan *source code* dari proses *filtering*:

```
path_stopword = "/content/drive/MyDrive/ANDRES/stopwordindo.txt"
with open(path_stopword, 'r', encoding='utf-8') as dataku:
    data_stopword = [i for i in dataku.read().splitlines()]

def Remove_Stopword(kalimat):
    hasil = [i for i in kalimat.split() if not i in data_stopword]
    return " ".join(hasil)

data['Comment Text'] = data['Comment Text'].apply(Remove_Stopword)
data
```

Gambar 4.8 Kode Program tahap *Filtering*

Didalam proses *filtering* bertujuan untuk menghilangkan kata yang dianggap tidak ada makna, contoh nya sebagai berikut:

Tabel 4.4 Hasil *Filtering* pada *comment text*

Ulasan	Setelah Proses <i>Filtering</i>
di puncak dingin bos	puncak dingin bos
menang kalah tetap gue persija semangat macan	menang kalah tetap persija semangat macan
semangat terus persijaku	semangat persijaku
jakarta full senyum	jakarta senyum

Kata-kata tidak bermakna seperti kata penghubung atau kata ganti orang akan dihapus, seperti kata di, gue, terus dan masih banyak lagi.

e. *Stemming*

Stemming adalah proses pengambilan kata yang memiliki makna dan mengubah kata ke bentuk dasarnya sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Pada tahap ini digunakan *library Python Sastrawi*. Berikut merupakan *source code* dari proses *stemming*:

```
factory = StemmerFactory()

stemmer = factory.create_stemmer()
def stemming(data, stemmer):
    return stemmer.stem(data)

data['Comment Text'] = data['Comment Text'].apply(lambda x: stemming(x, stemmer))
data
```

Gambar 4.9 Kode Program tahap *Stemming*

Bentuk kata tidak standar yang ditemukan pada data ulasan ada berbagai macam bentuk. Bentuk kata yang mengalami perbaikan adalah *slang word*, singkatan, salah penulisan kata (typo), bahasa asing, karakter berulang dan lainnya. Adapun daftar kata *slang*, singkatan, salah penulisan kata (*typo*), bahasa asing, dan kata karakter berulang beserta perbaikannya bisa dilihat pada Tabel 4.5, Tabel 4.6, Tabel 4.7, Tabel 4.8 dan Tabel 4.9, sebagai berikut:

Tabel 4.5 Contoh kata *slang* dan perbaikannya

<i>Slang word</i>	Perbaikan <i>Slang Word</i>
met	selamat
hahaha	tawa
salfok	salah fokus
sweger	segar
ucul	lucu

Contoh *slang word* yang diubah seperti kata met, hahaha, salfok, sweger dan ucul, diubah menjadi kata selamat, tawa, salah fokus, segar dan lucu.

Tabel 4.6 Contoh kata singkatan dan perbaikannya

Singkatan	Perbaikan Singkatan
sdh	sudah
smga	semoga
gtu	begitu
kpn	kapan
jg	juga

Tabel 4.7 Contoh kata *typo* dan perbaikannya

<i>Typo</i>	Perbaikan <i>Typo</i>
masi	masih
tetep	tetap
slalu	selalu
trus	terus
mantep	mantap

Contoh kata *typo* pada proses *stemming* adalah masi, tetep, slalu, trus dan mantep diperbaiki menjadi kata masih, tetap, selalu, terus dan mantap.

Tabel 4.8 Contoh kata asing dan perbaikannya

Kata Asing	Perbaikan Kata Asing
<i>answer</i>	jawab
<i>another</i>	lain
<i>good</i>	baik
<i>applied</i>	terapkan
<i>atleast</i>	paling sedikit

Tabel 4.9 Contoh karakter berulang dan perbaikannya

Karakter Berulang	Perbaikan Karakter Berulang
baaiikkk	baik
telaatt	telat
kereeeennnn	keren
asliiiiiiii	asli
gilaaaaaa	gila

4.3 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 3 kelas sentimen, yaitu sentimen positif, sentimen netral, dan 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 *Python* dengan cara menghitung skor jumlah kata positif dikurangi dengan skor jumlah istilah negatif dalam suatu kalimat ulasan. Bila suatu kalimat mempunyai skor label positif lebih banyak daripada label negatif maka akan diklasifikasikan dalam kelas positif. Jika kalimat mempunyai label negatif lebih banyak daripada label positif maka akan diklasifikasikan dalam kelas negatif. Dan jika jumlah label positif sama dengan label negatif maka akan diklasifikasikan sebagai label netral. Berikut *source code* proses pelabelan data pada data komentar Instagram Persija Jakarta:

```
def Kamus_Lexicon(text, positif, negatif):
    # Menghitung jumlah kata positif dan negatif dalam teks
    positive_count = sum(text.count(str(word)) for word in positif)
    negative_count = sum(text.count(str(word)) for word in negatif)

    # Menentukan label berdasarkan perbandingan jumlah kata positif dan negatif
    if positive_count > negative_count:
        return 'Positive'
    elif positive_count < negative_count:
        return 'Negative'
    else:
        return 'Neutral'

data['Label'] = data['Comment Text'].apply(lambda x: Kamus_Lexicon(x, positif,
negatif))
```

Gambar 4.10 Kode Program Proses Pelabelan Data

Jumlah kata pada kamus positif adalah 2151 kata dan pada kamus negatif berjumlah 2402 kata. Pelabelan sendiri dilakukan berdasarkan tiga kelas kategori yang berbeda yaitu kelas kategori sentimen positif, kelas kategori sentimen netral, dan kelas kategori sentimen negatif. Berikut adalah hasil proses pelabelan menggunakan kamus *lexicon* dari kamus positif dan negatif:

Tabel 4.10 Hasil Pelabelan kelas data

<i>Comment Text</i>	Label
jakarta senyum	Positive
alhamdulillah	Positive
puncak dingin bos	Neutral
selamat	Positive
semangat macan	Positive

Setelah melakukan pelabelan data, maka dihitung dan dibandingkan berapa banyak jumlah sentimen pada postingan ketika Persija Jakarta menang dan ketika Persija kalah, sebagai berikut:

```
positif = len([i for i in data['Label'] if i == 'Positive'])
negatif = len([i for i in data['Label'] if i == 'Negative'])
netral = len([i for i in data['Label'] if i == 'Neutral'])

positif
negatif
netral
```

Gambar 4.11 Kode Program Menentukan Jumlah Ulasan Sentimen

Source code yang terdapat pada Gambar 4.11 akan menjelaskan atau menentukan jumlah sentimen pada postingan Instagram Persija Jakarta. Jumlah ulasan sentimen akan diketahui karena sudah melalui tahap *data labelling* (pelabelan data). Pada Tabel 4.11 akan dijelaskan hasil dari jumlah ulasan sentimen pada postingan akun Instagram Persija Jakarta dalam kondisi Persija menang dan Persija kalah.

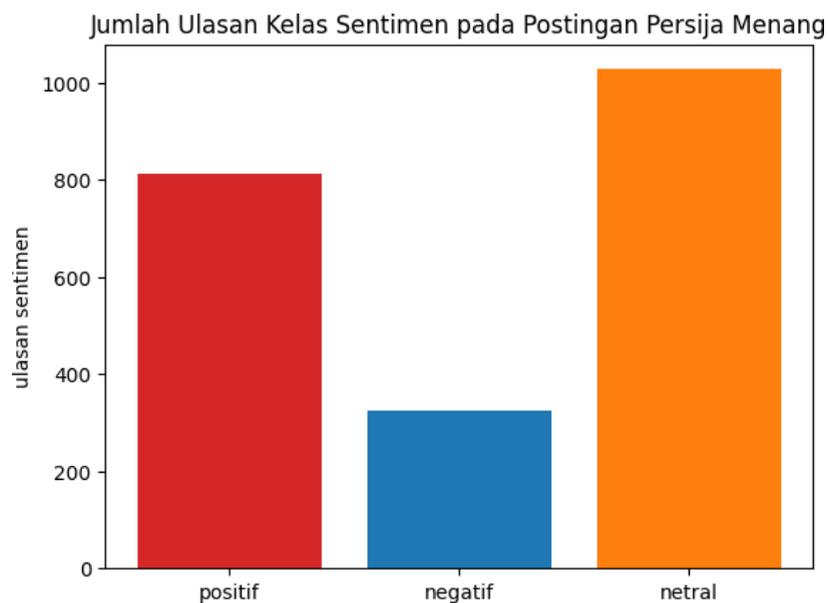
Berikut hasil dari *data train* jumlah setiap kelas sentimen positif, sentimen negatif, dan sentimen netral dari setiap postingan dalam kondisi Persija Jakarta menang dan kalah.

a. Postingan Persija Menang

Tabel 4.11 Jumlah ulasan setiap kelas sentimen pada postingan Persija menang

Kelas sentimen	Jumlah ulasan
POSITIF	811 (37,49%)
NEGATIF	324 (15,06%)
NETRAL	1.027 (47,44%)
TOTAL	2.162

Berikut merupakan grafik perbandingan ulasan positif, ulasan negatif dan ulasan netral pada komentar Instagram Persija Jakarta. Pada Gambar 4.12 menunjukkan bahwa pada saat Persija Jakarta menang dari total 2162 ulasan *data train* pada kelas sentimen terdapat 811 ulasan Positif, 324 ulasan Negatif dan 1.027 ulasan Netral.



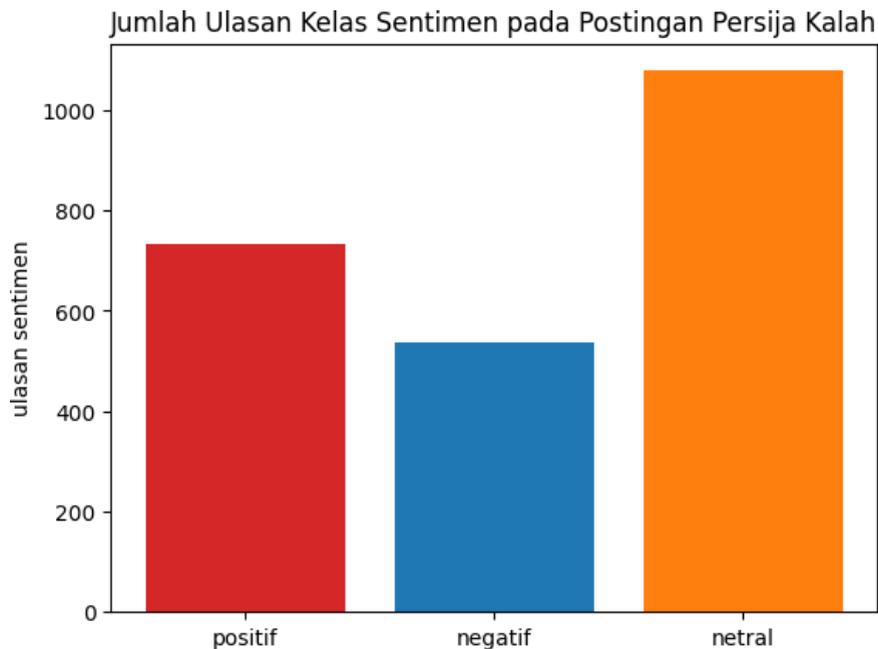
Gambar 4.12 Grafik Perbandingan Ulasan Sentimen pada Postingan Persija Menang

b. Postingan Persija Kalah

Tabel 4.12 Jumlah ulasan setiap kelas sentimen pada postingan Persija Kalah

Kelas sentimen	Jumlah ulasan
POSITIF	732 (33.94%)
NEGATIF	537 (22.03%)
NETRAL	1078 (44.02)
TOTAL	2.447

Bisa kita lihat pada Tabel 4.12, walaupun dalam kondisi Persija Jakarta sedang mengalami kekalahan, komentar pada akun Instagram Persija tetap lebih banyak sentimen positif daripada sentimen negatif. Berikut merupakan grafik perbandingan ulasan positif, ulasan negatif dan ulasan netral pada komentar Instagram Persija Jakarta. Pada Gambar 4.13 menunjukkan bahwa pada saat Persija Jakarta kalah dari total 2447 ulasan *data train* pada kelas sentimen terdapat 732 ulasan Positif, 537 ulasan Negatif dan 1.078 ulasan Netral.



Gambar 4.13 Grafik Perbandingan Ulasan Sentimen pada Postingan Persija Kalah

4.4 Klasifikasi dengan Metode Naïve Bayes Classifier

Sebelum melakukan klasifikasi terlebih dahulu menentukan atau membagi jumlah *data train* dan *data test*. Setelah dilakukan penentuan terkait data latih dan data uji selanjutnya dilakukan klasifikasi dengan metode *Naïve Bayes Classifier*. Analisis *Naïve Bayes Classifier* dilakukan dengan membuat *confusion matrix* agar dapat diketahui nilai akurasi, *recall* dan presisi. *Confussion matrix* merupakan salah satu *tools* penting yang digunakan pada *machine learning* untuk melakukan visualisasi dan biasanya memuat dua kategori atau lebih. *Confussion matrix* digunakan untuk mengevaluasi performansi model yang dibangun oleh algoritma klasifikasi.

a. Pembagian data latih dan data uji

```
X_train, X_test, y_train, y_test = train_test_split(data_vektor,
data['Label'], test_size=0.2, random_state=42)

len(y_train)

len(y_test)

positif = len([i for i in y_train if i == 2])
negatif = len([i for i in y_train if i == 0])
netral = len([i for i in y_train if i == 1])

print(f"- Jumlah Positif : {positif}\n- Jumlah Negatif : {negatif}\n-
Jumlah Netral : {netral}")
```

Gambar 4.14 Kode Program Proses membagi Data Latih dan Data Uji

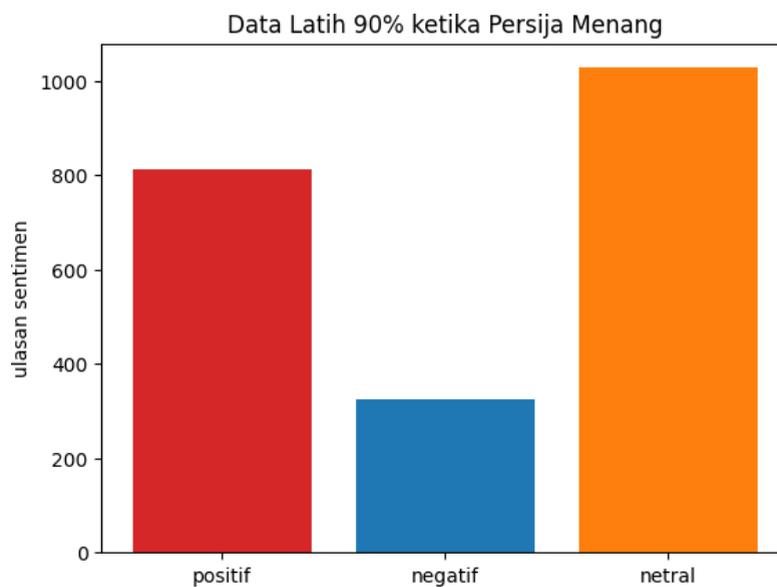
Semakin banyak data latih yang digunakan, maka akan semakin baik mesin dalam memahami pola data. Data uji digunakan untuk mengukur keakuratan *classifier* melakukan klasifikasi dengan benar. Data yang digunakan memiliki label kelas, dengan perbandingan data latih dan data uji 90% : 10%.

Berikut merupakan hasil perbandingan dari data sentimen positif, sentimen negatif, dan sentimen netral pada *data train*:

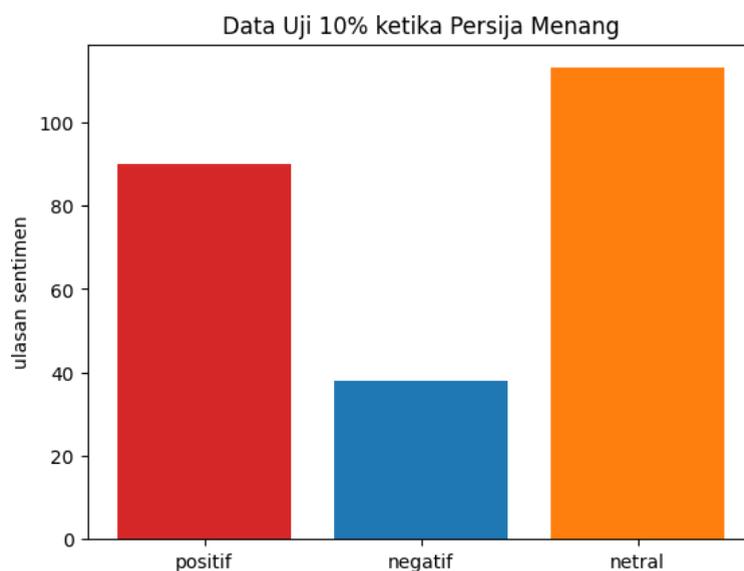
Tabel 4.13 Perbandingan data latih dan data uji ketika Persija Menang

Kelas/Label	Data latih 90%	Data Uji 10%
POSITIVE	811	90
NEGATIVE	324	38
NEUTRAL	1.027	113
TOTAL	2.162	241

Berikut merupakan Grafik perbandingan data latih 90% dan data uji 10% ketika Persija Jakarta menang, terdapat total 2.162 sentimen data latih dan 241 sentimen pada data uji. Pada Gambar 4.15 menunjukkan bahwa perbandingan sentimen positif, sentimen negatif dan sentimen netral pada data latih. Selanjutnya pada Gambar 4.16 menunjukkan bahwa perbandingan sentimen pada data uji, terdapat 90 sentimen positif, 38 sentimen negatif dan 113 sentimen netral.



Gambar 4.15 Grafik Perbandingan Jumlah Sentimen pada Data Latih ketika Persija Menang

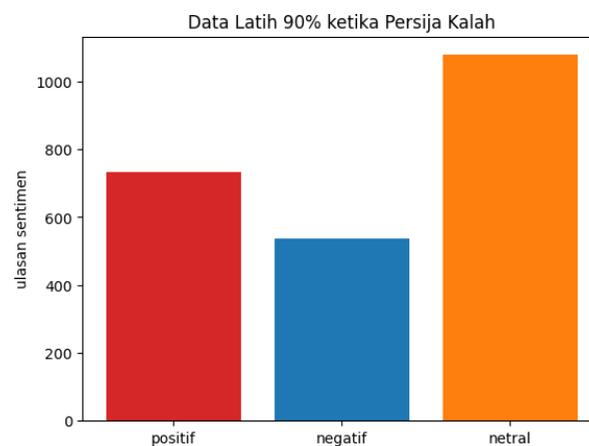


Gambar 4.16 Grafik Perbandingan Jumlah Sentimen pada Data Uji ketika Persija Menang

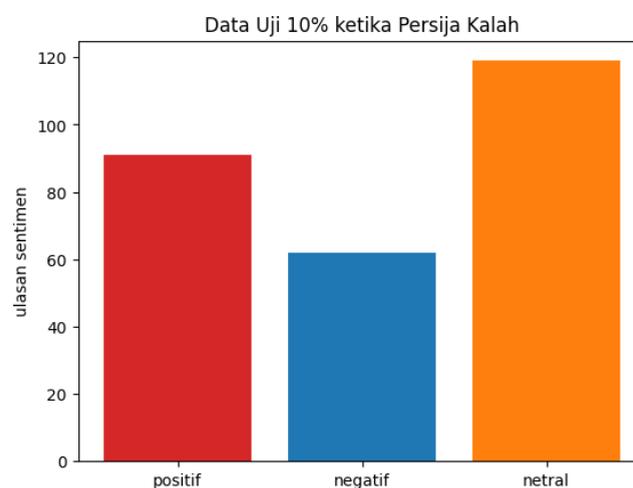
Tabel 4.14 Perbandingan data latih dan data uji ketika Persija Kalah

Kelas/Label	Data latih 90%	Data Uji 10%
POSITIVE	832	91
NEGATIVE	537	62
NEUTRAL	1.078	119
TOTAL	2.447	272

Berikut merupakan Grafik perbandingan data latih 90% dan data uji 10% ketika Persija Jakarta kalah, terdapat total 2.447 sentimen data latih dan 272 sentimen pada data uji. Pada Gambar 4.17 menunjukkan bahwa perbandingan sentimen positif, sentimen negatif dan sentimen netral pada data latih. Selanjutnya pada Gambar 4.18 menunjukkan bahwa perbandingan sentimen pada data uji, terdapat 91 sentimen positif, 62 sentimen negatif dan 119 sentimen netral.



Gambar 4.17 Grafik Perbandingan Jumlah Sentimen pada Data Latih ketika Persija Kalah



Gambar 4.18 Grafik Perbandingan Jumlah Sentimen pada Data Uji ketika Persija Kalah

b. Klasifikasi *Naïve Bayes Classifier*

Data latih yang telah dibentuk akan dipelajari melalui proses klasifikasi. Pada data latih terdapat ulasan negatif, ulasan netral, dan ulasan positif, kemudian pola data tersebut akan dipelajari menggunakan algoritma *Naïve Bayes Classifier* berdasarkan ciri-ciri data pada masing-masing label kelas yang telah dibentuk. Selanjutnya akan dilakukan proses *machine learning* yaitu hasil pelatihan pada algoritma *Naïve Bayes Classifier* akan diuji dengan pengujian *confussion matrix* untuk mengetahui nilai akurasi dalam memprediksi data baru. Pengujian dilakukan terbagi dua, pengujian pada data Persija Jakarta Menang dan saat Persija Jakarta Kalah. Berikut merupakan *source code* pengujian *confussion matrix*:

```
y_preds = model_g.predict(X_train)
akurasi = accuracy_score(y_train, y_preds)
print(confusion_matrix(y_train,y_preds))
print(classification_report(y_train,y_preds))
print(f'Akurasi Dari Model {akurasi*100} %')
```

Gambar 4.19 Kode Program Proses Pengujian *Confussion Matrix*

Proses evaluasi klasifikasi dilakukan dengan cara membuat *machine learning* menggunakan data latih dan data uji secara acak dan tentunya menggunakan metode *confussion matrix*. *Confussion matrix* digunakan untuk memudahkan dalam proses perhitungan akurasi dengan mengetahui jumlah data uji yang terklasifikasi dengan benar dan jumlah data uji yang salah pengklasifikasiannya. Hasil akurasi *data test* pada pengujian *confussion matrix* sebagai berikut:

Tabel 4.15 Hasil Akurasi data uji

Kondisi	Akurasi Data Train
Persija Menang	82,51%
Persija Kalah	80,09%

Tabel menjelaskan besar presentase akurasi pada *data test* pada postingan Persija Jakarta. Akurasi Tertinggi yaitu 82,51% didapatkan pada postingan ketika kondisi Persija Jakarta Menang dibandingkan ketika Persija Jakarta Kalah yang hanya mendapatkan 80,09%.

4.5 Visualisasi Data

Visualisasi dilakukan pada keseluruhan data, untuk *wordcloud* hanya pada kata yang paling sering muncul pada komentar Instagram Persija Jakarta. Visualisasi data dilakukan dengan tujuan untuk memperoleh informasi secara menyeluruh mengenai apa yang sering dibicarakan terkait topik yang dianalisis yaitu opini atau komentar pada postingan Instagram Persija Jakarta. Berikut pada Gambar 4.20 ditampilkan visualisasi kata yang paling sering muncul pada postingan Persija Jakarta:



Gambar 4.20 Visualisasi *Word Cloud* ketika Persija Jakarta Menang

Ketika Persija Jakarta menang, maka kata-kata yang paling sering muncul pada komentar Instagram seperti kata Persija, alhamdulillah, menang, main, mantap, senyum, semangat dan sebagainya. Semakin sering muncul kata-kata tersebut maka pada visualisasi data menggunakan *wordcloud* akan ditampilkan paling besar, begitu juga sebaliknya jika jarang muncul maka visualisasi datanya ditampilkan paling kecil. Berikutnya pada gambar 4.21 akan ditampilkan visualisasi menggunakan *word cloud* yaitu kata-kata yang paling sering muncul pada saat postingan kalah Persija Jakarta .



Gambar 4.21 Visualisasi *Word Cloud* ketika Persija Jakarta Kalah

Ketika Persija Jakarta kalah, maka kata-kata yang paling sering muncul pada kometar Instagram seperti kata Persija, kalah, menang, main, lawan, juara, semangat dan sebagainya. Semakin sering muncul kata-kata tersebut maka pada visualisasi data menggunakan *word cloud* akan ditampilkan paling besar, begitu juga sebaliknya jika jarang muncul maka visualisasi data nya ditampilkan paling kecil.

Visualisasi *word cloud* dapat di lihat dengan lebih jelas dengan kata-kata yang paling sering digunakan pengguna dalam memberikan komentar pada akun Instagram Persija Jakarta. Semakin besar ukuran kata pada *word cloud*, maka semakin tinggi pula frekuensi kata tersebut, artinya *followers* Instagram Persija Jakarta sering menggunakan kata tersebut sebagai topik yang sering dibicarakan atau kata yang paling banyak muncul pada komentar postingan akun Instagram Persija Jakarta.

4.5.1 Word Cloud Positif

Visualisasi dilakukan pada keseluruhan data, untuk *word cloud* positif hanya pada kata-kata positif yang paling sering muncul pada komentar Instagram Persija Jakarta. Visualisasi data dilakukan dengan tujuan untuk memperoleh informasi secara menyeluruh mengenai apa yang sering dibicarakan terkait topik yang dianalisis yaitu opini atau komentar pada postingan Instagram Persija Jakarta.

Berikut akan ditampilkan pada gambar 4.22 adalah hasil visualisasi data positif atau kata-kata positif yang selalu muncul pada postingan Persija Jakarta ketika menang, dan berikutnya pada gambar 4.23 adalah hasil visualisasi data positif ketika Persija Jakarta mengalami kekalahan.



Gambar 4.22 Visualisasi *Word Cloud* Positif ketika Persija Jakarta Menang



Gambar 4.23 Visualisasi *Word Cloud* Positif ketika Persija Kalah

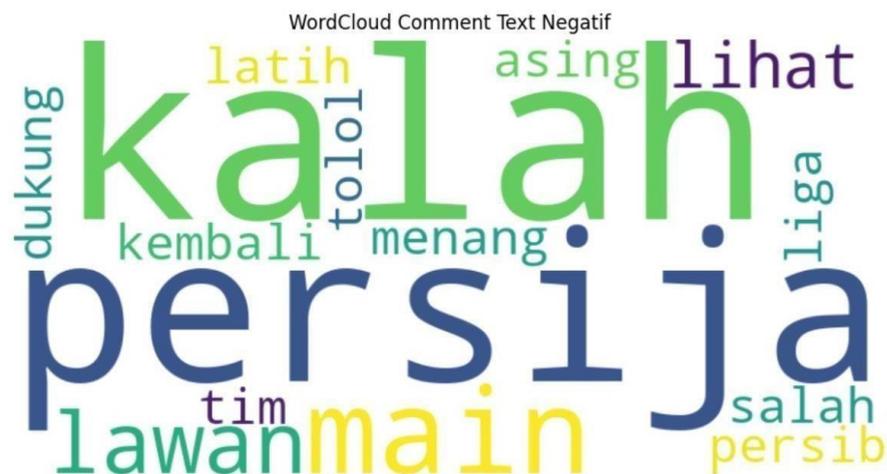
Kata-kata positif yang paling sering muncul pada postingan Instagram Persija Jakarta ketika kondisi Persija menang maupun kalah adalah kata alhamdulillah, semangat, selamat, baik dan lain-lain. Visualisasi *word cloud* positif dapat di lihat dengan lebih jelas dengan kata-kata yang paling sering digunakan pengguna dalam memberikan komentar pada akun Instagram Persija Jakarta. Semakin besar ukuran kata pada *word cloud*, maka semakin tinggi pula frekuensi kata tersebut, artinya *followers* Instagram Persija Jakarta sering menggunakan kata tersebut sebagai topik yang sering dibicarakan atau kata yang paling banyak muncul pada komentar postingan akun Instagram Persija Jakarta .

4.5.2 Word Cloud Negatif

Visualisasi dilakukan pada keseluruhan data, untuk *word cloud* negatif hanya pada kata-kata negatif yang paling sering muncul pada komentar Instagram Persija Jakarta. Visualisasi data dilakukan dengan tujuan untuk memperoleh informasi secara menyeluruh mengenai apa yang sering dibicarakan terkait topik yang dianalisis yaitu opini atau komentar pada postingan Instagram Persija Jakarta.



Gambar 4.24 Visualisasi *Word Cloud* Negatif ketika Persija Jakarta Menang



Gambar 4.25 Visualisasi *Word Cloud* Negatif ketika Persija Jakarta Kalah

Kata-kata negatif yang paling sering muncul pada postingan Instagram Persija Jakarta ketika kondisi Persija menang maupun kalah adalah kata kalah, asing, tolol, bau, gila dan lain-lain. Visualisasi *word cloud* negatif dapat di lihat dengan lebih jelas dengan kata-kata yang paling sering digunakan pengguna dalam memberikan komentar pada akun Instagram Persija Jakarta. Semakin besar ukuran kata pada *word cloud*, maka semakin tinggi pula frekuensi kata tersebut, artinya *followers* Instagram Persija Jakarta sering menggunakan kata tersebut sebagai topik yang sering dibicarakan atau kata yang paling banyak muncul pada komentar postingan akun Instagram Persija Jakarta.

4.6 Pengujian Analisis Sentimen

Dalam melakukan proses pengujian, peneliti menggunakan bahasa pemrograman Python pada *Google Colab*. Pengujian yang dilakukan bertujuan untuk memvalidasi atau menganalisis data ulasan, apakah ulasan tersebut termasuk ke dalam sentimen positif, sentimen negatif, maupun sentimen netral. Proses pengujian analisis sentimen bisa dilihat sebagai berikut:

```
v_data = vectorizer.transform(["akun siapa ini"]).toarray()
y_preds = model_g.predict(v_data)
if y_preds == 2:
    print('Positif')
elif y_preds== 1:
    print('Netral')
else:
    print("Negatif")

v_data = vectorizer.transform(["zonk"]).toarray()
y_preds = model_g.predict(v_data)
if y_preds == 2:
    print('Positif')
elif y_preds== 1:
    print('Netral')
else:
    print("Negatif")

v_data = vectorizer.transform(["main masih egois"]).toarray()
y_preds = model_g.predict(v_data)
if y_preds == 2:
    print('Positif')
elif y_preds== 1:
    print('Netral')
else:
    print("Negatif")
```

Gambar 4.26 Kode Program Pengujian Analisis Sentimen

Berikut hasil dari pengujian analisis sentimen atau proses prediksi, dapat dilihat pada Gambar 4.27 dan Gambar 4.28, sebagai berikut:

```
v_data = vectorizer.transform(["jakarta senyum"]).toarray()
y_preds = model_g.predict(v_data)
if y_preds == 2:
    print('Positif')
elif y_preds== 1:
    print('Netral')
else:
    print("Negatif")
```

Positif

Gambar 4.27 Uji Coba Sampel Data Positif

```
v_data = vectorizer.transform(["parah"]).toarray()
y_preds = model_g.predict(v_data)
if y_preds == 2:
    print('Positif')
elif y_preds== 1:
    print('Netral')
else:
    print("Negatif")
```

Negatif

Gambar 4.28 Uji Coba Sampel Data Negatif

Sistem dapat menunjukkan dan mengklasifikasi sampel data positif pada Gambar 4.27 dan dapat mengklasifikasikan data negatif yang terdapat pada Gambar 4.28. Jika di ambil beberapa komentar yang sudah berlabel positif dan negatif, sebagai contoh dicari komentar berlabel positif “jakarta senyum” maka akan didapatkan hasil Positif, sesuai pada pelabelan. Dan jika mencari kata “parah” maka hasil prediksi nya adalah Negatif sesuai dengan data ulasan berlabel Negatif. Jadi sistem dapat mengklasifikasikan data dengan baik dan dapat menampilkan hasilnya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis sentimen pada media sosial Instagram klub sepakbola Persija Jakarta yang telah dilakukan menggunakan algoritma *Naïve Bayes Classifier* dapat didapatkan kesimpulan yaitu:

- a. Dalam rentang waktu 22 Maret 2021 – 17 September 2023 dalam 60 postingan media sosial Instagram Persija Jakarta dengan perbandingan ketika Persija menang atau kalah didapatkan hasil analisis data komentar sebanyak 6.000 ulasan. 3.000 data ulasan untuk Persija Jakarta menang dan 3.000 data ulasan untuk Persija Jakarta kalah.
- b. Penerapan metode *Naïve Bayes Classifier*, dilakukan pembagian *data train* dan *data test* dengan perbandingan 90:10 yang artinya 90% *data train*, dan 10% *data test*.
- c. Ketika Persija Jakarta menang, jumlah kelas sentimen positif lebih banyak daripada sentiment negatif. Kemudian ketika Persija Jakarta kalah jumlah sentimen positif juga tetap lebih banyak daripada sentimen negatif. Perbandingan kelas sentimen positif atau negatif adalah 811:324 pada kondisi Persija Jakarta menang, dan perbandingan kelas sentimen positif dan negatif adalah 832:537 pada kondisi Persija Jakarta kalah.
- d. Hasil pengujian *confussion matrix* pada *data test* menghasilkan tingkat akurasi paling optimal sebesar 82,51% ketika kondisi Persija Jakarta menang.

5.2 Saran

Penulis menyarankan bahwa dengan adanya pendekatan menggunakan data jejak digital di media sosial dapat membantu untuk mendapatkan *insight* atau wawasan yang lebih baik terhadap apa yang sedang terjadi di media sosial. Lebih berhati-hati dan bijak dalam berkomentar di media sosial manapun, agar menjadi budaya positif untuk bersosmed kedepannya. Saran untuk penelitian selanjutnya dapat mengembangkan penelitian dengan menggunakan berbagai bahasa dikarenakan penelitian ini masih dibatasi oleh ulasan berbahasa Indonesia. Kemudian penelitian ini baru menggunakan satu algoritma yaitu *Naïve Bayes Classifier*, sehingga penelitian selanjutnya dapat dilakukan menggunakan pendekatan *machine learning* lainnya sebagai pembanding performa algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan ulasan.

Batasan dari penelitian ini yaitu akun yang dianalisis hanyalah akun resmi dari akun media sosial Instagram Klub sepak bola Persija Jakarta yang sudah *verified*, hal ini penting untuk dilakukan supaya terhindar dari miskonsepsi (salah paham atau salah sumber). Batasan selanjutnya yaitu data yang diperoleh memiliki scope yang berasal dari Indonesia saja. Akan lebih baik jika penelitian selanjutnya bisa direplikasikan dengan beberapa negara lain untuk dibandingkan hasil dan perbedaannya. Alhasil, penulis mengharapkan supaya dari penelitian ini dapat berguna untuk menjadikan sebuah acuan agar penelitian serupa dapat diteruskan dengan fokus yang berbeda di masa yang akan datang.

DAFTAR PUSTAKA

- Atenstaedt, R. (2012). Word cloud analysis of the BJGP. *British Journal of General Practice*, 62(596), 148. <https://doi.org/10.3399/bjgp12X630142>
- Bhardwaj, P. (2019). Types of sampling in research. *Journal of the Practice of Cardiovascular Sciences*, 5(3), 157. https://doi.org/10.4103/jpcs.jpcs_62_19
- Bustami, 2012, Teknik, D. I., & Bayes, N. (2018). “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah.” *Jurnal Penelitian Teknik Informatika Universitas Malikussaleh*, 146(Klasifikasi), 128–146.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- Imron, A. (2019). *ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DI KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*.
- Jang, H., & Lee, J. (2017). An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*, 6, 5427–5437. <https://doi.org/10.1109/ACCESS.2017.2779181>
- Juwiantho, H., Setiawan, E. I., Santoso, J., Purnomo, M. H., Informasi, D. T., Tinggi, S., & Surabaya, T. (2020). Sentiment Analysis Twitter Bahasa Indonesia Berbasis WORD2VEC Menggunakan Deep Convolutional Neural Network. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(1), 181–188. <https://doi.org/10.25126/jtiik.202071758>
- Liang, H., & Zhu, J. J. H. (2017). Big Data, Collection of (Social Media, Harvesting). *The International Encyclopedia of Communication Research Methods*, 1–18. <https://doi.org/10.1002/9781118901731.iecrm0015>
- Liu, B. (2013). A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS. *Sentiment Analysis*, 9781461432, 1–522. <https://doi.org/10.1007/978-1-4614-3223>
- Longobardi, C., Settanni, M., Fabris, M. A., & Marengo, D. (2020). Follow or be followed: Exploring the links between Instagram popularity, social media addiction, cyber victimization, and subjective happiness in Italian adolescents. *Children and Youth Services Review*, 113 (December 2019), 104955. <https://doi.org/10.1016/j.childyouth.2020.104955>

- Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- Molina, A., Gómez, M., Lyon, A., Aranda, E., & Loibl, W. (2020). What content to post? Evaluating the effectiveness of Facebook communications in destinations. *Journal of Destination Marketing and Management*, 18(September). <https://doi.org/10.1016/j.jdmm.2020.100498>
- Naradhipa, A. R., & Purwarianti, A. (2012). Sentiment classification for Indonesian message in social media. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*. <https://doi.org/10.1109/ICEEI.2011.6021696>
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Naïve Bayes.
- Nugroho, D. G., & , Yulison Herry Chrisnanto, A. W. (2016). *Analisis Sentimen pada JasaOjek Online*. 156–161.
- Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016). Text Mining Dalam Analisis SentimenAsuransi Menggunakan Metode Niave Bayes Classifier. *Prosiding SNST*, 7, 37–42.
- P Priatna. (2019). *Peran Persija pers dalam membangun opini publik persija jakarta*. 2, 1–13.
- Pertiwi, M. W. (2019). 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*, 14(1), 27–32.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *JurnalTeknologi Informasi Dan Ilmu Komputer*, 5(3), 269. <https://doi.org/10.25126/jtiik.201853810>
- Pranandha Syah, A., Adiwijaya, & Al Faraby, S. (2017). Analisis Sentimen Pada Data Ulasan Produk Toko Online Dengan Metode Maximum Entropy Sentiment Analysis on Online Store Product Reviews With Maximum. *Proceeding of Engineering (E-Proceeding)*, 4(3), 4632–4640.

- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2789–2797.
- Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91. <https://doi.org/10.22146/ijccs.3499>
- Riefky, M., & Pramesti, W. (2020). Sentiment Analysis of Southeast Asian Games (SEA Games) in Philippines 2019 Based on Opinion of Internet User of Social Media Twitter with K-Nearest Neighbor and Support Vector Machine. *Jurnal Matematika, Statistika Dan Komputasi*, 17(1), 26–41. <https://doi.org/10.20956/jmsk.v17i1.9947>
- Roy, K., Simatupang, P., Studi, P., Komunikasi, I., Komunikasi, F., Informatika, D. A. N., & Surakarta, U. M. (2022). *Strategi Digital Branding Tim Persija Jakarta Melalui*.
- Rozi, I. F., Hadi, S., & Achmad, E. (2012). *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi Imam*. 6(1), 37–43.
- S Kurniawan, W Gata, D. P. (2017). Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI*, 1(3), 176–183.
- Sabily, A. F., Adikara, P. P., & Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(5), 4204–4209.
- Sapozhnikova, L. E., & Gordeeva, O. A. (2019). Text classification using convolutional neural network. *CEUR Workshop Proceedings*, 2416, 219–226. <https://doi.org/10.18287/1613-0073-2019-2416-219-226>
- Setiawan, A., Astuti, I. F., & Kridalaksana, A. H. (2016). Klasifikasi Dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier (NBC) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur). *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 10(1), 1. <https://doi.org/10.30872/jim.v10i1.17>
- Sharma, D., Sabharwal, M., Goyal, V., & Vij, M. (2020). *Sentiment analysis techniques for social media data: A review. Advances in Intelligent Systems and Computing*.

- Sohrabi, M. K., & Hemmatian, F. (2019). An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-019-7586-4>
- Spillane, T. E., Wong, B. A., & Giovenco, D. P. (2021). Content analysis of instagram posts by leading cannabis vaporizer brands. *Drug and Alcohol Dependence*, 218(June), 108353. <https://doi.org/10.1016/j.drugalcdep.2020.108353>
- Sugiyono. (2015). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D*. Suryani, P. S. M., Linawati, L., & Saputra, K. O. (2019). Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, 18(1), 145. <https://doi.org/10.24843/mite.2019.v18i01.p22>
- Umi Rofiqoh, Rizal Setya Perdana, M. A. F. (2017). 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. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12)(October), 1725–1732.
- Van Den Rul, C. (2019). How to Generate Word Clouds in R. *Data Science*.
- Wicaksono, B. I. (2019). Sejarah persepakbolaan di yogyakarta (1929-1943). *Jurnal Prodi Ilmu Sejarah UNY*, 1–15.
- Yoo, S. Y., Song, J. I., & Jeong, O. R. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, 105, 102–111. <https://doi.org/10.1016/j.eswa.2018.03.055>

LAMPIRAN

Lampiran A. Source Code Penelitian

Untuk dataset, kamus alay, kamus positif dan negatif, *stopwords* pada penelitian ini dapat diakses di tautan dibawah ini.

Source Code Persija Menang:

https://drive.google.com/file/d/14l_R6lhLGVcNDmzEn3_QaSty8G5yf7cO/view?usp=sharing

Source Code Persija Kalah:

https://drive.google.com/file/d/1Ym3Cj3_4pXZ2P8RNJxWbmEX0e34i01qf/view?usp=sharing

Dataset menang :

<https://drive.google.com/drive/folders/1RvzDceYKkEKD4svks7pe5mfn5AioYZyO?usp=sharing>

Dataset Kalah :

https://drive.google.com/drive/folders/1MBFouekxNvSkCdfVdGrGrwGWzwex_wXd?usp=sharing

Kamus Positif :

<https://drive.google.com/file/d/1r130249OKQpM4FhfBDHG3FwWBYA1e-by/view?usp=sharing>

Kamus Negatif :

<https://drive.google.com/file/d/1Uk7gUptmr3tjUM-zumIbdMcCxLmlfp6k/view?usp=sharing>

Kamus Alay :

https://drive.google.com/file/d/1IyosXmS9GrMQPESiUxF_PjsqPaocdnYW/view?usp=sharing

Stopwords :

https://drive.google.com/file/d/1viXX6pp_fvC-ZBQiVwj7THx_R0-321Qn/view?usp=sharing