

BAB III HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Pengambilan Data

Data ulasan aplikasi SIREKAP 2024 diambil menggunakan library *Google-play-scrapper* yang diambil pada tanggal 6 februari 2024. Berikut hasil data ulasan yang terambil dapat dilihat pada gambar 3.1.

| | userName | score | at | content | thumbsUpCount |
|------|-----------------|-------|---------------------|---|---------------|
| 6671 | Nency Miranda | 5 | 2024-02-05 14:07:12 | Mantap | 0 |
| 5783 | ichal sabiel | 1 | 2024-02-05 14:06:17 | Sudah tau pemakainya masyarakat biasa Malah sp... | 0 |
| 7909 | Evandra Aditya | 1 | 2024-02-05 14:06:16 | Apk nya nggk bisa buat log in | 0 |
| 6307 | Ganesh insann | 1 | 2024-02-05 14:05:38 | Susah masuk dih | 0 |
| 8008 | Yohana Frediana | 1 | 2024-02-05 14:04:32 | Tidak bisa masuk inisiliasi | 0 |

Gambar 3.1 Hasil Pengambilan Data Ulasan SIREKAP 2024

Hasil data yang telah terambil diekspor menggunakan format *Comma Separated Values (CSV)* dengan nama “scrapped_data.csv” untuk dianalisis lebih lanjut. Selanjutnya, data ulasan diinput menggunakan format CSV ke *Python* dan hasilnya ditunjukkan pada Gambar 3.2.

```

      Nama Rating Waktu \
0      Nency Miranda 5 05/02/2024 14:07
1      ichal sabiel 1 05/02/2024 14:06
2      Evandra Aditya 1 05/02/2024 14:06
3      Ganesh insann 1 05/02/2024 14:05
4      Yohana Frediana 1 05/02/2024 14:04
...
8353      Tn A Muntaha 5 23/01/2024 09:05
8354      Adhi Nugroho 1 23/01/2024 08:51
8355      TATANG RUSDIANA 5 23/01/2024 07:03
8356      vie 5 23/01/2024 06:50
8357      Ahmad ferdiansyah 5 23/01/2024 04:26

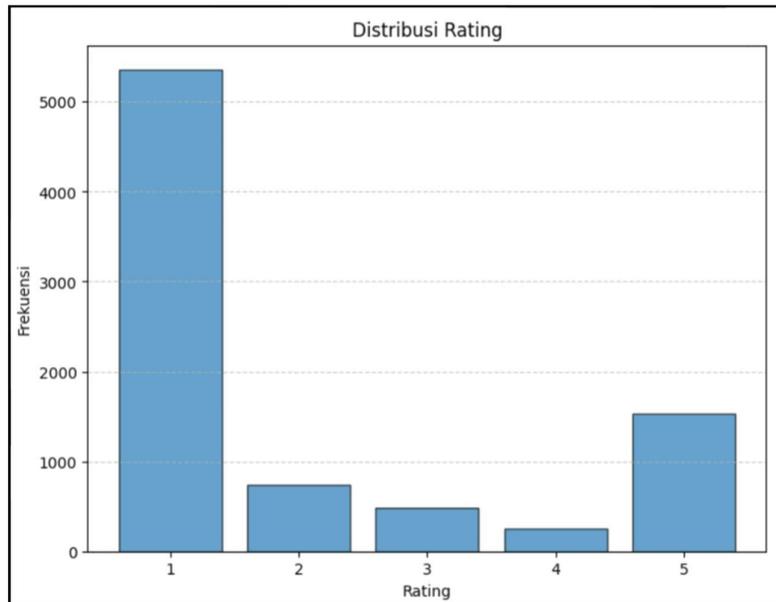
      Komentar Like
0      Mantap 0.0
1      Sudah tau pemakainya masyarakat biasa Malah sp... 0.0
2      Apk nya nggk bisa buat log in 0.0
3      Susah masuk dih 0.0
4      Tidak bisa masuk inisiliasi 0.0
...
8353      Semoga Tambah Baik,Mudah Dalam Input Dan Share... 21.0
8354      Cara masuk nya bagaimana? Pakai email kok gag ... 448.0
8355      Mbuh 6.0
8356      Semoga tidak seperti sirekap 2020 yang saat di... 42.0
8357      Semoga dengan adanya aplikasi ini pemilu akan ... 186.0

[8358 rows x 5 columns]
```

Gambar 3.2 Hasil Input Data

Dapat dilihat pada gambar diatas, data yang digunakan terdiri dari 8358 baris dan 5 kolom. Yang berisikan mulai dari Nama, *Rating*, Waktu, Komentar, dan Like. Setelah data diinput, tahap pertama dalam menganalisis data yang dilakukan adalah Distribusi *Rating*. Distribusi *rating* dilakukan untuk

melihat berapa banyak ulasan yang terdapat pada setiap *rating*. Hasil distribusi *rating* dapat dilihat pada gambar 3.3.



Gambar 3.3 Distribusi *Rating*

Terlihat pada gambar 3.3, *rating* 1 memiliki frekuensi kemunculan yang sangat tinggi dibandingkan *rating* lainnya. Selain itu, frekuensi kemunculan pada *rating* 2, 3, dan 4 yang berada diantara *rating* 1 dan *rating* 5 sangat rendah. Ini menunjukkan bahwa data ulasan memiliki data yang tidak seimbang. Jumlah kemunculan data pada setiap *rating* dapat dilihat pada tabel 3.1.

Tabel 3.1 Jumlah Kemunculan Data Pada Setiap *Rating*

| <i>Rating</i> | Jumlah kemunculan data |
|---------------|------------------------|
| 1 | 5361 |
| 2 | 736 |
| 3 | 478 |
| 4 | 255 |
| 5 | 1528 |

Tabel 3.1 menunjukkan jumlah data pada setiap *rating*. *Rating* 1 berisi 5361 data, *Rating* 2 berisi 736 data, *Rating* 3 berisi 478 data, *Rating* 4 berisi 255 data, dan *Rating* 5 berisi 1528 data. Selanjutnya digunakan *Wordcloud* untuk mengvisualisasikan data ulasan pada setiap *rating* dalam bentuk gambar dan melihat frekuensi kata yang sering muncul.

memproses teks tersebut. Untuk itu perlu dilakukan tahap *Pra-proses* agar data ulasan ini menjadi lebih efisien untuk digunakan pada model.

3.1.2 *Pra-proses* Data

Setiap data diberikan id untuk mempermudah pelacakan dan identifikasi setiap baris secara individual. Hasilnya dapat dilihat pada gambar 3.6.

| | ID | Nama | Rating | Waktu | \ |
|------|-------|---|--------|------------------|-----------------|
| 0 | d0001 | Nency Miranda | 5 | 05/02/2024 14:07 | |
| 1 | d0002 | ichal sabiel | 1 | 05/02/2024 14:06 | |
| 2 | d0003 | Evandra Aditya | 1 | 05/02/2024 14:06 | |
| 3 | d0004 | Ganesh insann | 1 | 05/02/2024 14:05 | |
| 4 | d0005 | Yohana Frediana | 1 | 05/02/2024 14:04 | |
| ... | ... | ... | ... | ... | ... |
| 8353 | d8354 | Tn A Muntaha | 5 | 23/01/2024 09:05 | |
| 8354 | d8355 | Adhi Nugroho | 1 | 23/01/2024 08:51 | |
| 8355 | d8356 | TATANG RUSDIANA | 5 | 23/01/2024 07:03 | |
| 8356 | d8357 | vie | 5 | 23/01/2024 06:50 | |
| 8357 | d8358 | Ahmad ferdiansyah | 5 | 23/01/2024 04:26 | |
| | | | | | Komentar |
| 0 | | | | | Mantap |
| 1 | | Sudah tau pemakainya masyarakat biasa | | | Malah sp... |
| 2 | | Apk nya nggk bisa buat log in | | | |
| 3 | | | | | Susah masuk dih |
| 4 | | Tidak bisa masuk inisiliasi | | | |
| ... | | | | | ... |
| 8353 | | Semoga Tambah Baik,Mudah Dalam Input Dan Share... | | | 21.0 |
| 8354 | | Cara masuk nya bagaimana? Pakai email kok gag ... | | | 448.0 |
| 8355 | | | | | Mbuh 6.0 |
| 8356 | | Semoga tidak seperti sirekap 2020 yang saat di... | | | 42.0 |
| 8357 | | Semoga dengan adanya aplikasi ini pemilu akan ... | | | 186.0 |

[8358 rows x 6 columns]

Gambar 3.6 Hasil Input Id

Terlihat pada gambar 3.6 telah ditambahkan kolom id kumpulan data. Kolom id terdapat pada kolom ke 1 dengan menambahkan id “d0001” dan seterusnya yang kemudian ditambahkan pada data ulasan masing-masing. Tahap berikutnya hanya menggunakan kolom Komentar, karena hanya ulasan pengguna dalam kolom Komentar yang diperlukan pada proses klasifikasi teks.

1. *Lower Case*

Data ulasan diubah dari kata yang mengandung huruf kapital menjadi huruf kecil. Gambar 3.7 merupakan perbandingan antara sebelum dilakukan *Lower Case* dan setelah dilakukan *Lower Case*.

| | Komentar | \ | komentar_lowercase |
|------|---|---|---|
| 0 | Mantap | | mantap |
| 1 | Sudah tau pemakainya masyarakat biasa | | sudah tau pemakainya masyarakat biasa |
| 2 | Apk nya nggk bisa buat log in | | apk nya nggk bisa buat log in |
| 3 | Susah masuk dih | | susah masuk dih |
| 4 | Tidak bisa masuk inisiliasi | | tidak bisa masuk inisiliasi |
| ... | ... | | ... |
| 8353 | Semoga Tambah Baik,Mudah Dalam Input Dan Share... | | semoga tambah baik,mudah dalam input dan share... |
| 8354 | Cara masuk nya bagaimana? Pakai email kok gag ... | | cara masuk nya bagaimana? pakai email kok gag ... |
| 8355 | | | mbuh |
| 8356 | Semoga tidak seperti sirekap 2020 yang saat di... | | semoga tidak seperti sirekap 2020 yang saat di... |
| 8357 | Semoga dengan adanya aplikasi ini pemilu akan ... | | semoga dengan adanya aplikasi ini pemilu akan ... |

Gambar 3.7 Hasil *Lower Case*

Terlihat pada gambar diatas, *komentar_lowercase* berisi data ulasan yang telah dilakukan proses *lower_case* yang mengubah huruf kapital menjadi huruf kecil.

2. Remove Unnecessary Character

Atribut spesifik dan semua karakter alphanumeric atau karakter yang tidak diperlukan pada data ulasan seperti simbol, emoji, karakter yang berulang, tanda baca, dan karakter spesial lainnya dihapus menggunakan *remove_unnecessary_character*. Hasil dari *Remove Unnecessary Character* dapat dilihat pada gambar 3.8.

| komentar_lowercase \ | | komentar_remove_char | |
|----------------------|---|----------------------|---|
| 0 | mantap | 0 | mantap |
| 1 | sudah tau pemakainya masyarakat biasa malah sp... | 1 | sudah tau pemakainya masyarakat biasa malah sp... |
| 2 | apk nya nggk bisa buat log in | 2 | apk nya nggk bisa buat log in |
| 3 | susah masuk dih | 3 | susah masuk dih |
| 4 | tidak bisa masuk inisiliasi | 4 | tidak bisa masuk inisiliasi |
| ... | ... | ... | ... |
| 8353 | semoga tambah baik,mudah dalam input dan share... | 8353 | semoga tambah baik mudah dalam input dan share... |
| 8354 | cara masuk nya bagaimana? pakai email kok gag ... | 8354 | cara masuk nya bagaimana pakai email kok gag b... |
| 8355 | mbuh | 8355 | mbuh |
| 8356 | semoga tidak seperti sirekap 2020 yang saat di... | 8356 | semoga tidak seperti sirekap 2020 yang saat di... |
| 8357 | semoga dengan adanya aplikasi ini pemilu akan ... | 8357 | semoga dengan adanya aplikasi ini pemilu akan ... |

Gambar 3.8 Hasil *Remove Unnecessary Character*

komentar_remove_char adalah data ulasan yang telah dilakukan proses *Remove Unnecessary Character* pada data *komentar_lowercase*. Contoh perubahan yang terlihat pada *komentar_remove_char* adalah tanda baca seperti “,” dan “?” yang sebelumnya terdapat di *komentar_lowercase* pada baris 8353 dan baris 8354 berhasil dihapus.

3. Spell Checker

Hasil dari proses pembuatan kamus bernama “*kamus_tidak_baku*” yang digunakan sebagai referensi proses *Spell Checker* ditampilkan pada gambar berikut.

| | Original | Replacement |
|-------------------------|-----------------|-------------|
| 0 | Kata_Tidak_Baku | Kata_Baku |
| 1 | | 0 |
| 2 | | 00 |
| 3 | | 000d |
| 4 | | 1 |
| ... | | ... |
| 4404 | hahaha | ketawa |
| 4405 | oalah | ambigu |
| 4406 | ayok | ayo |
| 4407 | system | sistem |
| 4408 | geogle | google |
| [4409 rows x 2 columns] | | |

Gambar 3.9 Kamus Tidak Baku

Pada gambar 3.9, terlihat isi dari “*kamus_tidak_baku*” yang terdiri dari 4409 baris dan 2 kolom. Kolom Original merupakan kata – kata tidak baku atau kata yang tidak terdeteksi oleh file “4 lebih list kata KBBI.txt”. Sedangkan kolom Replacment merupakan kata – kata yang telah dibakukan.

Selanjutnya, dilakukan *Spell Checker* untuk memeriksa ejaan pada data ulasan dan mengubah ejaan menjadi lebih baik. Hasilnya dapat dilihat pada gambar 3.10.

| komentar_remove_char \ | | komentar_spellchecker | |
|------------------------|---|-----------------------|---|
| 0 | mantap | 0 | mantap |
| 1 | sudah tau pemakainya masyarakat biasa malah sp... | 1 | sudah tahu pemakainya masyarakat biasa malah s... |
| 2 | apk nya ngggk bisa buat log in | 2 | aplikasi nya tidak bisa buat masuk ini |
| 3 | susah masuk dih | 3 | susah masuk dih |
| 4 | tidak bisa masuk inisiliasi | 4 | tidak bisa masuk inisialisasi |
| ... | ... | ... | ... |
| 8353 | semoga tambah baik mudah dalam input dan share... | 8353 | semoga tambah baik mudah dalam memasukkan dan ... |
| 8354 | cara masuk nya bagaimana pakai email kok gag b... | 8354 | cara masuk nya bagaimana pakai email kok tidak... |
| 8355 | mbuh | 8355 | tidak tahu |
| 8356 | semoga tidak seperti sirekap 2020 yang saat di... | 8356 | semoga tidak seperti sirekap 2020 yang saat di... |
| 8357 | semoga dengan adanya aplikasi ini pemilu akan ... | 8357 | semoga dengan adanya aplikasi ini pemilu akan ... |

Gambar 3.10 Hasil *Spell Checker*

Proses *Spell Checker* menggunakan “kamus_tidak_baku” yang telah diinput. Seperti yang ditunjukkan Gambar 3.10, *komentar_spellchecker* merupakan data ulasan setelah dilakukan *Spell Checker* seperti "apk" menjadi "aplikasi", "inisiliasi" menjadi "inisialisasi", dan perubahan lainnya.

4. *Stemming*

Stemming menggunakan library *Sastrawi* untuk mengubah kata-kata yang ada pada data ulasan menjadi kata dasar. Hasil proses ini dapat dilihat pada gambar berikut.

| komentar_spellchecker | | komentar_stemming | |
|-----------------------|---|-------------------|---|
| 0 | mantap | 0 | mantap |
| 1 | sudah tahu pemakainya masyarakat biasa malah s... | 1 | sudah tahu maka masyarakat biasa malah spesifi... |
| 2 | aplikasi nya tidak bisa buat masuk ini | 2 | aplikasi nya tidak bisa buat masuk ini |
| 3 | susah masuk dih | 3 | susah masuk dih |
| 4 | tidak bisa masuk inisialisasi | 4 | tidak bisa masuk inisial |
| ... | ... | ... | ... |
| 8353 | semoga tambah baik mudah dalam memasukkan dan ... | 8353 | moga tambah baik mudah dalam masuk dan bagi data |
| 8354 | cara masuk nya bagaimana pakai email kok tidak... | 8354 | cara masuk nya bagaimana pakai email kok tidak... |
| 8355 | tidak tahu | 8355 | tidak tahu |
| 8356 | semoga tidak seperti sirekap 2020 yang saat di... | 8356 | moga tidak seperti sirekap 2020 yang saat guna... |
| 8357 | semoga dengan adanya aplikasi ini pemilu akan ... | 8357 | moga dengan ada aplikasi ini milu akan lebih b... |

Gambar 3.11 Hasil *Stemming*

Terlihat pada gambar 3.11, *komentar_stemming* adalah data ulasan setelah dilakukan proses *Stemming* dimana kata-kata pada data ulasan diubah menjadi kata dasar seperti kata “semoga” mejadi “moga”, “memasukkan” menjadi “masuk”, “inisialisasi” menjadi “inisial”, dan perubahannya lainnya.

5. Hapus Data Kosong

Setelah data ulasan melalui tahap *stemming*, dilakukan cek data kosong ‘NaN’ untuk melihat berapa data yang memiliki nilai ‘NaN’ yang terdapat pada *komentar_stemming*. Didapatkan sebanyak 60 data yang memilki nilai ‘NaN’ pada *komentar_stemming*. Selanjutnya, ‘NaN’ dihapus untuk menghilangkan data kosong tersebut. Hasilnya dapat dilihat pada gambar 3.12.

| ID | Nama | Rating | Waktu | Komentar | Like | komentar_lowerCase | komentar_remove_char | komentar_spellchecker | komentar_stemming |
|------|-------|-------------------|-------|------------------|---|--------------------|---|---|--|
| 0 | d0001 | Nency Miranda | 5 | 05/02/2024 14:07 | Mantap | 0.0 | mantap | mantap | mantap |
| 1 | d0002 | ichal sabiel | 1 | 05/02/2024 14:06 | Sudah tau pemakainya masyarakat biasa Malah sp... | 0.0 | sudah tau pemakainya masyarakat biasa malah sp... | sudah tau pemakainya masyarakat biasa malah sp... | sudah tahu maka masyarakat biasa malah spesifik... |
| 2 | d0003 | Evandra Aditya | 1 | 05/02/2024 14:06 | Apk nya nggk bisa buat log in | 0.0 | apk nya nggk bisa buat log in | apk nya nggk bisa buat log in | aplikasi nya tidak bisa buat masuk ini |
| 3 | d0004 | Ganesh insann | 1 | 05/02/2024 14:05 | Susah masuk dih | 0.0 | susah masuk dih | susah masuk dih | susah masuk dih |
| 4 | d0005 | Yohana Frediana | 1 | 05/02/2024 14:04 | Tidak bisa masuk inisiliasi | 0.0 | tidak bisa masuk inisiliasi | tidak bisa masuk inisiliasi | tidak bisa masuk inisial |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8353 | d8354 | Tn A Muntaha | 5 | 23/01/2024 09:05 | Semoga Tambah Baik,Mudah Dalam Input Dan Share... | 21.0 | semoga tambah baik,mudah dalam input dan share... | semoga tambah baik mudah dalam memasukan dan ... | moga tambah baik mudah dalam masuk dan bagi data |
| 8354 | d8355 | Adhi Nugroho | 1 | 23/01/2024 08:51 | Cara masuk nya bagaimana? Pakai email kok gag ... | 448.0 | cara masuk nya bagaimana? pakai email kok gag ... | cara masuk nya bagaimana pakai email kok gag b... | cara masuk nya bagaimana pakai email kok tidak... |
| 8355 | d8356 | TATANG RUSDIANA | 5 | 23/01/2024 07:03 | Mbuh | 6.0 | mbuh | mbuh | tidak tahu |
| 8356 | d8357 | vie | 5 | 23/01/2024 06:50 | Semoga tidak seperti sirekap 2020 yang saat di... | 42.0 | semoga tidak seperti sirekap 2020 yang saat di... | semoga tidak seperti sirekap 2020 yang saat di... | moga tidak seperti sirekap 2020 yang saat guna... |
| 8357 | d8358 | Ahmad ferdiansyah | 5 | 23/01/2024 04:26 | Semoga dengan adanya aplikasi ini pemilu akan ... | 186.0 | semoga dengan adanya aplikasi ini pemilu akan ... | semoga dengan adanya aplikasi ini pemilu akan ... | moga dengan ada aplikasi ini milu akan lebih b... |

8298 rows x 10 columns

Gambar 3.12 Hasil Hapus Data Kosong

Dapat dilihat pada gambar 3.12, data ulasan menjadi 8298 baris setelah dilakukan proses hapus data kosong. Ini menandakan bahwa penghapusan data kosong berhasil dilakukan. Perbandingan data ulasan sebelum dan setelah dilakukan tahap *Pra-proses* dapat dilihat pada gambar berikut.

| | | |
|------|---|--|
| 0 | | Mantap |
| 1 | Sudah tau pemakainya masyarakat biasa Malah sp... | |
| 2 | Apk nya nggk bisa buat log in | |
| 3 | Susah masuk dih | |
| 4 | Tidak bisa masuk inisiliasi | |
| ... | ... | ... |
| 8353 | Semoga Tambah Baik,Mudah Dalam Input Dan Share... | 8353 moga tambah baik mudah dalam masuk dan bagi data |
| 8354 | Cara masuk nya bagaimana? Pakai email kok gag ... | 8354 cara masuk nya bagaimana pakai email kok tidak... |
| 8355 | Mbuh | 8355 tidak tahu |
| 8356 | Semoga tidak seperti sirekap 2020 yang saat di... | 8356 moga tidak seperti sirekap 2020 yang saat guna... |
| 8357 | Semoga dengan adanya aplikasi ini pemilu akan ... | 8357 moga dengan ada aplikasi ini milu akan lebih b... |

Name: Komentar, Length: 8358, dtype: object

Name: komentar stemming, Length: 8298, dtype: object

Gambar 3.13 Hasil *Pra-proses*

Dapat dilihat pada gambar 3.13, gambar disisi kiri adalah data ulasan sebelum dilakukan Tahap *Pra-proses*, Sedangkan gambar disisi kiri merupakan data ulasan yang telah melalui tahap *Pra-proses*.

3.1.3 Wordnet

Tahapan selanjutnya setelah *Pra-proses* adalah melakukan tahapan analisis sentimen menggunakan metode *Wordnet*. Sebelum melakukan analisis sentimen menggunakan *Wordnet*, data ulasan diterjemahkan ke dalam bahasa Inggris. Hasil penerjemahan dapat dilihat pada gambar 3.14.

| komentar_stemming \ | | Translated | |
|---------------------|--|--------------------------------------|---|
| 0 | mantap | 0 | Excellent |
| 1 | sudah tahu maka masyarakat biasa malah spesifik... | 1 | You already know that ordinary people even hav... |
| 2 | aplikasi tidak bisa buat masuk in | 2 | the application can't do this entry |
| 3 | susah masuk dih | 3 | it's hard to get in |
| 4 | tidak bisa masuk inisial | 4 | Can't enter initials |
| ... | ... | ... | ... |
| 8297 | moga tambah baik mudah dalam masuk dan share data | 8293 | Hopefully it will be easier to enter and share... |
| 8298 | cara masuk bagaimana pakai email kok tidak bis... | 8294 | How come you can log in using email? How come ... |
| 8299 | tidak tahu | 8295 | Don't know |
| 8300 | moga tidak seperti sirekap 2020 yang saat guna... | 8296 | Hopefully it won't be like the 2020 recap wher... |
| 8301 | moga dengan ada aplikasi ini milu akan lebih b... | 8297 | Hopefully with this application Milu will be m... |
| | | [8298 rows x 2 columns] | |
| | | Waktu proses: 6387.12087059021 detik | |

Gambar 3.14 Hasil *Deep-translator*

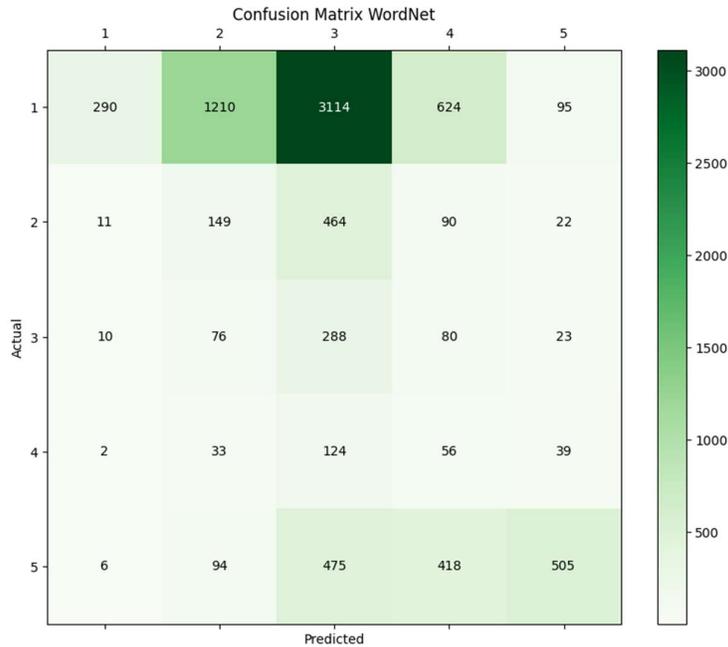
Terlihat pada gambar 3.14, *Translated* merupakan data ulasan dari *komentar_stemming* yang telah melalui tahap penerjemahan dari bahasa Indonesia ke bahasa Inggris. Proses ini membutuhkan waktu sekitar 107 menit untuk menerjemahkan data ulasan. Selanjutnya, mencari nilai polaritas menggunakan *TextBlob* dengan memanfaatkan *Wordnet* dan mengklasifikasinya dalam 5 kelas peringkat berdasarkan tabel 2.2. Hasilnya dapat dilihat pada gambar berikut.

| ID | Rating | komentar_stemming | Translated | Score_Wordnet | Sentiment_Score | |
|----|--------|-------------------|--|---|-----------------|-----------|
| 0 | d0001 | 5 | mantap | Excellent | 5 star | 1.000000 |
| 1 | d0002 | 1 | sudah tahu maka masyarakat biasa malah spesifik... | You already know that ordinary people even hav... | 3 star | -0.045000 |
| 2 | d0003 | 1 | aplikasi nya tidak bisa buat masuk ini | the application can't do this entry | 3 star | 0.000000 |
| 3 | d0004 | 1 | susah masuk dih | it's hard to get in | 2 star | -0.291667 |
| 4 | d0005 | 1 | tidak bisa masuk inisial | Can't enter initials | 3 star | 0.000000 |

Gambar 3.15 Hasil *Wordnet*

Terlihat pada gambar 3.15, setiap komentar memiliki nilai polaritas masing – masing yang ada pada kolom “*Sentiment_Score*” dan kemudian diklasifikasikan ke dalam 5 kelas peringkat yang ditampilkan pada kolom “*Score_Wordnet*”. Seperti pada komentar “mantap” yang memiliki nilai polaritas 1 dan masuk kedalam kelas peringkat 5 *star* karena memiliki nilai lebih dari 0,5. Terlihat juga bahwa *rating* yang diberikan pengguna dengan kelas peringkat berdasarkan nilai polaritas ada yang tidak sama. Contohnya pada komentar “tidak bisa masuk inisial” yang diberikan *rating* 1 oleh pengguna namun memiliki nilai polaritas 0 dan masuk kedalam kelas peringkat 3 *star* karena memiliki nilai yang kurang dari 0,2 dan lebih dari -0,2.

Wordnet dievaluasi menggunakan *Confusion Matrix* untuk mencari nilai *F1-Score*. *Confusion Matrix* terdiri dari *Actual* dan *Predicted*. *Actual* merupakan label kelas sebenarnya dari data ulasan, sedangkan *Predicted* merupakan kelas yang diklasifikasi oleh model pada data ulasan. Hasil *Confusion Matrix* pada klasifikasi menggunakan *Wordnet* dapat dilihat pada gambar 3.16.



Gambar 3.16 *Confusion Matrix* dengan *Wordnet*

Dapat dilihat pada gambar 3.16, warna hijau yang lebih intens menandakan data test yang paling banyak diklasifikasi, sedangkan warna hijau yang lebih pudar menandakan data test yang paling sedikit diklasifikasi. data test yang paling banyak diklasifikasi adalah 3114 data test yang seharusnya dilabelkan pada *rating* 1, namun diklasifikasikan kedalam kelas peringkat 3 *star*. Sedangkan data test yang paling sedikit diklasifikasi adalah 2 data test yang seharusnya dilabelkan pada *rating* 4, namun diklasifikasikan kedalam kelas peringkat 1 *star*.

Dari gambar 3.16 dapat disimpulkan bahwa, dari 5361 data pada *rating* 1, hanya 298 data test yang berhasil diklasifikasi benar. 736 data pada *rating* 2, hanya 149 data test yang berhasil diklasifikasi benar. 478 data pada *rating* 3, yang berhasil diklasifikasi benar hanya 288 data test. 255 data pada *rating* 4, hanya 56 data test yang berhasil diklasifikasi benar. Sedangkan, 1528 data pada *rating* 5, hanya 505 data test yang berhasil diklasifikasi benar.

Hasil dari *Confusion Matrix*, dilakukan perhitungan performa dan didapatkan hasil *F1-Score* yaitu 17.70%. Proses ini membutuhkan waktu sekitar 1 detik untuk memproses evaluasi *F1-Score* pada model *Wordnet*. Nilai tersebut menunjukkan hasil evaluasi *F1-Score* yang didapatkan dari *Wordnet* dalam menganalisis sentimen pada ulasan aplikasi SIREKAP 2024.

3.1.4 Klasifikasi *Decision Tree* CART

Dalam ekstraksi fitur TF-IDF, data dibagi menjadi 2 bagian, yaitu X (Gambar 3.17) sebagai kolom komentar *_stemming* dan y (Gambar 3.18) sebagai kolom *rating*. Berikut gambar dari kolom X dan y.

```

1 X
✓ 0.0s
0
1          mantap
2  sudah tahu maka masyarakat biasa malah spesifi...
3          aplikasi nya tidak bisa buat masuk ini
4          susah masuk dih
          tidak bisa masuk inisial
          ...
8293  moga tambah baik mudah dalam masuk dan bagi data
8294  cara masuk nya bagaimana pakai email kok tidak...
8295          tidak tahu
8296  moga tidak seperti sirekap 2020 yang saat guna...
8297  moga dengan ada aplikasi ini milu akan lebih b...
Name: komentar_stemming, Length: 8298, dtype: object

```

Gambar 3.17 komentar_stemming

```

1 y
✓ 0.0s
0      5
1      1
2      1
3      1
4      1
      ..
8293   5
8294   1
8295   5
8296   5
8297   5
Name: Rating, Length: 8298, dtype: int64

```

Gambar 3.18 Rating

Dapat dilihat pada gambar 3.17 dan gambar 3.18, data telah dibagi menjadi dua bagian yang ditandai dengan variabel X sebagai data komentar dan variabel y sebagai data *rating* dari data ulasan. Selanjutnya dilakukan perhitungan TF-IDF pada variabel X untuk mengubah . Hasil perhitungan TF-IDF dapat dilihat pada gambar 3.19.

| | |
|--------------|---------------------|
| (0, 1376) | 1.0 |
| (1, 259) | 0.6792391025684885 |
| (1, 887) | 0.10383584445945196 |
| (1, 206) | 0.07652109295763325 |
| (1, 1614) | 0.14506596822529783 |
| (1, 2410) | 0.13373232476819363 |
| (1, 2320) | 0.27802569670255944 |
| (1, 825) | 0.16356998608982817 |
| (1, 2155) | 0.27307728845508056 |
| (1, 1359) | 0.13781584933616398 |
| (1, 380) | 0.25004926716628667 |
| (1, 1394) | 0.29022619285877926 |
| (1, 1348) | 0.2746602935493359 |
| (1, 2207) | 0.22574459106729391 |
| (1, 2173) | 0.11088639682045809 |
| (2, 1393) | 0.3060772188862647 |
| (2, 439) | 0.5368957909043959 |
| (2, 400) | 0.3187697434660648 |
| (2, 2307) | 0.287055929249031 |
| (2, 1573) | 0.4683618676203409 |
| (2, 887) | 0.3729945196202908 |
| (2, 206) | 0.27487567956070985 |
| (3, 600) | 0.9096105750431945 |
| (3, 2193) | 0.34717119421206155 |
| (3, 1393) | 0.22821210239371564 |
| ... | |
| (8297, 1061) | 0.26376030223663394 |
| (8297, 887) | 0.15294509486441463 |
| (8297, 206) | 0.11271180855185466 |
| (8297, 1394) | 0.4274889160867539 |

Gambar 3.19 Hasil TF-IDF

Hasil pada gambar 3.19 menunjukkan nilai yang telah dihitung menggunakan TF-IDF. Contohnya (0, 1376), yang menunjukkan pada indeks 0 terdapat kata dengan term 1376 yang berisi kata ‘mantap’ dengan nilai TF-IDF yaitu 1,0 dan seterusnya.

Setelah menghitung ekstraksi fitur TF-IDF, performa model dievaluasi dengan membagi data menjadi 10 *fold*. *Fold* 1 sampai *fold* 8 membagi data menjadi 7468 sampel sebagai data train dan data 830 sampel sebagai data test. Sedangkan, *fold* 9 dan *fold* 10 membagi data menjadi 7469 sampel sebagai data train dan 829 sampel sebagai data test. Berikut pembagian indeks data yang digunakan pada *fold* 1.

Tabel 3.2 *Fold* 1 *Cross-Validation*

| Train Index | Test Index |
|--|-------------------------------|
| [830, 831, 832, ..., 8295, 8296, 8297] | [0, 1, 2, ..., 827, 828, 829] |

Tabel 3.2 menunjukkan *index* yang digunakan dalam melatih dan menguji model pada *fold* 1. *Train Index* merupakan sampel yang digunakan pada data train *fold* 1, yaitu *index* 830 sampai 8279. Sedangkan *Test Index* merupakan sampel yang digunakan pada data test *fold* 1, yaitu *index* 0 sampai 829. Selanjutnya, pembagian indeks data yang digunakan pada *fold* 10 yang memiliki perbedaan penggunaan sampel data dengan *fold* 1.

Tabel 3.3 *Fold 10 Cross-Validation*

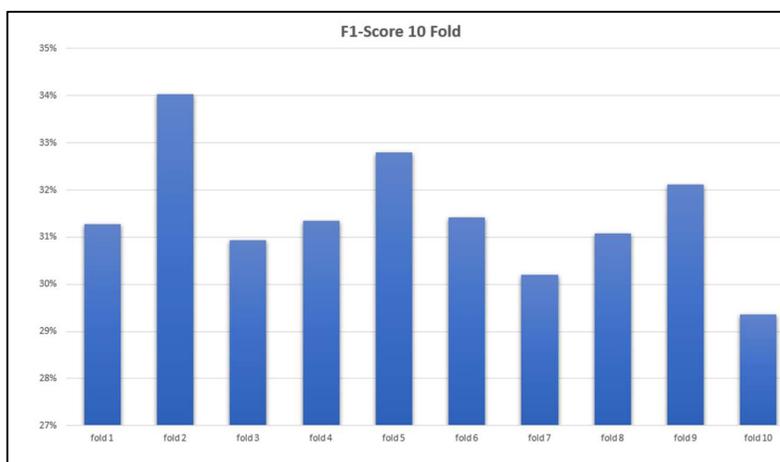
| <i>Train Index</i> | <i>Test Index</i> |
|----------------------------------|---|
| [0, 1, 2, ..., 7466, 7467, 7468] | [7469, 7470, 7471, ..., 8295, 8296, 8297] |

Tabel 3.3 menunjukkan *index* yang digunakan dalam melatih dan menguji model pada *fold* 10. *Train Index* merupakan sampel yang digunakan pada data train *fold* 10, yaitu *index* 0 sampai 7468. Sedangkan *Test Index* merupakan sampel yang digunakan pada data test *fold* 10, yaitu *index* 7469 sampai 8297. Setelah itu, dilakukan perhitungan performa menggunakan *F1-Score* pada masing – masing *fold*. Hasilnya dapat dilihat pada tabel 3.4.

Tabel 3.4 Hasil *F1-Score* pada Setiap *Fold*

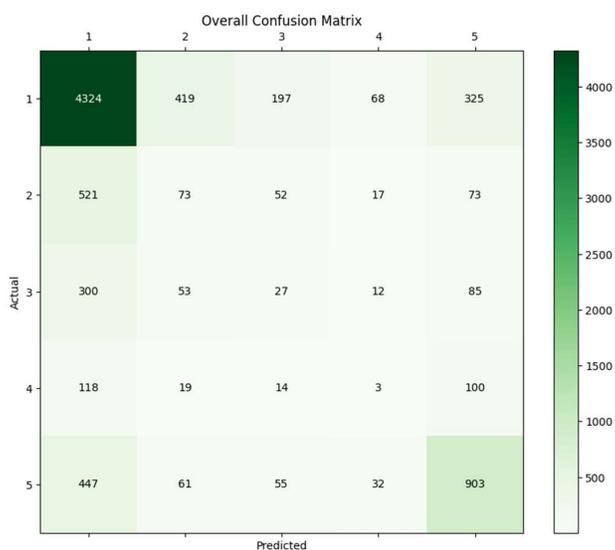
| <i>Fold</i> | <i>F1-Score</i> |
|-------------|-----------------|
| 1 | 31.27% |
| 2 | 34.03% |
| 3 | 30.94% |
| 4 | 31.35% |
| 5 | 32.79% |
| 6 | 31.42% |
| 7 | 30.19% |
| 8 | 31.07% |
| 9 | 32.11% |
| 10 | 29.36% |

Tabel diatas menunjukkan hasil *F1-Score* dari masing – masing *fold* menggunakan *Cross-Validation* 10 *fold*. dapat dilihat *fold* yang memiliki nilai *F1-Score* tertinggi adalah *fold* 2 dengan nilai *F1-Score* sebesar 34.03% dan *fold* yang memiliki nilai *F1-Score* terendah adalah *fold* 10 dengan nilai *F1-Score* sebesar 29.36%. Perbandingan nilai *F1-Score* pada setiap *fold* dapat dilihat pada gambar 3.20.



Gambar 3.20 *F1-Score 10 Fold*

Gambar 3.20 merupakan grafik performa pada setiap *fold* yang dihitung menggunakan *F1-Score*. Dari 10 *fold* tersebut, *fold 2* memiliki performa tertinggi, sementara *fold 10* memiliki performa terendah dibandingkan *fold* lainnya. Ini menunjukkan bahwa pembagian data train dan data test yang menghasilkan performa tertinggi berada pada *fold 2*, sedangkan pembagian data dengan performa terendah berada pada *fold 10* berdasarkan hasil *K-fold Cross-Validation* $K=10$. Selanjutnya nilai dari setiap *fold* digunakan untuk menghitung hasil keseluruhan nilai *fold*. Berikut hasil keseluruhan nilai *fold* yang divisualisasikan dengan *Confusion Matrix*.



Gambar 3.21 *Confusion Matrix Overall* dengan *Decision Tree CART*

Gambar 3.21 merupakan tampilan *Confusion Matrix* dari hasil keseluruhan tahapan klasifikasi dari *Decision Tree* dengan algoritma *CART*. Terlihat data ulasan yang paling banyak benar diklasifikasikan oleh *Decision Tree* dengan algoritma *CART* ada pada ulasan dengan *rating 1* dan *5* yaitu sebanyak 4324 data untuk *rating 1* dan 903 data untuk *rating 5*. Dari *Confusion Matrix* tersebut, didapatkan hasil dari *Macro F1-Score*, yaitu 31.62%. Waktu yang dibutuhkan untuk proses ini adalah 32 detik. Hal ini menunjukkan hasil evaluasi *F1-Score* menggunakan metode *Decision Tree* dengan algoritma *CART* dalam menganalisis sentimen ulasan aplikasi SIREKAP 2024.

3.2 Pembahasan

3.2.1 Pengambilan dan Analisis Data

Penelitian ini menggunakan data ulasan dari pengguna aplikasi SIREKAP 2024 di *Google Playstore* yang diambil pada tanggal 6 februari 2024. Setelah dilakukan analisis data, data pada penelitian ini merupakan data tidak seimbang dimana jumlah kemunculan data pada *rating* 1 dan 5 jauh lebih banyak dibandingkan *rating* 2, 3, dan 4. Menurut (Provost, 2000; Utami, 2022), data yang tidak seimbang menyebabkan kesalahan klasifikasi pada kelas minoritas, karena data cenderung lebih mendukung kelas mayoritas.

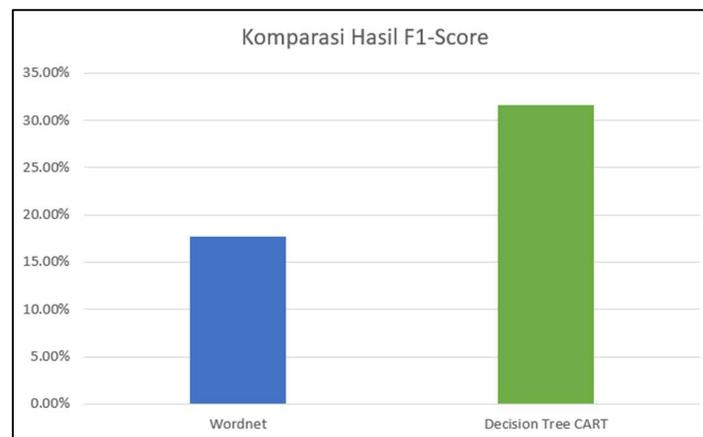
3.2.2 Analisis Hasil *Pra-proses*

Pada hasil tahap *Pra-proses*, tahap ini berhasil menghapus karakter non-alfanumerik, membenarkan sebagian besar kesalahan ejaan kata, serta membersihkan data kosong yang pada data ulasan. Namun, masih terdapat masalah pada hasil *Pra-proses* dipenelitian ini. Dapat dilihat pada gambar 3.13, terdapat kata “dih” yang tidak memiliki arti atau disebut dengan kata *slank* yang seharusnya dihapus. Selain itu, terdapat juga kesalahan pengubahan kata imbuhan menjadi kata dasar, seperti kata “pemainnya” menjadi “maka” dimana dua kata tersebut memiliki makna yang berbeda. Hal ini dapat mempengaruhi hasil klasifikasi *Wordnet* dan *Decision Tree* CART karena menggunakan data ulasan yang telah melalui tahap *Pra-proses*.

3.2.3 Analisis Komparasi Model Klasifikasi *Wordnet* dan CART

Model *Wordnet* dan *Decision Tree* dengan algoritma CART memiliki kelebihan dan kekurangannya masing-masing dalam mengklasifikasi sentimen pada data ulasan di penelitian ini. Pada model *Wordnet*, sebenarnya klasifikasi dan evaluasi *F1-Score* yang dilakukan lebih unggul dari segi kecepatan dibandingkan *Decision Tree* dengan algoritma CART. Namun, model *Wordnet* memerlukan waktu yang lama dalam menerjemahkan data ulasan dari bahasa Indonesia ke bahasa Inggris sehingga model *Decision Tree* dengan algoritma CART menjadi lebih cepat.

Selain itu, berdasarkan hasil evaluasi *F1-Score* yang didapatkan dari masing masing model, performa yang dihasilkan model *Decision Tree* dengan algoritma CART lebih tinggi dibandingkan model *Wordnet*. perbandingannya dapat dilihat pada gambar 3.22 dan tabel 3.5.



Gambar 3.22 Komparasi Hasil *F1-Score*

Tabel 3.5 Hasil *F1-Score* Kedua Metode

| Metode | F1 - Score |
|---------------------------|------------|
| <i>Wordnet</i> | 17.70% |
| <i>Decision Tree</i> CART | 31.62% |

Berdasarkan tabel 3.5 dan gambar 3.22, dapat dilihat bahwa model *Decision Tree* dengan algoritma CART memiliki performa yang lebih tinggi dengan nilai *F1-Score* 31.62%, dibandingkan dengan model *Wordnet* yang memiliki nilai *F1-Score* 17.70% dalam menganalisis sentimen pada ulasan aplikasi SIREKAP 2024.

Rendahnya performa yang dihasilkan oleh hasil evaluasi *F1-Score* pada model *Wordnet* dibandingkan model *Decision Tree* CART, menandakan klasifikasi pada model *Wordnet* kurang baik dibandingkan model *Decision Tree* CART. Secara keseluruhan, model *Wordnet* banyak melakukan kesalahan dalam mengklasifikasi data ulasan, terutama pada data test yang seharusnya berada pada *rating* 1 namun dilabelkan ke dalam kelas peringkat 3 *star*, yaitu sebanyak 3114 data. Sedangkan data ulasan yang diklasifikasi benar pada *Wordnet* berada pada data test yang memiliki *rating* 5 dan dilabelkan ke dalam kelas peringkat 5 *star*, yaitu 505 data. Ini menandakan bahwa klasifikasi yang dilakukan pada model *Wordnet* lebih banyak yang salah dibandingkan yang benar.

Selanjutnya pada model *Decision Tree* CART, data ulasan yang diklasifikasi benar lebih banyak dibandingkan yang klasifikasi salah. Ini ditunjukkan pada gambar 3.21, dimana data test yang paling banyak diklasifikasi benar berada pada data test yang seharusnya memiliki *rating* 1 dan diklasifikasi ke dalam *rating* 1, yaitu sebanyak 4324 data. Sedangkan data ulasan yang paling banyak diklasifikasi salah berada pada data test yang seharusnya memiliki *rating* 5 namun diklasifikasi ke dalam *rating* 1, yaitu sebanyak 447 data. Hal ini membuktikan bahwa klasifikasi yang dilakukan pada model *Decision Tree* CART lebih banyak yang benar dibandingkan yang salah.

Hasil klasifikasi yang didapatkan dari *Confussion Matrix* pada masing-masing model menunjukkan bahwa, klasifikasi ulasan yang dilakukan oleh *Decision Tree* dengan algoritma CART lebih unggul dibandingkan model *Wordnet*. Hal ini menjadi alasan mengapa performa *F1-Score* pada model *Wordnet* lebih rendah dibandingkan performa *F1-Score* pada model *Decision Tree* dengan algoritma CART.

Hasil penelitian ini menunjukkan bahwa model *Wordnet* dan *Decision Tree* dengan algoritma CART terbukti dapat digunakan dalam menganalisis sentimen pada ulasan aplikasi SIREKAP 2024 yang memiliki data yang tidak seimbang. Setelah dianalisis komparasi dari kedua model, metode *Decision Tree* dengan algoritma CART menjadi metode yang lebih baik dibandingkan dengan metode *Wordnet* dalam mengklasifikasi sentimen ulasan pada aplikasi SIREKAP 2024.