

BAB III METODE PENELITIAN

3.1 Waktu Dan Tempat Penelitian

Waktu penelitian ini berlangsung dari Januari 2024 hingga 2025. Peningkatan perhatian terhadap Maarten Paes sebagai pemain diaspora dimulai setelah beberapa pencapaian pentingnya yang viral di media sosial pada Awal 2024. Penelitian ini dilakukan di platform Twitter dengan memanfaatkan data yang relevan mengenai sentimen netizen Indonesia terhadap Maarten Paes dan interaksinya dengan publik.



Gambar 3.1 Gambar Platform Twitter

3.1.1 *Twitter*

Penelitian ini berfokus pada Twitter, yang juga dikenal sebagai X, salah satu platform media sosial yang paling populer untuk berbagi opini dan diskusi di kalangan masyarakat Indonesia. Twitter memberikan ruang untuk pengguna berbagi pendapat secara real-time melalui cuitan, balasan, dan interaksi lainnya seperti retweet dan likes. Keunggulan Twitter sebagai platform utama dalam penelitian ini terletak pada kecepatan penyebaran informasi dan kemampuan untuk menjangkau berbagai segmen masyarakat, termasuk penggemar sepak bola. Dengan karakteristik ini, Twitter merupakan medium yang ideal untuk menganalisis sentimen publik terhadap maarten paes sebagai pemain diaspora.

Tautan: <https://twitter.com/login>

Tabel 3. 1 Jadwal Penelitian

NO	Jenis Penelitian	Tahun 2024												Tahun 2025											
		Oktober				November				Desember				Januari				Februari							
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
1.	Pengusulan Judul skripsi																								
2.	Implementasi Penelitian																								
3.	Penyusunan Proposal Skripsi																								
4.	Seminar Proposal Skripsi																								
5.	Observasi dan pengumpulan Data																								
6.	Penyusunan Skripsi																								
7.	Ujian Skripsi																								

3.2 Alat dan Bahan

3.2.1 Alat Penelitian

Perangkat keras laptop dengan spesifikasi minimumnya sebagai berikut:

1. *Processor* Intel(R) Core(TM) i3-10300H CPU @ 1.80GHz 900 GHz
2. Ram 4 GB
3. Penyimpanan SSD dengan kapasitas 256 GB
4. Jenis sistem operasi 64-bit, prosesor berbasis x64
5. Menggunakan edisi windows 11

1. Perangkat Lunak dan Alat Digital

a. **ORANGE data Mining**

Sebuah Platform *open source* untuk *machine learning* dan analisis data yang menggunakan pendekatan pemrograman visual (*visual programming*). Orange digunakan sebagai alat utama untuk seluruh alur kerja penelitian, mulai dari membaca data, *preprocessing* teks (termasuk *stopword removal*), membangun model klasifikasi naïve bayes, hingga evaluasi performa model dan visualisasi hasil seperti *Word Cloud*.

b. **Library Stopword Bahasa Indonesia dari GitHub**

Untuk mendukung proses pembersihan data, peneliti mengunduh daftar stopwords bahasa Indonesia dari repositori GitHub. Daftar ini berisi kata-kata umum yang tidak membawa informasi penting untuk analisis sentimen, seperti "yang", "dan", "di", dan lainnya. File stopwords diintegrasikan ke dalam workflow Orange, khususnya pada tahap Text Preprocessing.

c. **Selenium dengan Visual Studio Code**

Digunakan untuk proses scraping data komentar dari Twitter secara otomatis.

d. **Browser Web**

Untuk mengakses twitter serta mengunduh data

e. **Koneksi Internet**

Dibutuhkan untuk proses scraping dan pengunduhan data tambahan.

3.2.2 Bahan Penelitian

1. **Data Komentar Twitter**

Sebanyak 999 tweet yang berisi opini publik terhadap Maarten Paes, baik yang bersentimen positif, negatif, maupun netral.

2. **Kamus Stopwords Bahasa Indonesia**

Daftar stopwords yang diunduh dari GitHub untuk proses filtering data teks.

3.3 Jenis dan Pendekatan Penelitian

Tujuan dari studi ini adalah untuk mengevaluasi performa sebuah alat analisis sentimen otomatis dalam mengkategorikan komentar berbahasa Indonesia, dengan menggunakan pendekatan kuantitatif. Penelitian ini secara spesifik mengukur tingkat akurasi dari perangkat lunak Orange yang berbasis leksikon. Pengukuran kinerja ini dilakukan dengan cara memvalidasi hasil klasifikasi otomatis terhadap sebuah dataset acuan (*ground truth*) yang telah dilabeli secara manual oleh peneliti ke dalam kelas Positif, Negatif, dan Netral.

Alur kerja penelitian ini berpedoman pada kerangka CRISP-DM, dengan penyesuaian pada beberapa fase agar selaras dengan pendekatan analisis berbasis leksikon. Tahapannya adalah sebagai berikut:

1. Pemahaman Konteks Bisnis (*Business Understanding*)

Fase ini difokuskan pada perumusan masalah dan penetapan tujuan penelitian, yakni untuk menganalisis dinamika sentimen publik terhadap pemain diaspora seperti Maarten Paes, serta memahami implikasinya pada citra atlet di media sosial.

2. Pemahaman Data (*Data Understanding*)

Mencakup kegiatan akuisisi data mentah dari platform Twitter berdasarkan kata kunci yang relevan. Eksplorasi awal dilakukan pada data ini untuk memperoleh pemahaman mendasar mengenai volume dan karakteristik umum dari opini yang ada.

3. Persiapan Data (*Data Preparation*)

Tahap ini terdiri dari dua aktivitas krusial. Pertama, pembersihan teks (*preprocessing*) dilakukan pada keseluruhan dataset untuk mengeliminasi *noise* (seperti URL dan tanda baca) dan melakukan normalisasi teks. Kedua, sebuah subset data 999 tweet dianotasi secara manual oleh peneliti sesuai Panduan Anotasi untuk membangun sebuah dataset referensi (*ground truth*) yang valid.

4. Anotasi Sentimen Otomatis (*Automated Sentiment Analysis*)

Teks yang telah melalui tahap pembersihan kemudian diproses oleh widget *Sentiment Analysis* pada platform Orange. Perangkat lunak ini secara

otomatis memberikan "Label Prediksi" pada setiap tweet dengan memanfaatkan metode berbasis leksikon internalnya.

5. Evaluasi Kinerja (*Performance Evaluation*)

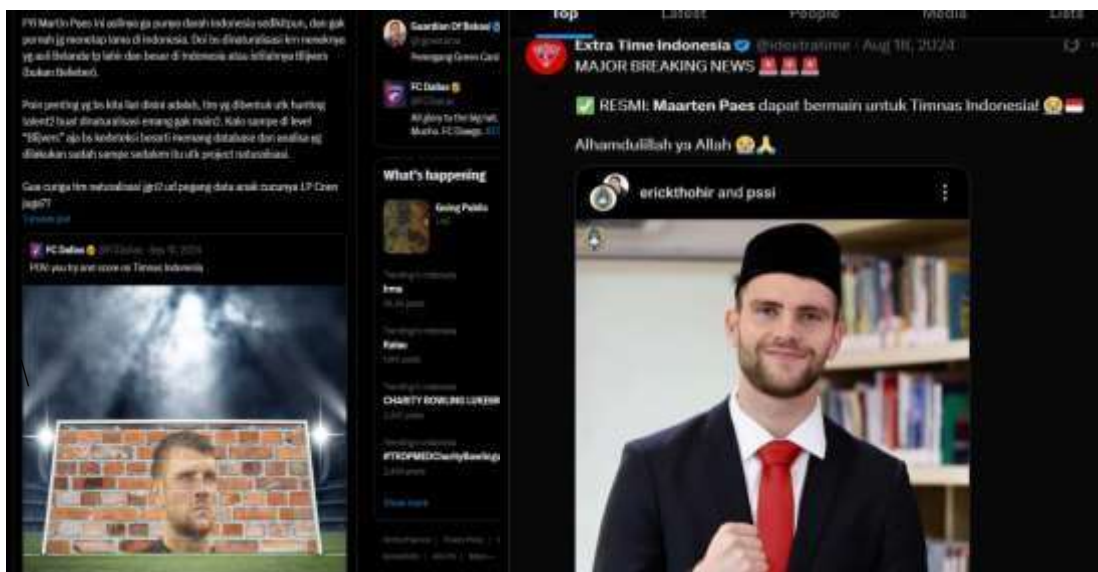
Merupakan tahap inti di mana "Label Prediksi" dari mesin divalidasi dengan cara membandingkannya terhadap "Label Manual" dari dataset *ground truth*. Pengujian ini menggunakan widget *Test & Score* untuk menghasilkan *Confusion Matrix*, yang kemudian digunakan untuk mengkalkulasi metrik-metrik evaluasi kuantitatif seperti akurasi, presisi, dan *recall*.

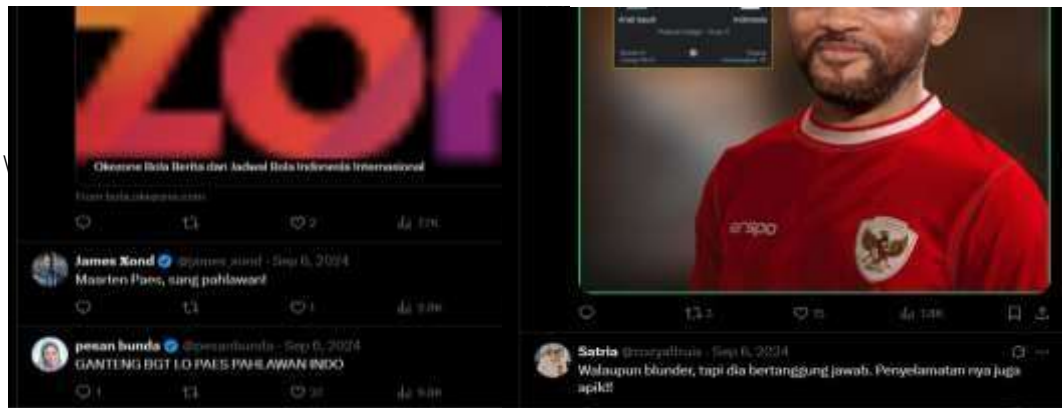
6. Penyajian Hasil (*Deployment*)

Tahap akhir yang meliputi penyajian temuan dari evaluasi kinerja dalam format laporan. Data pendukung seperti visualisasi distribusi sentimen dan *Word Cloud* juga disertakan untuk memberikan konteks serta membantu dalam interpretasi dan penarikan kesimpulan mengenai reliabilitas alat analisis sentimen otomatis pada studi kasus ini.

3.4 Sumber Data

Data untuk penelitian ini diperoleh dari opini publik di platform media sosial Twitter yang membahas Maarten Paes. Pemilihan platform tersebut didasarkan pada kemampuannya menyajikan informasi secara real-time dan statusnya sebagai media yang umum digunakan publik untuk merespons berbagai isu, termasuk olahraga.





Gambar 3.2 Sumber Data Dari Halaman Twitter

3.4.1 Alasan Pemilihan Maarten Paes serta Pertimbangannya sebagai Objek Penelitian

Maarten Paes merupakan pesepakbola profesional kelahiran Belanda yang menjalani proses naturalisasi sebagai Warga Negara Indonesia (WNI) pada awal tahun 2024. Sebelum bergabung dengan Tim Nasional Indonesia, ia meniti karier di kancah sepak bola Eropa, bermain untuk *FC Utrecht* di *Eredivisie*, dan kemudian melanjutkan kariernya di klub *Major League Soccer (MLS)*, *FC Dallas*. Tak hanya itu, Paes juga pernah memperkuat Tim Nasional Belanda kelompok usia U-21, yang menunjukkan kapasitasnya sebagai pemain dengan kualitas internasional.

Pemilihan Maarten Paes sebagai objek dalam penelitian ini didasari oleh sejumlah pertimbangan kontekstual dan akademik sebagai berikut:

1. Fenomena Naturalisasi Atlet *Diaspora*

Kehadiran Maarten Paes menjadi bagian dari tren naturalisasi pemain diaspora yang memperkuat Tim Nasional Indonesia. Fenomena ini telah menimbulkan berbagai diskursus publik, khususnya mengenai makna nasionalisme dan keterbukaan terhadap identitas baru dalam ranah olahraga.

2. Sorotan Publik dan Viralitas Digital

Proses naturalisasi serta debut Maarten Paes di bawah bendera Indonesia mendapat perhatian luas di berbagai platform media sosial, khususnya Twitter. Keberadaannya secara aktif diperbincangkan, baik dalam bentuk dukungan, antisipasi, hingga kritik terhadap kebijakan naturalisasi itu sendiri.

3. Signifikansi Analisis Sentimen Netizen Indonesia

Melalui keputusannya membela Indonesia, Maarten Paes menjadi cermin bagaimana publik Indonesia merespons figur asing yang memilih untuk menjadi bagian dari identitas nasional. Analisis sentimen terhadap dirinya mencerminkan dinamika penerimaan sosial dan konstruksi citra seorang atlet diaspora dalam ruang digital.

4. Relevansi Kontekstual dan Akademik

Tokoh seperti Maarten Paes menawarkan nilai strategis sebagai subjek kajian dalam konteks komunikasi publik, studi media, dan representasi sosial. Dengan posisinya yang unik—baik secara karier maupun identitas—Paes membuka ruang untuk menggali persepsi publik terhadap isu kebangsaan, performa atlet, dan dinamika pencitraan di media sosial.



Gambar 3.3 Maarten Paes resmi menjadi Warga Negara Indonesia (kiri) dan saat bermain bersama klubnya FC Dallas (kanan). Status resminya sebagai WNI diumumkan pada Agustus 2024 dan dikonfirmasi oleh PSSI serta Ketua

Umum Erick Thohir, yang menjadi titik awal keterlibatannya bersama Tim Nasional Indonesia.

(Sumber: @FaktaSepakbola via Twitter/X, 18 Agustus 2024)

Dengan demikian, peneliti memandang bahwa tokoh Maarten Paes relevan secara empiris dan konseptual untuk ditelaah melalui pendekatan analisis sentimen. Pemilihan ini diharapkan dapat memberikan kontribusi dalam memahami bagaimana masyarakat Indonesia membentuk dan memproyeksikan opini terhadap figur diaspora di era keterbukaan informasi digital.

3.5 Metode Pengumpulan Data

Penelitian ini menggunakan dua jenis metode pengumpulan data, yaitu studi literatur dan pengumpulan data sekunder dari media sosial.

a. Studi Literatur

Studi literatur dilakukan untuk mengumpulkan informasi dan teori yang relevan dengan topik penelitian. Peneliti melakukan review pada berbagai sumber literatur, termasuk jurnal ilmiah, buku, dan artikel yang terkait dengan analisis sentimen, media sosial, dan olahraga. Studi literatur ini bertujuan untuk memahami konsep dan teori yang terkait dengan sentimen masyarakat terhadap Maarten Paes sebagai pemain diaspora, serta untuk mengidentifikasi gap yang ada dalam penelitian sebelumnya.

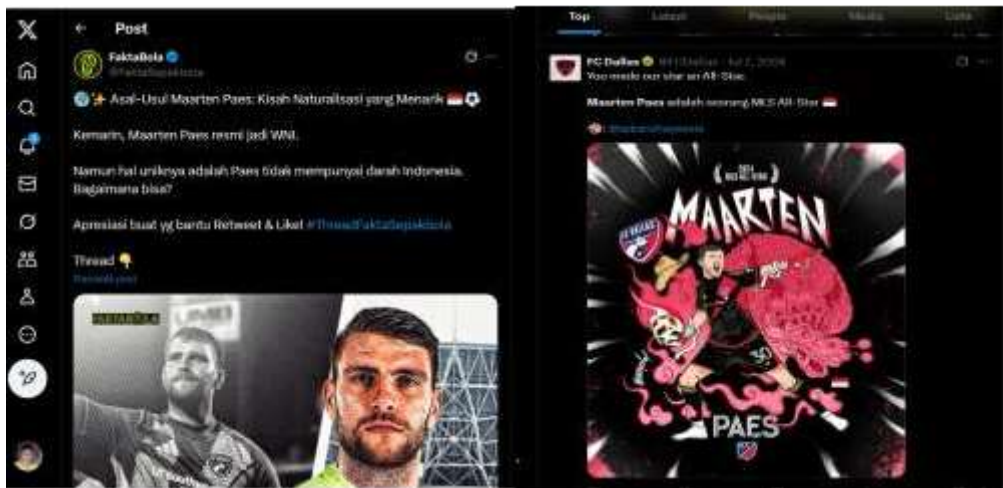
b. Akuisisi Data dari Twitter melalui Crawling

Fokus utama penelitian adalah pada interaksi digital dari pengguna Twitter di Indonesia untuk memahami perspektif netizen lokal terhadap Maarten Paes. Meskipun pengumpulan data bersifat daring dan tidak terikat lokasi fisik, data yang dianalisis secara geografis berasal dari konteks Indonesia dan dianggap merepresentasikan pandangan publik secara luas.

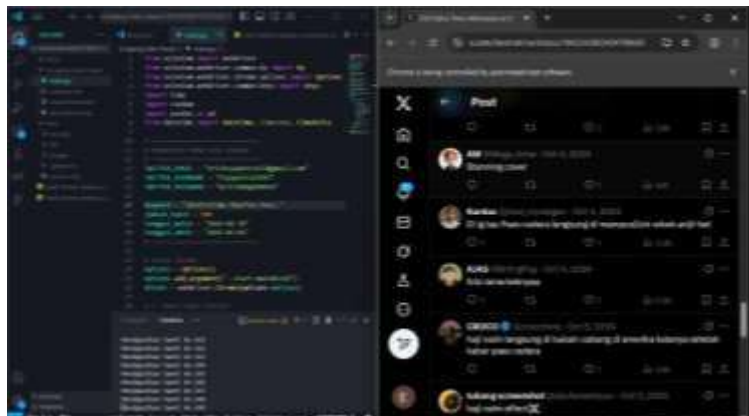
Pengumpulan data untuk penelitian ini dilakukan dengan metode *scraping* secara otomatis dari platform media sosial Twitter. Untuk mengatasi keterbatasan akses pada API resmi Twitter/X, penelitian ini menerapkan sebuah skrip *web scraping* yang dijalankan melalui Visual Studio Code. Skrip tersebut dibangun

menggunakan bahasa pemrograman Python dengan memanfaatkan library Selenium untuk mengekstraksi data komentar secara efisien.

Fokus *scraping* adalah pada komentar dan interaksi yang berkaitan dengan Maarten Paes khususnya mengenai proses naturalisasi, prestasi, dan performanya dengan menggunakan kata kunci yang relevan. Data mentah yang terkumpul kemudian divalidasi dan disaring untuk memastikan relevansi konten sebelum dilanjutkan ke tahap analisis.dengan menggunakan kata kunci yang relevan. Dengan memanfaatkan Selenium melalui Visual Studio Code, data dikumpulkan dari Twitter menggunakan kata kunci seperti Maarten Paes, Timnas Indonesia, dan diaspora. Data disimpan dalam format ExCeL.



Gambar 3.4 Postingan yang membahas pada kisah unik Maarten Paes dalam pengambilan sumpah WNI-nya serta prestasi dan performanya, Gambar menunjukkan postingan yang berkaitan dengan pembahasan yang akan dianalisis.



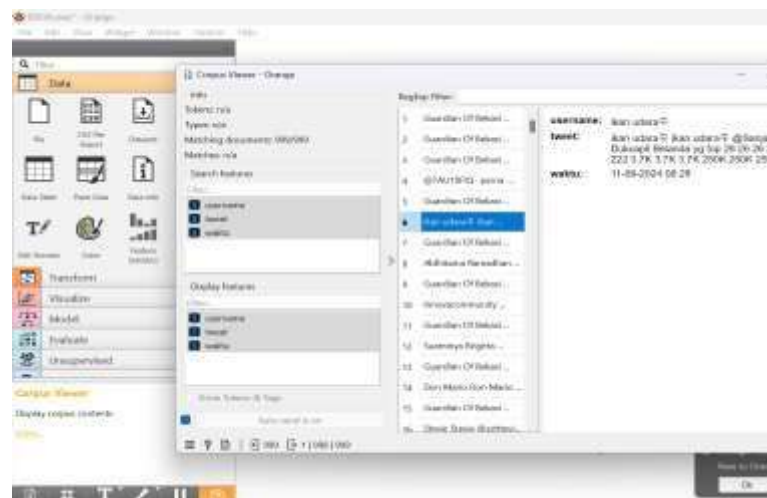
Gambar 3.5 Proses *Scraping* Data Menggunakan Skrip *Python* dan *Selenium*

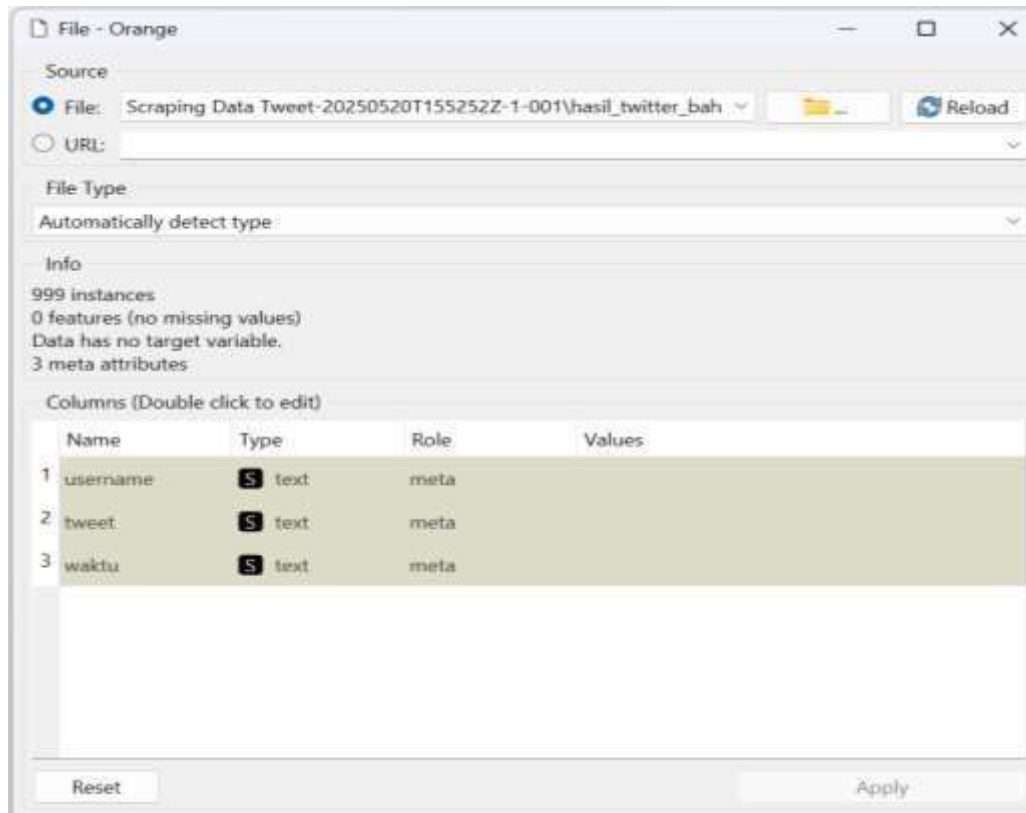
Gambar 3.4 menunjukkan implementasi dari metode pengumpulan data menggunakan skrip otomatis. Proses ini dijalankan di lingkungan Visual Studio Code dengan memanfaatkan library Selenium pada Python. Skrip secara otomatis membuka peramban, menavigasi ke URL target di Twitter, dan melakukan *scrolling* untuk memuat semua komentar yang ada. Setelah semua data termuat, skrip kemudian mengekstraksi teks komentar beserta metadata relevan lainnya dan menyimpannya ke dalam sebuah file Excel untuk tahap analisis selanjutnya.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
548	Petiri	Petiri	Fri 19-11-2024	20:38										
549	jar444	jar444	Fri 19-11-2024	20:38										
550	Andi harte	Andi harte	Fri 19-11-2024	20:38										
551	s	s	Fri 19-11-2024	20:39										
552	ameng	ameng	Fri 19-11-2024	20:39										
553	lvoe	lvoe	Fri 19-11-2024	20:39										
554	sm_sm	sm_sm	Fri 19-11-2024	20:39										
555	Paman Sr	Paman Sr	Fri 19-11-2024	20:39										
556	55655555	55655555	Fri 19-11-2024	20:39										
557	Hey you	Hey you	Fri 19-11-2024	20:39										
558	Afforda	Afforda A	Fri 19-11-2024	20:39										
559	nuel	nuel	Fri 19-11-2024	20:39										
560	8-888%	8-888%	Fri 19-11-2024	20:39										
561	naufal	naufal	Fri 19-11-2024	20:39										
562	Dh	Dh Dh	Fri 19-11-2024	20:43										
563	Al	Al	Fri 19-11-2024	20:44										
564	yoyol	yoyol	Fri 19-11-2024	20:44										
565	PositivTh	PositivTh	Fri 19-11-2024	20:46										
566	misy	misy	Fri 19-11-2024	20:47										
567	spidermar	spidermar	Fri 19-11-2024	21:11										
568	aditya	aditya	Fri 19-11-2024	21:15										
569	Nanda Kti	Nanda Kti	Fri 19-11-2024	21:36										
570	Rev	Rev	Fri 19-11-2024	20:41										
571	Memedaz	Memedaz	Fri 19-11-2024	20:39										

Gambar 3.6 Hasil *crawling* data dalam bentuk format Excel

Bisa kita lihat di gambar 3.5 hasil *crawling* yang telah dicrawling sebelumnya dapat diakses melalui *Microsoft Excel*.



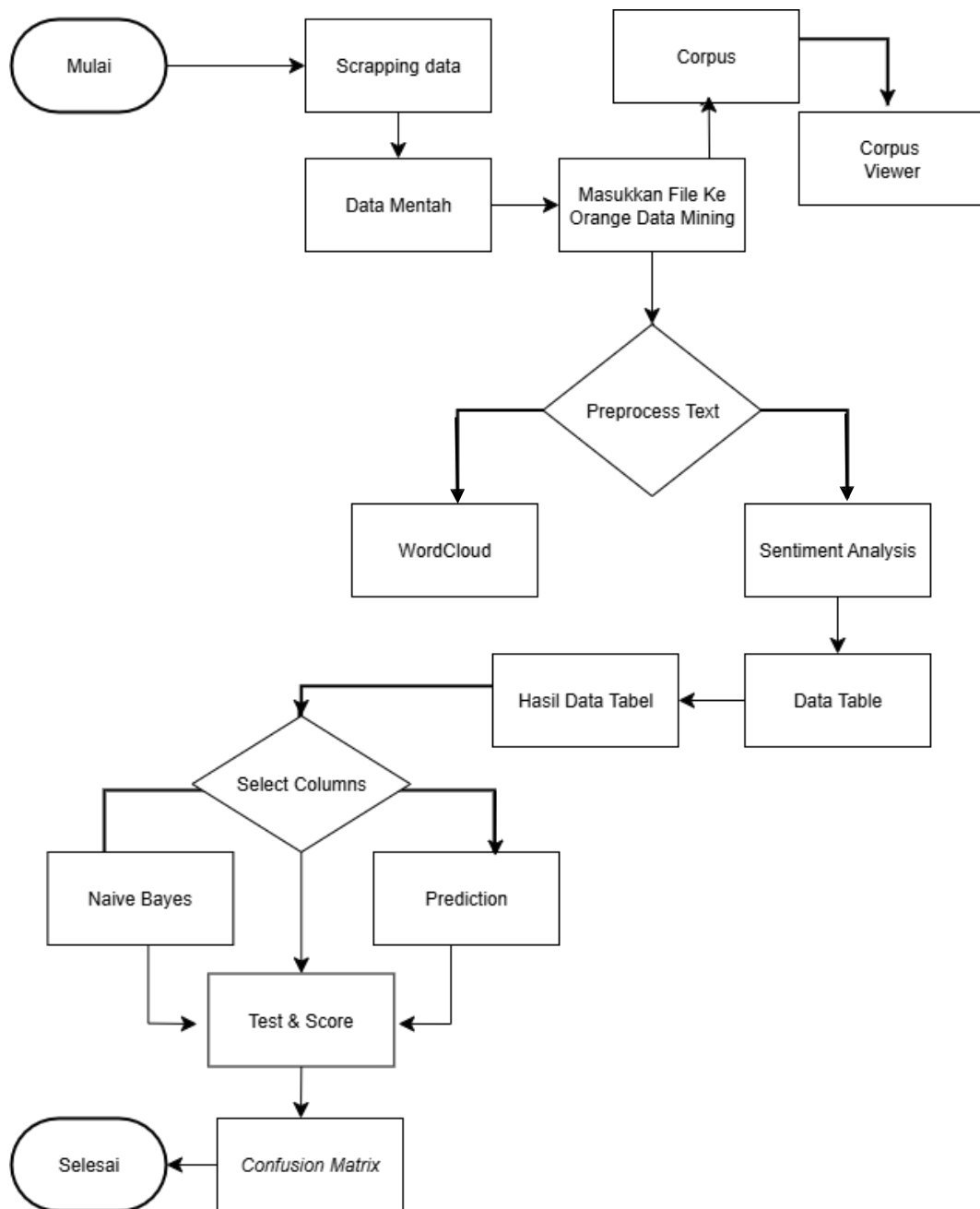


Gambar 3.7 Contoh Data latih setelah Proses pelabelan

Data mentah yang diperoleh tersebut kemudian melalui tahap preprocessing dan pelabelan manual untuk membentuk data latih. Proses ini bertujuan untuk menyiapkan data yang akan digunakan oleh model Naive Bayes. Format data setelah melalui proses pelabelan dapat dilihat pada Gambar 3.6 diatas.

3.6 Tahapan Kerja penelitian

Penelitian ini dilakukan dengan mengikuti langkah-langkah sistematis untuk memastikan data dikumpulkan, diproses, dianalisis, dan disajikan secara efektif. Prosedur penelitian mengacu pada kerangka kerja CRISP-DM, serta menggunakan Platform *Orange Data Mining* untuk melakukan pemrosesan data secara visual dan interaktif, Alur lengkapnya dapat dilihat pada Gambar 3.7 dibawah ini :



Gambar 3.8 Tahapan Kerja Penelitian

1. Pengumpulan Data (*Data Collection*)

Data diambil dari platform media sosial Twitter menggunakan metode *crawling* dengan kata kunci yang relevan, seperti Maarten Paes, Timnas Indonesia, dan diaspora. Komentar-komentar yang terkumpul kemudian disimpan sebagai data mentah untuk proses selanjutnya.

2. Persiapan Data dan Pembuatan *Ground Truth*

Data mentah yang terkumpul dibersihkan dari *noise* seperti mention, hashtag, dan tanda baca menggunakan widget *Preprocess Text* di *Orange*. Selain itu, pada tahap krusial ini, sebagian data 999 tweet juga dilabeli secara manual oleh peneliti berdasarkan Panduan Anotasi untuk menciptakan sebuah dataset *ground truth* atau kunci jawaban yang valid.

3. Analisis Sentimen Otomatis dan Evaluasi Kinerja

Setelah data bersih, widget *Sentiment Analysis* digunakan untuk memberikan Label Prediksi secara otomatis pada setiap tweet. Selanjutnya, sebagai inti dari penelitian, kinerja alat otomatis ini dievaluasi dengan cara membandingkan Label Prediksi dari mesin dengan Label Manual (*ground truth*) untuk mengukur tingkat akurasinya.

4. Pengukuran Metrik dan Visualisasi Hasil

Proses evaluasi kinerja menghasilkan *Confusion Matrix* dari widget *Test & Score*, yang kemudian digunakan untuk menghitung metrik kuantitatif seperti akurasi, presisi, dan *recall*. Selain itu, hasil distribusi sentimen dari analisis otomatis juga divisualisasikan dalam bentuk grafik dan *Word Cloud* untuk interpretasi lebih mendalam.

