

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Tahapan Analisis Hasil Penelitian

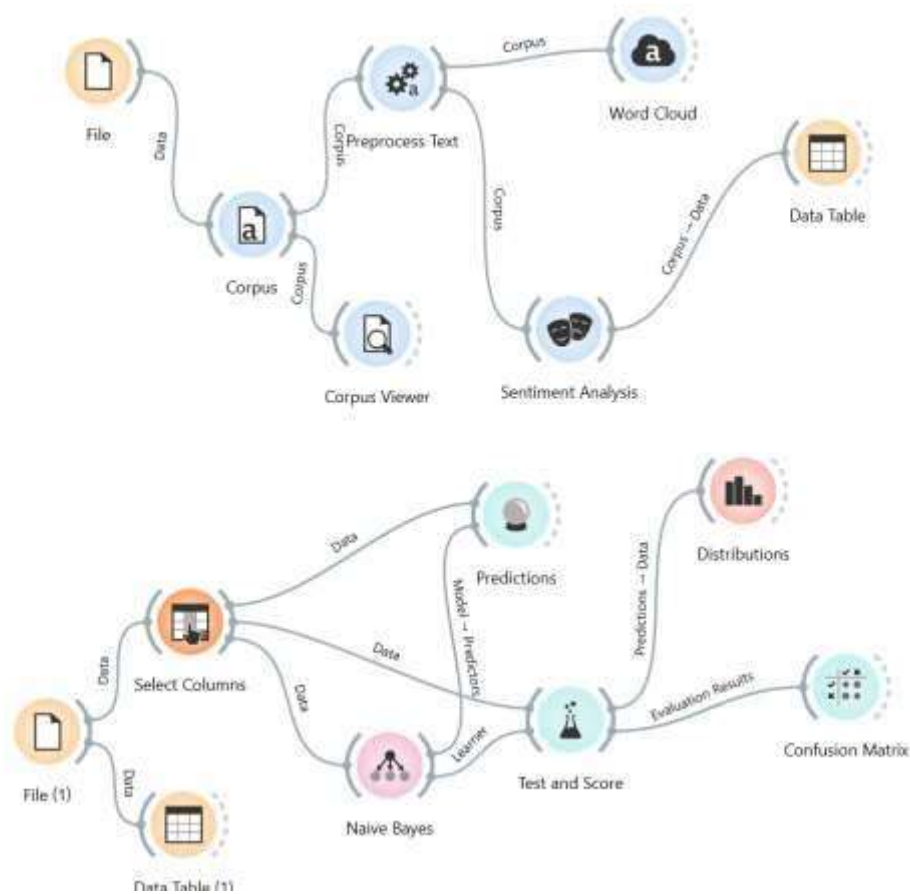
Penelitian ini menganalisis sebanyak 999 data *tweet* yang membahas Maarten Paes sebagai pemain diaspora. Data diperoleh melalui proses web scraping menggunakan *Python* dan *Selenium*. Selanjutnya, data tersebut dimasukkan ke dalam platform Orange Data Mining untuk dianalisis secara visual dan interaktif.

| A  | B   | C        |
|----|---|----------|
| 1  | username tweet  | waktu    |
| 2  | string string   | string   |
| 3  | meta meta include=True title=True   | meta     |
| 4  | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - FYI Martin Paes ini aslinya ga punya darah Indonesia sedikitpun, dan gak pernah jg menetap lama di Indon    | 11-09-21 |
| 5  | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama FYI Martin Paes ini aslinya ga punya darah Indonesia sedikitpun, dan gak pernah jg menetap lama di Indonesi   | 10-09-21 |
| 6  | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Geert Wilders ibunya turunan Sukabumi. Tp ogah lah punya politisi macem dia wkwk. XEC XEC @xaveriusen       | 11-09-21 |
| 7  | @TAU19FIQ - pernah denger podcast lupa channelnya... di sisi lain emang orang Belanda rapi bgt soal pengarsipan, pernah sampe pihak KAI untuk pembebasan lah            | 11-09-21 |
| 8  | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Termasuk arsip tanda bukti bahwa mrk emang udah bayar gaji pekerja jalan anyer panarukan, tapi 60 60 60     | 11-09-21 |
| 9  | ikan udarakan udara 🇮🇩 ikan udara 🇮🇩 @Sanjayawati - Dukacpil Belanda yg top 26 26 26 222 222 222 3.7K 3.7K 3.7K 250K 250K 250K  | 11-09-21 |
| 10 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Dan datanya gak pake bocor2 11 11 11 57 57 57 2.6K 2.6K 2.6K 238K 238K 238K                                 | 11-09-21 |
| 11 | AldhitamaAldhitama Ramadhan Aidhitama Ramadhan @AldhitamaR - Kenapa dirut BUMN ga nyari dari naturalisasi aja ya? Siapa tau bisa pada untung, 24 24 24 15 15 15 72      | 11-09-21 |
| 12 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Coba deh tanyain ke CEO2 starap asing, mereka untung gak? 4 4 4 3 3 3 336 336 153K 153K 153K                | 11-09-21 |
| 13 | InnovacorInnovacommunity Innovacommunity @innovacommunity - Jangan2 Mbappe neneknya dulu lahir di Indonesia jg?ikut arturnya agresi militer di Jogja. 19 19 19 26 26 26 | 11-09-21 |
| 14 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Karena ini sih 4 4 4 14 14 14 476 476 476 174K 174K 174K  | 11-09-21 |
| 15 | Suaminya Suaminya Brigitta Suaminya Brigitta @yokamudi - Ga masalah bngt sih, brahim aja yg lahir dan besar di Spanyol bisa wakilin Maroko 14 14 14 3 3 3 92 92 107K 1  | 11-09-21 |
| 16 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Bkn ttg bisa apie engga (krn pasti bs dan ada aturannya). Cuma buat colist info soal pemain bola Belanda    | 11-09-21 |
| 17 | Don MarioDon Mario Don Mario @mariomboys - Anak FM cuma tersenyum melihat ini 4 4 4 2 2 2 37 37 37 72K 72K 72K  | 11-09-21 |
| 18 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Aduh gua anak CM lagi 3 3 3 26 26 26 63K 63K 63K  | 11-09-21 |
| 19 | Stevie Stevie Stevie @stststevie - Charlie van houten jg naturalisasi kan 2 2 2 1 1 1 34 34 34 9.7K 9.7K 9.7K   | 11-09-21 |
| 20 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Damn knp gua bs sampe lupain dia ya 17 17 17 8.7K 8.7K 8.7K   | 11-09-21 |
| 21 | za riza za riza @dprfrzni - masa iya nakata ga ada darah Indonesiannya? jepang jajah luamayan juga 4 4 4 19 19 19 44K 44K 44K   | 11-09-21 |
| 22 | Guardian Guardian Of Bekasi Guardian Of Bekasi @rgoestama - Nama depan Nakata jg udah pake gelar Haj 7 7 7 1 1 1 52 52 52 41K 41K 41K                                   | 11-09-21 |
| 23 | Aulia S.E., M.Pak., Tbk. Aulia S.E., M.Pak., Tbk. @hanmula - Jangan2 sampe jaman daendels 3 3 3 1 1 1 216 216 216 105K 105K 105K  | 11-09-21 |
| 24 | XEC XEC XEC @xaveriuservan - Harus ditingkatkan ini, naturalisasi untuk level menteri, ketua KPK, kepolisian, dpr dll. 9 9 9 52 52 52 207 207 207 387K 387K 387K        | 11-09-21 |
| 25 | Muhamm Muhammad Bardan Salam Muhammad Bardan Salam @barjoou - Mas bro diluar tim resmi PSSI, komunitas FM @tdt_fm malah udah ngelist duluan bahkan samp                 | 11-09-21 |

Gambar 4.1 Data mentah hasil *scraping*

Proses analisis data ini dilakukan secara bertahap dan sistematis melalui berbagai tahapan penting. Dimulai dengan melakukan *preprocessing teks*, yaitu tahap pembersihan dan pengolahan data teks mentah agar siap untuk dianalisis.

Setelah itu, data divisualisasikan untuk mengidentifikasi kata-kata yang paling sering muncul menggunakan teknik *WordCloud*. Kemudian, data dianalisis secara sentimen dengan membagi hasil opini publik menjadi tiga kategori utama: positif, netral, dan negatif, yang dapat diamati melalui *widget Sentiment Analysis* di Orange. Langkah selanjutnya adalah menerapkan algoritma klasifikasi *Naïve Bayes* untuk mengelompokkan data berdasarkan sentimen tersebut, serta melakukan evaluasi akurasi model menggunakan *Test & Score* dan *Confusion Matrix*.



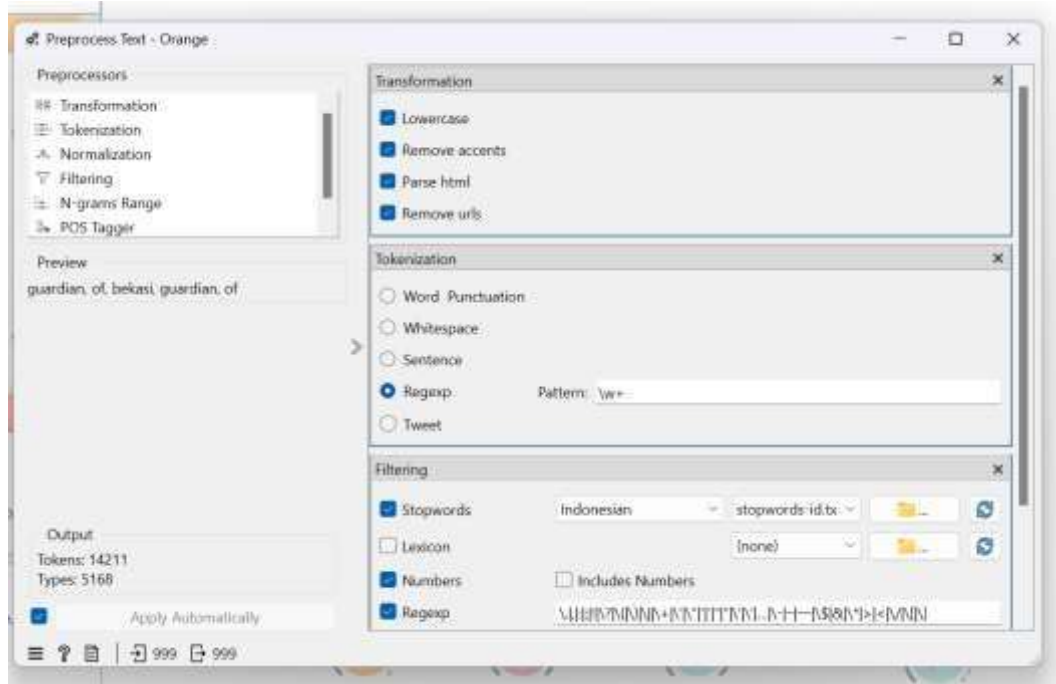
**Gambar 4.2** Alur proses pada *Orange Data Mining*

Pada tahap preprocessing, proses pembersihan data dilakukan secara menyeluruh dengan memanfaatkan widget Preprocess Text di aplikasi *Orange Data Mining*. Widget ini menyediakan berbagai fitur pengolahan teks yang sistematis, dimulai dengan konversi semua teks menjadi huruf kecil (*lowercase*), penghapusan aksent (*accent removal*), pembersihan elemen *HTML*, hingga penghilangan *URL* yang terdapat dalam isi *tweet*.

Langkah berikutnya adalah proses tokenisasi yang dilakukan menggunakan pola *regular expression (regex)* dengan format `\w+`, yang bertujuan untuk mengekstrak kata-kata utama dari setiap *tweet*. Setelah tokenisasi, dilakukan proses filtering dengan cara menghapus kata-kata umum (*stopwords*) yang dinilai tidak memiliki makna penting dalam konteks analisis sentimen. Daftar *stopwords* Bahasa Indonesia yang digunakan dalam penelitian ini diperoleh dari repository open-source *stopwords-iso* yang tersedia di GitHub [17].

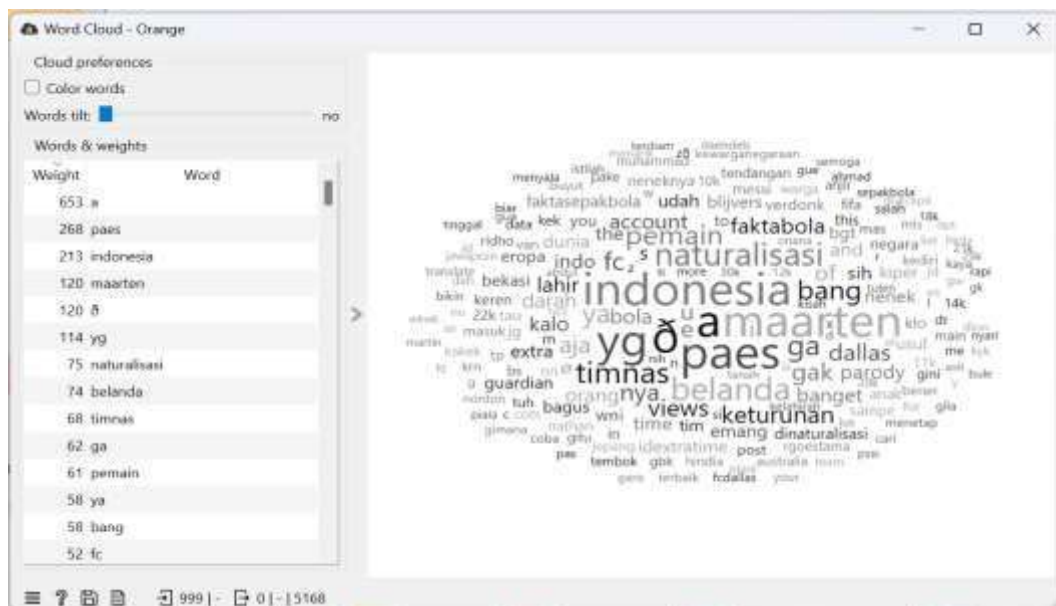
*tautan: <https://github.com/stopwords-iso/stopwords-id>.*

Selain penghapusan *stopwords*, dilakukan juga penyaringan terhadap simbol, angka, dan karakter non-alfabet lainnya agar hasil representasi data teks menjadi lebih bersih, konsisten, dan terstruktur. Keseluruhan tahapan ini bertujuan untuk menghasilkan data teks yang optimal dan siap digunakan dalam proses klasifikasi sentimen serta visualisasi data analisis.



**Gambar 4.3** Tampilan konfigurasi dan hasil *preprocessing data tweet* menggunakan *widget Preprocess Text* pada *Orange Data Mining*

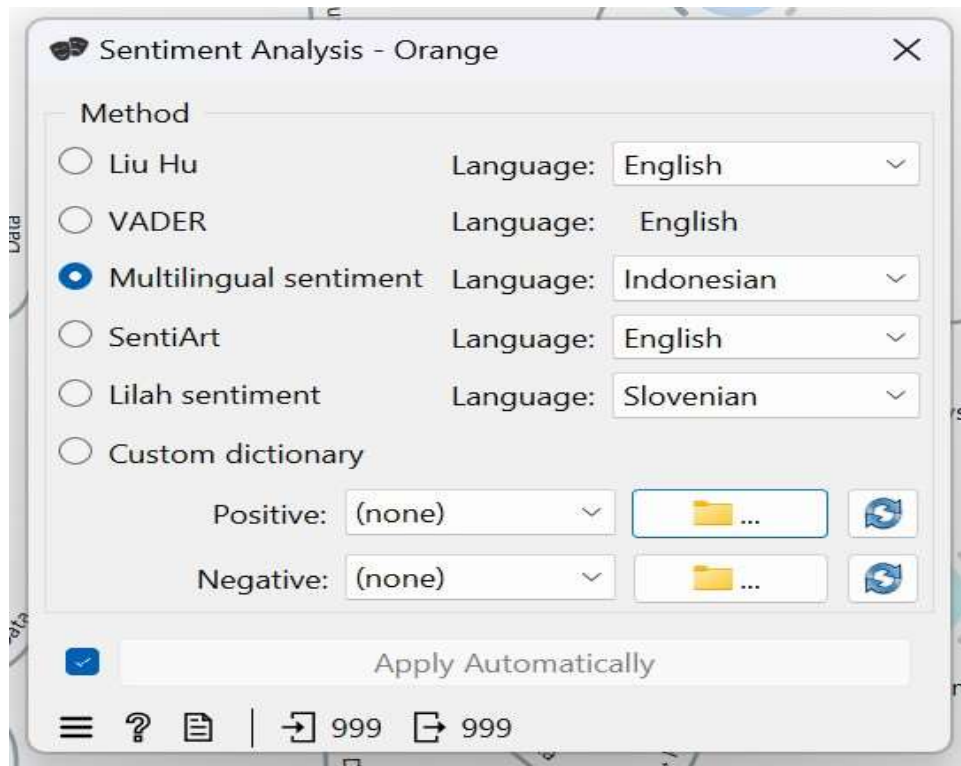
Selanjutnya, hasil *preprocessing* divisualisasikan dalam bentuk *WordCloud*, yang membantu mengidentifikasi kata-kata yang paling sering muncul dalam kumpulan *tweet*.



#### Gambar 4.4 Visualisasi WordCloud

Proses berikutnya adalah analisis sentimen yang dilakukan dengan memanfaatkan *widget Sentiment Analysis* pada *Orange Data Mining*. Metode yang digunakan adalah *Multilingual sentiment* dengan pengaturan bahasa Indonesia untuk menyesuaikan konteks lokal dari *tweet* yang dianalisis. Setiap *tweet* kemudian diklasifikasikan secara otomatis ke dalam tiga kategori utama, yaitu positif, netral, dan negatif. Hasil dari proses klasifikasi ini divisualisasikan pada *widget Data Table*, yang menampilkan detail data seperti username, isi *tweet*, waktu, serta skor sentimen numerik yang dihasilkan oleh sistem klasifikasi otomatis. Dengan adanya dua tampilan ini pengaturan metode dan hasil data peneliti dapat memastikan kesesuaian proses analisis serta mengevaluasi hasil klasifikasi secara lebih detail.

a.



| include title | username             | tweet                 | waktu             | sentiment |
|---------------|----------------------|-----------------------|-------------------|-----------|
| 69            | helmiscout           | helmiscout hel...     | 11-09-2024 11:... | 0         |
| 70            | A R A I              | A R A I   A R A I ... | 11-09-2024 11:... | 7.14286   |
| 71            | Rifqi Irsyadi        | Rifqi Irsyadi Rif...  | 11-09-2024 11:... | 0         |
| 72            | Mantan Sarjana       | Mantan Sarjana...     | 11-09-2024 11:... | -6.25     |
| 73            | Weirdo               | Weirdo Weirdo ...     | 11-09-2024 11:... | 9.09091   |
| 74            | pengedar obat ...    | pengedar obat ...     | 11-09-2024 11:... | 8.33333   |
| 75            | drgsj.eth            | drgsj.eth drgs...     | 11-09-2024 12:... | 0         |
| 76            | kimmy                | kimmy.kimmy ...       | 11-09-2024 13:... | 0         |
| 77            | Dion                 | Dion Dion @di...      | 11-09-2024 13:... | 0         |
| 78            | Jual Kuota 3         | Jual Kuota 3 Ju...    | 11-09-2024 21:... | 0         |
| 79            | dunia ini, berat ... | dunia ini, berat ...  | 11-09-2024 22:... | -2.94118  |
| 80            | @ @ @AllianR         | @ @ @AllianR...       | 12-09-2024 07:... | -12.5     |
| 81            | Islamic diaries      | Islamic diaries I...  | 12-09-2024 11:... | 0         |
| 82            | hmm                  | hmm hmm @p...         | 12-09-2024 11:... | 0         |
| 83            | Ihza                 | Ihza. Ihza. @_Ih...   | 11-09-2024 18:... | 0         |
| 84            | doddy ari susan...   | doddy ari susan...    | 11-09-2024 18:... | 6.66667   |
| 85            | HNDR                 | HNDR HNDR @...        | 11-09-2024 18:... | 0         |
| 86            | HyloSun              | HyloSun HyloS...      | 11-09-2024 18:... | -12.5     |
| 87            | Bayu 5 Minggu        | Bayu 5 Minggu ...     | 11-09-2024 18:... | 0         |
| 88            | James Alexander      | James Alexande...     | 11-09-2024 10:... | 18.1818   |
| on            | Mahesl Ala           | Mahesl Ala Mah...     | 11-09-2024 10:... | 4.54545   |

b. **Gambar 4.5 a) Tampilan konfigurasi metode analisis sentimen menggunakan widget *Sentiment Analysis* dengan bahasa Indonesia.**

**b) Tampilan hasil klasifikasi sentimen dalam *Data Table* yang menunjukkan skor dan kategori tiap *tweet*.**

Distribusi sentimen ditampilkan dalam bentuk tabel dan grafik sebagai berikut:

a. ***Tweet* Positif**

Sebanyak 247 *tweet* (24,7%) menunjukkan dukungan atau pandangan optimis terhadap Maarten Paes. Beberapa *tweet* ini memuji performa atau menyambut baik status naturalisasi sang pemain. Seluruh *tweet* ini dapat ditemukan di *Data Table* setelah proses sentimen analysis selesai, dengan label positive.

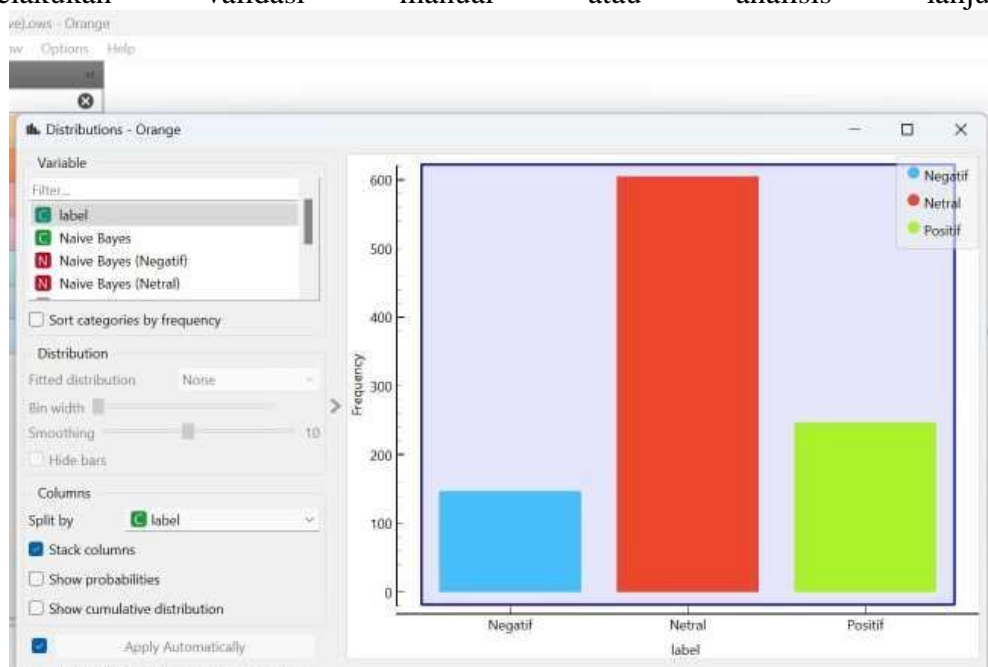
b. ***Tweet* Netral**

Sejumlah 605 *tweet* (60,6%) bersifat informatif atau tidak menunjukkan opini emosional yang jelas. Sentimen netral mendominasi dan menandakan bahwa sebagian besar netizen hanya menyampaikan informasi atau komentar ringan tanpa arah dukungan atau kritik yang eksplisit. *Tweet* ini juga ditandai sebagai netral di *Data Table*.

c. ***Tweet Negatif***

Sejumlah 147 *tweet* (14,7%) menunjukkan sentimen kritis terhadap proses naturalisasi maupun performa yang ditampilkan oleh Maarten Paes. Sebagian di antaranya mengandung nada skeptis dan memperbandingkan dirinya dengan pemain lokal. Seluruh *tweet* dalam kategori ini telah diklasifikasikan sebagai sentimen negatif oleh *Orange Data Mining*, dan hasilnya dapat diekspor untuk keperluan analisis lebih lanjut.

Dengan fitur *Data Table* di *Orange*, pengguna dapat memfilter ketiga kategori ini, mengekspor hasilnya, dan meninjau teks *tweet* satu per satu untuk melakukan validasi manual atau analisis lanjutan.



**Gambar 4.6 Diagram distribusi sentimen publik Pada Widget Orange Data Mining**

Selanjutnya, data yang telah dilabeli digunakan sebagai input dalam pembangunan model klasifikasi *Naïve Bayes*. Proses ini mencakup penggunaan widget "Select Columns", *Naïve Bayes*, *Test & Score*, dan *Confusion Matrix*.

**Tabel 4.1 *Confusion Matrix* hasil klasifikasi sentimen menggunakan *Naïve Bayes***

|                       | <b>Pred. Negatif</b> | <b>Pred. Netral</b> | <b>Pred. Positif</b> | <b>Total Aktual</b> |
|-----------------------|----------------------|---------------------|----------------------|---------------------|
| <b>Negatif</b>        | 146                  | 1                   | 0                    | 147                 |
| <b>Netral</b>         | 1                    | 604                 | 0                    | 605                 |
| <b>Positif</b>        | 0                    | 0                   | 247                  | 247                 |
| <b>Total Prediksi</b> | 147                  | 605                 | 247                  | 999                 |

Tabel 4.1 menyajikan hasil evaluasi performa model Naive Bayes dalam melakukan klasifikasi sentimen terhadap 999 data tweet terkait Maarten Paes. Berikut adalah penjelasan rinci dari masing-masing komponen dalam confusion matrix:

1. *True Negative* (TN) = 146: Merupakan jumlah *tweet* yang benar-benar bersentimen negatif dan berhasil diklasifikasikan dengan tepat oleh model sebagai negatif. Artinya, dari 147 *tweet* negatif, sebanyak 146 *tweet* teridentifikasi dengan benar sebagai negatif.

2. *False Positive* (FP) = 1: Menunjukkan adanya 1 *tweet* netral yang secara keliru diklasifikasikan oleh model sebagai negatif. Ini menyebabkan jumlah total prediksi negatif menjadi 147 (146 TN + 1 FP).
3. *False Negative* (FN) = 1: Menggambarkan adanya 1 *tweet* negatif yang salah diklasifikasikan oleh model sebagai netral. Ini berarti, dari 147 *tweet* sebenarnya negatif, hanya 146 yang terprediksi benar, sementara 1 sisanya menjadi FN.
4. *True Neutral* (T-Neutral) = 604: Yaitu jumlah *tweet* yang memang bersentimen netral dan diklasifikasikan dengan tepat sebagai netral oleh model. Dari total 605 *tweet* netral, sebanyak 604 berhasil diklasifikasikan dengan benar.
5. *False Neutral* (F-Neutral) = 1: Ada 1 *tweet* negatif yang salah diklasifikasikan sebagai netral (sudah disebut di FN tadi), sehingga total prediksi netral menjadi 605 (1 FN + 604 benar).
6. *True Positive* (TP) = 247: Semua *tweet* yang memiliki sentimen positif sebanyak 247 berhasil diklasifikasikan dengan benar oleh model tanpa ada kesalahan klasifikasi. Tidak terdapat *False Positive* maupun *False Negative* untuk kategori positif ini.
7. Jumlah Total Data = 999: Merupakan total keseluruhan *tweet* yang dianalisis dalam penelitian ini, yang terdiri dari 147 *tweet* negatif, 605 *tweet* netral, dan 247 *tweet* positif.

Berdasarkan tabel ini, dapat dilihat bahwa model Naive Bayes memiliki tingkat akurasi yang sangat tinggi dengan jumlah prediksi yang hampir seluruhnya tepat pada masing-masing kategori sentimen.

### **Perhitungan Akurasi, Precision, Recall, dan F1-Score**

Berdasarkan data confusion matrix pada Tabel 4.1, berikut adalah hasil perhitungan metrik evaluasi model berikut ini.

**1. Akurasi (Accuracy):** Akurasi dihitung dengan rumus:

$$Accuracy = \frac{TP + TN + T-Neutral}{Total} \times 100\%$$
$$Accuracy = \frac{247 + 146 + 604}{999} \times 100\% = \frac{997}{999} \times 100\% \approx 99,8\%$$

**2. Precision untuk Sentimen Negatif:** Precision dihitung dengan rumus:

$$Precision_{Negatif} = \frac{TN}{TN + FP} = \frac{146}{146 + 1} \approx 99,3\%$$

**3. Recall untuk Sentimen Negatif,** Recall dihitung dengan rumus:

$$Recall_{Negatif} = \frac{TN}{TN + FN} = \frac{146}{146 + 1} \approx 99,3\%$$

**4. F1-Score untuk Sentimen Negatif:**

$$F1_{Negatif} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \approx 99,3\%$$

**5. Precision untuk Sentimen Positif,** Karena tidak ada FP atau FN pada kategori positif:

$$Precision_{Positif} = \frac{TP}{TP} = 100\%$$

## 6. Recall untuk Sentimen Positif:

$$Recall_{Positif} = \frac{TP}{TP} = 100\%$$

## 7. F1-Score untuk Sentimen Positif:

$$F1_{Positif} = 100\%$$

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Naive Bayes* memiliki performa yang sangat baik dalam mengklasifikasikan sentimen *tweet* terkait Maarten Paes, dengan akurasi yang mencapai 99,8% serta nilai precision, recall, dan F1-score yang sangat tinggi pada semua kategori sentiment.

## 4.2 Pembahasan

### 4.2.1 Proses Pengumpulan Data

Data penelitian ini dikumpulkan melalui proses *scraping Twitter* dengan bantuan *Python* dan *Selenium* untuk mengatasi keterbatasan akses API. Fokus pengambilan data diarahkan pada *tweet* yang membahas Maarten Paes, khususnya terkait naturalisasi dan performanya sebagai pemain *diaspora*. Penggunaan kata kunci relevan membantu memastikan bahwa data yang diperoleh sesuai dengan topik penelitian dan merepresentasikan opini netizen Indonesia. Dokumentasi lengkap mengenai proses *scraping* menggunakan *Python* dan *Selenium* dapat dilihat pada **Lampiran 1** yang terletak di bagian akhir dokumen ini.

#### 4.2.2 Pelabelan Otomatis Menggunakan Rumus Excel

Sebelum data dianalisis lebih lanjut di Orange Data Mining, dilakukan proses pelabelan sentimen secara manual berbantuan otomatisasi sederhana di Microsoft Excel. Setiap tweet diberikan label sentimen positif, netral, atau negatif berdasarkan nilai skor sentimen yang diperoleh dari preprocessing sebelumnya. Penentuan label dilakukan menggunakan rumus logika IF untuk menyesuaikan nilai dengan klasifikasi tertentu.

Skema penetapan label menggunakan logika sebagai berikut:

- A. Jika nilai skor lebih dari 0 → positif
- B. Jika nilai sama dengan 0 → netral
- C. Jika nilai kurang dari 0 → negatif

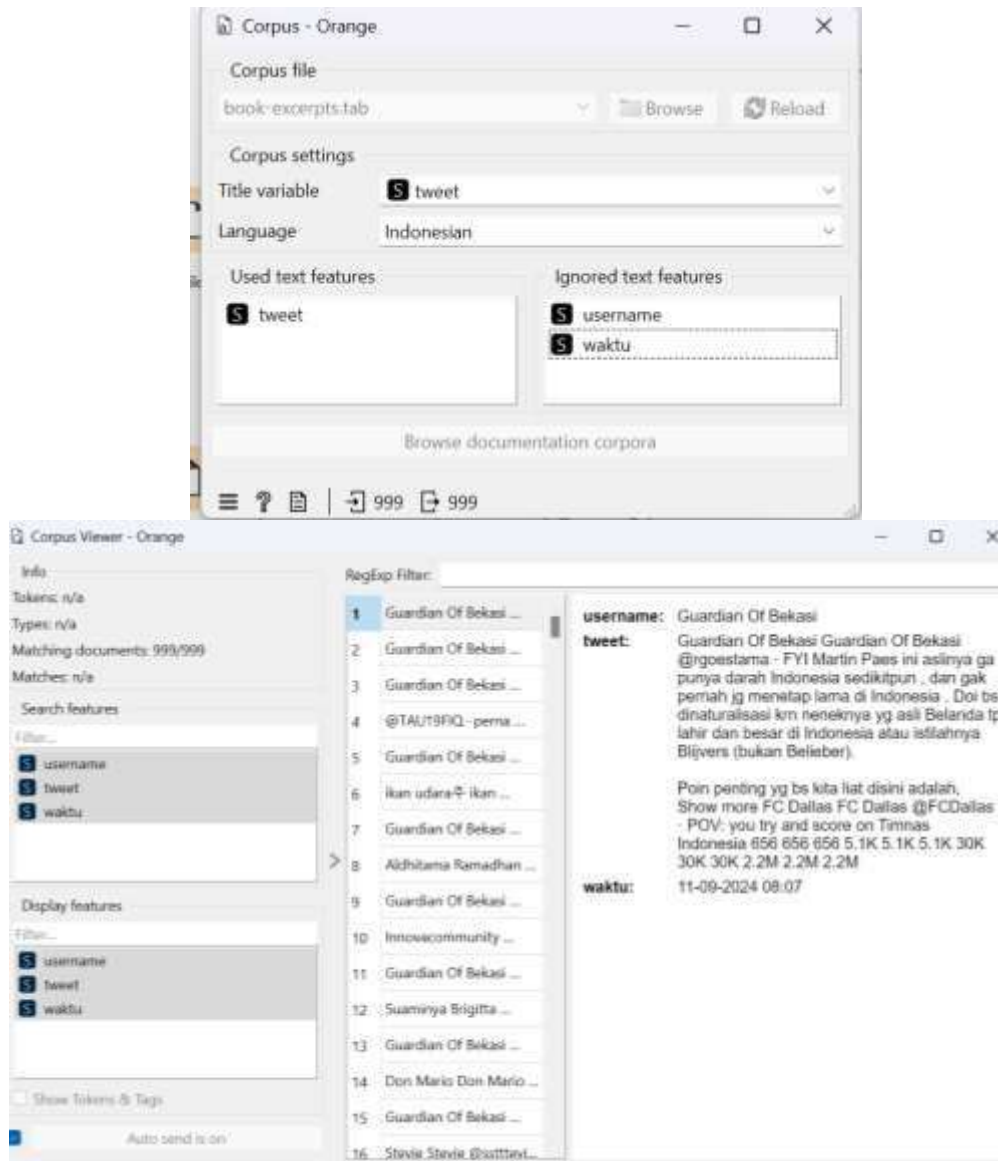
Adapun rumus yang digunakan dalam Excel adalah sebagai berikut:

**=IF(VALUE(D4)>0, "Positif", IF(VALUE(D4)<0, "Negatif", "Netral"))**

Rumus ini memastikan setiap baris data memperoleh label yang konsisten dan sesuai dengan polaritas sentimen yang tercermin dari nilai numeriknya. Proses ini mempermudah pemetaan data secara sistematis sebelum dimasukkan ke dalam sistem klasifikasi machine learning. Dengan demikian, hasil labeling ini menjadi acuan awal sebelum dilakukan analisis sentimen otomatis menggunakan algoritma Naive Bayes di tahap selanjutnya.

#### 4.2.3 Proses Pemasukan Data

Data hasil scraping disimpan dalam format Excel dan diunggah ke Orange Data Mining. Data diproses melalui widget "Corpus" untuk diubah menjadi format teks.



**Gambar 4.7** Tampilan pengaturan *widget Corpus* untuk memuat file Excel *tweet* hasil *scraping*, dengan kolom teks dipilih sebagai sumber utama untuk **analisis sentimen**

#### 4.2.4 Proses Pembersihan Data dan *Preprocessing*

Proses preprocessing dalam penelitian ini dilakukan menggunakan widget Preprocess Text di Orange Data Mining. Widget ini secara otomatis menjalankan

beberapa tahapan utama dalam pengolahan teks, yaitu tokenisasi, stopwords removal, normalisasi, dan stemming, dengan fungsi masing-masing sebagai berikut:

a. Tokenisasi

Memecah teks menjadi kata-kata individual.

b. Stopword removal

menghapus kata yang tidak memiliki makna penting.

c. Normalisasi

mengubah kata tidak baku menjadi bentuk baku.

d. Stemming

mengubah kata ke bentuk dasarnya.

#### **4.2.5 Pelabelan Data (*Labeling*)**

Pelabelan sentimen dilakukan menggunakan *widget Sentiment Analysis* pada *Orange Data Mining*. Setiap tweet secara otomatis diklasifikasikan ke dalam salah satu dari tiga kategori sentimen: positif, netral, atau negatif, berdasarkan isi teks. Hasil labeling ini kemudian digunakan sebagai variabel target dalam proses klasifikasi untuk membangun model *Naive Bayes*.

#### **4.2.6 WordCloud**

*Widget WordCloud* dimanfaatkan sebagai media visualisasi untuk menampilkan kumpulan kata yang paling sering muncul dalam dataset yang diteliti. Dalam tampilan *WordCloud*, ukuran setiap kata mencerminkan tingkat frekuensi kemunculannya dalam data semakin sering sebuah kata muncul,



The screenshot shows the 'Predictions - Orange' window. At the top, it displays 'Show probabilities for: Classes in data' and 'Show classification errors' (checked). Below this is a table with columns: 'Naive Bayes', 'error', 'label', 'username', 'tweet', and 'waktu'. Three rows are visible, all showing a prediction of 'Netral' with an error rate of 0.005. Below the table, there is a section for 'Show performance scores' with 'Target class: (Average over classes)'. A table shows performance metrics for the Naive Bayes model: AUC (0.998), CA (0.998), F1 (0.998), Prec (0.998), Recall (0.998), and MCC (0.996). At the bottom, there are icons and a status bar showing '999 | 999 | 1x999'.

| Naive Bayes                 | error | label  | username      | tweet               | waktu      |
|-----------------------------|-------|--------|---------------|---------------------|------------|
| 0.00 : 1.00 : 0.00 → Netral | 0.005 | Netral | R             | R R @Ramadha...     | 11-09-2024 |
| 0.00 : 1.00 : 0.00 → Netral | 0.005 | Netral | piculararimba | piculararimba pi... | 11-09-2024 |
| 0.00 : 1.00 : 0.00 → Netral | 0.005 | Netral | Netizen       | Netizen Netize...   | 11-09-2024 |

| Model       | AUC   | CA    | F1    | Prec  | Recall | MCC   |
|-------------|-------|-------|-------|-------|--------|-------|
| Naive Bayes | 0.998 | 0.998 | 0.998 | 0.998 | 0.998  | 0.996 |

**Gambar 4.9 Diagram evaluasi model Naïve Bayes**

Berdasarkan hasil klasifikasi, distribusi sentimen dari total 999 *tweet* adalah sebagai berikut: 247 *tweet* (24,7%) termasuk dalam kategori positif, 605 *tweet* (60,6%) termasuk netral, dan 147 *tweet* (14,7%) termasuk negatif.

#### 4.2.8 Analisis Kesalahan Klasifikasi

Dari *confusion matrix*, terdapat dua kasus kesalahan klasifikasi satu *tweet* netral diprediksi sebagai negatif dan satu *tweet* negatif diprediksi sebagai netral. Kesalahan ini kemungkinan besar disebabkan oleh gaya bahasa ambigu, ironi, atau konteks yang tidak eksplisit.

#### 4.2.9 Kaitan Hasil dengan Hubungan Atlet dan Publik

Mayoritas *tweet* netral mencerminkan bahwa netizen belum memiliki keterlibatan emosional yang kuat terhadap Maarten Paes. Sentimen positif menunjukkan adanya dukungan terhadap program naturalisasi, sedangkan sentimen negatif cenderung muncul dalam konteks kontroversi atau perbandingan dengan pemain lokal. Temuan ini menunjukkan bahwa opini publik terhadap atlet

diaspora bersifat dinamis dan dipengaruhi oleh persepsi, performa, serta konteks kebangsaan.

#### **4.2.10 Implikasi Penelitian**

Penelitian ini memiliki beberapa implikasi strategis. Bagi pihak universitas, penelitian ini menunjukkan potensi mahasiswa dalam memanfaatkan teknologi analisis data seperti *Orange Data Mining* untuk isu-isu sosial, olahraga, dan komunikasi digital. Ini mendukung pengembangan literasi data dan keterampilan digital mahasiswa sebagai bagian dari kurikulum Merdeka Belajar.

Bagi federasi seperti PSSI, hasil penelitian ini dapat dijadikan pertimbangan dalam memahami dinamika opini publik terhadap pemain diaspora. Dengan mengetahui distribusi sentimen publik, federasi dapat mengelola komunikasi publik secara lebih bijak, meningkatkan kepercayaan masyarakat, serta mengantisipasi reaksi terhadap kebijakan naturalisasi pemain.

Dengan demikian, penelitian ini tidak hanya relevan dalam konteks akademis, tetapi juga memiliki nilai praktis dalam pengambilan kebijakan komunikasi dan keolahragaan nasional.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Penelitian ini menganalisis persepsi publik terhadap Maarten Paes sebagai pemain diaspora melalui 999 tweet yang diolah menggunakan platform Orange Data Mining. Hasil klasifikasi sentimen menunjukkan bahwa mayoritas tanggapan dari netizen bersifat netral (60,6%), diikuti oleh positif (24,7%), dan negatif (14,7%).

Dominasi sentimen netral mencerminkan sikap publik yang cenderung berhati-hati dan mempertimbangkan konteks sebelum menyampaikan opini bernuansa emosional. Sementara itu, sentimen positif menandakan adanya dukungan terhadap kehadiran Maarten Paes di skuad Timnas Indonesia. Sebaliknya, adanya sentimen negatif mengindikasikan bahwa sebagian masyarakat masih menunjukkan resistensi terhadap kebijakan naturalisasi pemain asing.

Dari sisi metodologis, penelitian ini berhasil menunjukkan bahwa Orange Data Mining merupakan alat yang efektif dalam menganalisis data teks secara visual dan terstruktur. Seluruh proses analisis, mulai dari pra-pemrosesan hingga evaluasi model, dapat dijalankan dengan baik menggunakan alur kerja yang tersedia. Penggunaan algoritma Naïve Bayes dalam klasifikasi sentimen juga terbukti mampu menghasilkan tingkat akurasi tinggi, yakni sebesar 99,8%, yang menunjukkan keandalan pendekatan yang digunakan.

Secara keseluruhan, hasil penelitian ini memperkuat argumen bahwa opini warganet dapat memengaruhi citra dan penerimaan publik terhadap pemain

diaspora. Dinamika ini menjadi refleksi dari hubungan antara atlet dan masyarakat yang terbangun melalui interaksi di media sosial.

## **5.2 Saran**

Penelitian ini masih memiliki keterbatasan, terutama dalam hal ruang lingkup data yang hanya bersumber dari satu platform media sosial, yaitu Twitter, serta penggunaan satu metode klasifikasi, yakni algoritma Naïve Bayes. Oleh karena itu, penulis menyarankan agar penelitian selanjutnya dapat memperluas cakupan data dari berbagai platform media sosial seperti Instagram, TikTok, atau YouTube yang juga menjadi wadah diskusi publik, serta menguji efektivitas model klasifikasi lain seperti SVM, Decision Tree, atau bahkan pendekatan berbasis deep learning untuk perbandingan dan validasi hasil.

Secara substantif, penulis mendorong agar analisis terhadap fenomena atlet diaspora seperti Maarten Paes tidak hanya dilihat dari segi statistik sentimen, tetapi juga konteks sosial dan budaya yang memengaruhinya. Kajian ini dapat dikembangkan lebih lanjut untuk mengeksplorasi representasi digital atlet diaspora serta keterkaitannya dengan persepsi nasionalisme dan identitas publik.

Dalam konteks penelitian ini, pendekatan melalui Orange Data Mining terbukti efektif dalam memvisualisasikan dan memproses data teks secara praktis. Namun, sebelum memilih pengolahan melalui alat di atas penulis menghadapi kendala teknis berupa gangguan koneksi (network error) pada perangkat komputer yang digunakan, yang menyebabkan proses eksekusi kode di Google Colab menjadi terhambat dan tidak dapat berjalan dengan lancar. Permasalahan ini berdampak pada efisiensi dan kelancaran selama proses analisis data berlangsung.

Melalui penelitian ini, diharapkan dapat membuka cakrawala baru bagi mahasiswa dan akademisi untuk terus memanfaatkan teknologi analisis data, serta menjadi bahan pertimbangan bagi institusi olahraga seperti PSSI dalam memahami dan merespons opini publik terkait kebijakan pemain diaspora. Penelitian ini juga menunjukkan bahwa mahasiswa memiliki kontribusi penting dalam menjembatani data digital dan isu-isu nasional secara akademik dan aplikatif.

## DAFTAR PUSTAKA

- [1] M. K. Sinaga and G. Y. Lubis, -Analisis Pengaruh Media Sosial Terhadap Persepsi Masyarakat Pada Olahraga Sepakbola,|| *J. Ilm. Kaji. Multidisipliner*, vol. 8, no. 6, pp. 834–837, 2024.
- [2] M. Zulhidayat, A. Suganda, and I. B. Razif, -Political Law of The Government in A Special Naturalization of Indonesian Football Players Based on Welfare State Theory,|| *Eduvest - J. Univers. Stud.*, vol. 2, no. 4, pp. 622–629, 2022, doi: 10.59188/eduvest.v2i4.419.
- [3] Rani, Tamara Cindy Samsita, dan Eka Sahputra. 2024. *Penerapan Machine Learning terhadap Analisis Sentimen Masyarakat*. E-book. Jakarta: Penerbit NEM.Diakses dari [https://books.google.com/books/about/Penerapan Machine Learning terhadap Anal.html?id=KRALEQAAQBAJ](https://books.google.com/books/about/Penerapan_Machine_Learning_terhadap_Anal.html?id=KRALEQAAQBAJ)
- [4] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, -Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal),|| *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] E. S. S. P. SARLI, NURHADI, -ANALISIS PENGGUNAAN GAYA BAHASA SARKASME NETIZEN DI MEDIA SOSIAL TIKTOK,|| vol. 3, no. 1, pp. 84–92, 2023.
- [6] U. Hasanah, A. R. Rahim, and A. S. Syamsuri, -Analisis Penggunaan Gaya Bahasa Sarkasme Netizen di Media Sosial Instagram Pendahuluan,|| vol. 7, no. 2, pp. 411–423, 2021.
- [7] A. Prasetya *et al.*, -POLA PERILAKU BERMEDIA SOSIAL NETIZEN INDONESIA,|| pp. 1–12, 2022.
- [8] N. Bahar and R. Latief, -CYBERACTIVISM SEBAGAI BUDAYA POPULER : ANALISIS WACANA HASHTAG #PERCUMALAPORPOLISI,|| vol. 2021, pp. 39–65.
- [9] D. S. Puspitarini and R. Nuraeni, -PEMANFAATAN MEDIA SOSIAL SEBAGAI MEDIA PROMOSI ( Studi Deskriptif pada Happy Go Lucky House ),|| vol. 3, 2019.
- [10] M. E. Hambrick, J. M. Simmons, G. P. Greenhalgh, and T. C. Greenwell, -Understanding Professional Athletes’ Use of Twitter: A Content Analysis of Athlete Tweets,|| *Int. J. Sport Commun.*, vol. 3, no. 4, pp. 454–471, 2016, doi: 10.1123/ijsc.3.4.454.
- [11] A. Pak and P. Paroubek, -Twitter as a corpus for sentiment analysis and opinion mining,|| *Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010*, no.

December, pp. 1320–1326, 2010, doi: 10.17148/ijarccce.2016.51274.

- [12] F. N. Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, –Jurnal KomtekInfo Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan,|| vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [13] P. A. Permatasari, L. Linawati, and L. Jasa, –Survei Tentang Analisis Sentimen Pada Media Sosial,|| Maj. Ilm. Teknol. Elektro, vol. 20, no. 2, p. 177, 2021, doi: 10.24843/mite.2021.v20i02.p01. [14]B. Chalifa Chazar1 and us E. W. Program, –INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020,|| vol. 12, no. 1, pp. 67–80, 2020.
- [15] A. Géron (2019), *Hands-On Machine Learning with Scikit-Learn , Keras , and. O’Reilly Media.*
- [16] Y. Nurdiansyah, S. Bukhori, and R. Hidayat, –Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method,|| *J. Phys. Conf. Ser.*, vol. 1008, no. 1, 2018, doi: 10.1088/1742-6596/1008/1/012011.
- [17] genediazjr, "*stopwords-iso: Indonesian stopwords list*", GitHub, 2016. [Online]. Available: <https://github.com/stopwords-iso/stopwords-id>

## LAMPIRAN

### Lampiran 1

```
#main.py
from selenium import webdriver

from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.keys import Keys

import time

import random

import pandas as pd

from datetime import datetime, timezone, timedelta

# =====
# PARAMETER YANG BISA DIUBAH
# =====

TWITTER_EMAIL = "ericksyaputra337@gmail.com"
TWITTER_USERNAME = "ESyaputra55947"
TWITTER_PASSWORD = "8z7CiR88gXVWPa4"

keyword = "Maarten Paes"

jumlah_tweet = 999

tanggal_mulai = "2024-01-23"
tanggal_akhir = "2025-05-01"

# =====

# Setup Chrome

options = Options()

options.add_argument("--start-maximized")
```

```

driver = webdriver.Chrome(options=options)

# 1. Buka login Twitter
driver.get("https://twitter.com/login")
time.sleep(5)

# 2. Masukkan email
email_input = driver.find_element(By.NAME, "text")
email_input.send_keys(TWITTER_EMAIL)
email_input.send_keys(Keys.RETURN)
time.sleep(3)

# 3. Masukkan username
username_input = driver.find_element(By.NAME, "text")
username_input.send_keys(TWITTER_USERNAME)
username_input.send_keys(Keys.RETURN)
time.sleep(3)

# 4. Masukkan password
password_input = driver.find_element(By.NAME, "password")
password_input.send_keys(TWITTER_PASSWORD)
password_input.send_keys(Keys.RETURN)
time.sleep(5)

# 5. Masuk ke pencarian tweet berbahasa Indonesia
query = f"{keyword} since:{tanggal_mulai} until:{tanggal_akhir} lang:id"
search_url = f"https://twitter.com/search?q={query}&src=typed_query&f=live"
driver.get(search_url)

```