



**SENTIMEN ANALISIS REVIEW APLIKASI RUANG GURU PADA  
GOOGLE PLAYSTORE DENGAN METODE NAÏVE BAYES**

**SKRIPSI**

**Oleh:**

Christoper Andrian

20211000013

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS BUDDHI DHARMA  
TANGERANG**

**2025**



**SENTIMENT ANALISIS REVIEW APLIKASI RUANG GURU PADA  
GOOGLE PLAYSTORE DENGAN METODE NAÏVE BAYES**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada  
Program Studi Teknik Informatika Sains dan Teknologi  
Universitas Buddhi Dharma  
Jenjang Pendidikan Strata 1**

**Oleh:**

**CHRISTOPER ANDRIAN**

**20211000013**

**FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS BUDDHI DHARMA  
TANGERANG**

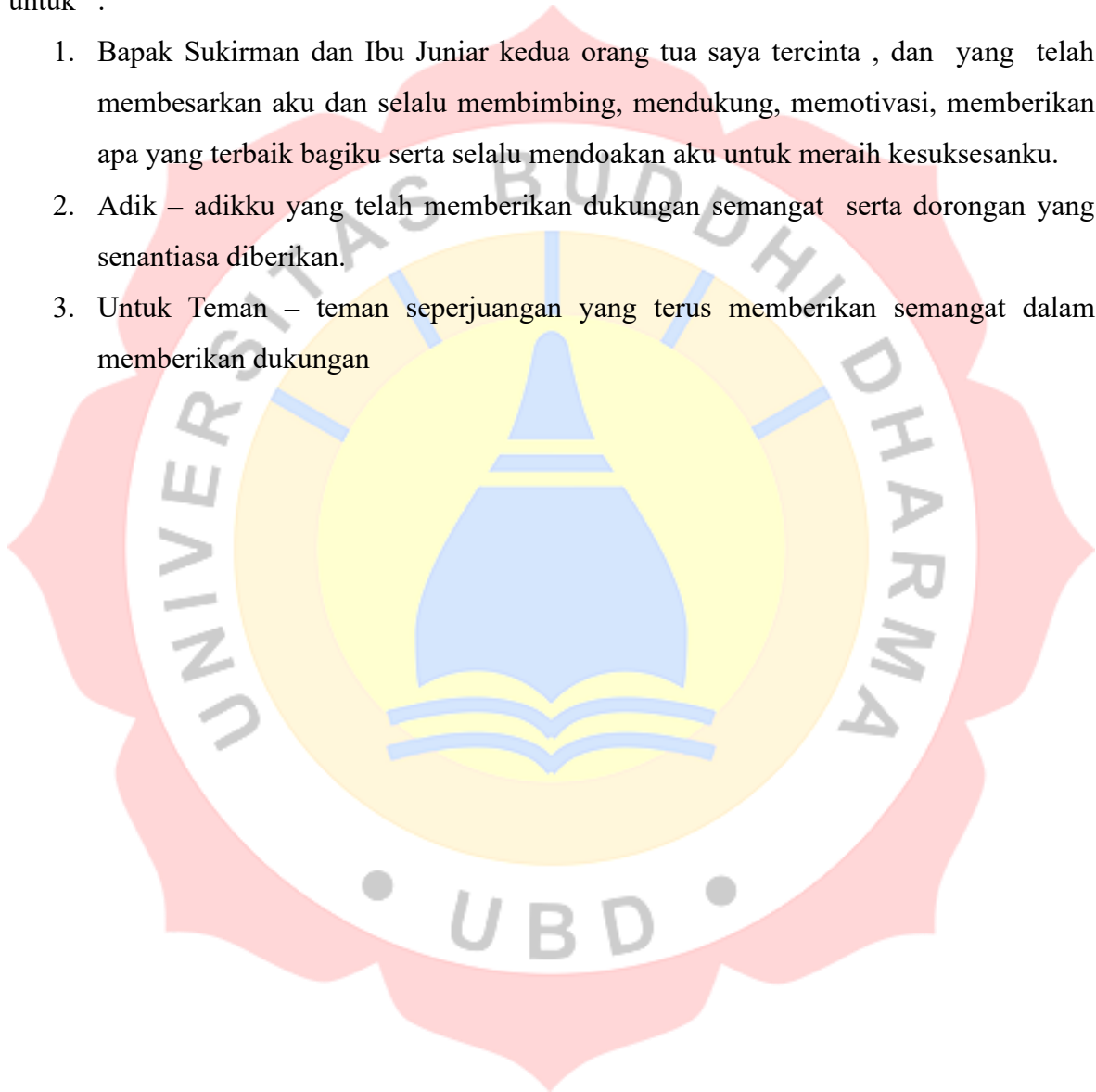
**2025**

## LEMBAR PERSEMBAHAN

“Keberhasilan bukan milik orang pintar, tapi milik mereka yang senantiasa konsisten dalam berusaha”

Dengan mengucapkan puji syukur kepada Tuhan Yang Maha Esa, SKRIPSI ini ku persembahkan untuk :

1. Bapak Sukirman dan Ibu Juniar kedua orang tua saya tercinta , dan yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberikan apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
2. Adik – adikku yang telah memberikan dukungan semangat serta dorongan yang senantiasa diberikan.
3. Untuk Teman – teman seperjuangan yang terus memberikan semangat dalam memberikan dukungan



# UNIVERSITAS BUDDHI DHARMA

## LEMBARAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini.

NIM : 20211000013  
Nama : CHRISTOPER ANDRIAN  
Jejang Studi : Strata1  
Program Studi : Teknik Informatika  
Peminatan : Data Sciene

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademi Sarjana atau kelengkapan, studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembaran pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ke tidak benaran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya per oleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 06- 08-2025

Yang membuat pernyataan



CHRISTOPER ANDRIAN

20211000013

# UNIVERSITAS BUDDHI DHARMA

## LEMBARAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini .

NIM : 20211000013  
Nama : CHRISTOPER ANDRIAN  
Jejang Studi : Strata1  
Program Studi : Teknik Informatika  
Peminatan : Data Sciene

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non- Eksklusif ( *Non -exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul : “SENTIMEN ANALISIS REVIEW APLIAKSI RUANG GURU PADA GOOGLE PLAYSTORE DENGAN METHODE NAIVE BAYES”, beserta alat yang diperlukan (apabila ada). Dengan Hak bebas Royalti Non- Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis pertama atau pencipta karya ilmiah tersebut. Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 06- 08-2025

Yang membuat pernyataan



CHRISTOPER ANDRIAN

20211000013

**LEMBARAN PENGESAHAN PEMBIMBING**  
**SENTIMENT ANALISIS REVIEW APLIKASI RUANG GURU PADA**  
**GOOGLE PLAYSTORE DENGAN METODE NAÏVE BAYES**

Dibuat Oleh:

NIM : 20211000013

Nama : CHRISTOPER ANDRIAN

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan Oleh,

Tangerang, 06-08-2025

**Pembimbing,**



Yo Ceng Giap, M.Kom., CPS

NUPTK. 7044758659131143

**LEMBARAN PENGESAHAN SKRIPSI**  
**SENTIMENT ANALISIS REVIEW APLIKASI RUANG GURU PADA**  
**GOOGLE PLAYSTORE DENGAN METODE NAÏVE BAYES**

Dibuat Oleh:

NIM : 20211000013

Nama : CHRISTOPER ANDRIAN

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan Oleh,

Tangerang, 06-08-2025

**Dekan,**

**Ketua Program Studi**

Dr. Yakub, M.Kom., M.M.

NUPTK. 183674764813017

Hartana Wijaya, M.Kom.

NUPTK. 3844759660131192

## LEMBAR PENGESAHAN TIM PENGUJI

Nama : CHRISTOPER ANDRIAN  
NIM : 20211000013  
Fakultas : Sains dan Teknologi  
Judul Skripsi : Sentimen Analisi Review Aplikasi Ruang Guru  
Pada Google Playstore Dengan Metode Naïve  
Bayes

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji Pada hari Rabu, 06-08-2025.

Nama Penguji :  
Ketua Sidang : Yusuf Kurnia, M.Kom  
NUPTK. 4551765666130223  
Penguji I : Dram Renaldi, M.Kom  
NUPTK. 0443768669130322  
Penguji II : Yo Ceng Giap, M.Kom., CPS  
NUPTK. 7044758659131143

Tanda Tangan :

Mengetahui,  
**Dekan Fakultas Sains dan Teknologi**

Dr. Yakub, M.Kom., M.M.  
NUPTK. 1836747648130172



## KATA PENGANTAR

Dengan mengucapkan Puji kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul SENTIMEN ANALISIS REVIEW APLIKASI RUANG GURU PADA GOOGLE PLAY DENGAN METODE NAÏVE BAYES. Tujuan utama dari Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P. sebagai Rektor Universitas Buddhi Dharma.
2. Bapak Dr. Yakub, M.Kom., M.M. Dekan Fakultas Sains dan Teknologi.
3. Bapak Hartana Wijaya, M.Kom sebagai Ketua Program Studi Teknik Informatika.
4. Bapak Yo Ceng Giap, M.Kom.,CPS. sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan Skripsi ini.
5. Kedua orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
6. Teman – teman yang selalu membantu dan memberikan semangat.

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang. Akhir kata semoga Penelitian Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 06-08-2025

Penulis

## ABSTRAK

Pesatnya teknologi berkembang dalam bidang pendidikan, terutama dibidang *internet* telah mengubah orang mencari informasi, media belajar secara *online* dengan bantuan produk digital atau biasa dikenal sebagai aplikasi Ruang Guru, di *Google PlayStore* sebagai platform penyedia untuk pengguna menemukan aplikasi atau produk digital. Setelah inovasi terhadap aplikasi pendidikan ini, masyarakat atau pengguna memperdebatkan kelebihan dan kekurangan aplikasi Ruang Guru. Pada setiap ulasannya komentar *review* memiliki *rating* skala 1 sampai 5, namun *rating* yang diberikan oleh pengguna sering kali tidak sesuai dengan ulasan komentar *review* sehingga tidak cukup untuk menggambarkan kelebihan dan kekurangan pada aplikasi Ruang Guru. Untuk mengatasi masalah tersebut diperlukan analisa sentimen untuk memahami sentimen secara sistematis. Maka dari itu dirancang aplikasi analisa sentimen dengan mengimplementasikan metode *Naïve Bayes* dalam mengklasifikasikan sentimen menjadi kategori positif, netral, dan negatif. Aplikasi ini menampilkan analisis dalam bentuk grafik yang memberikan gambaran visual tentang sentimen positif, netral dan negatif yang di dapatkan informasi dari dalam ulasan dan juga untuk sentimen analisa komentar. Dengan melakukan perbandingan tiga skenario yaitu 70:30, 80:20, dan 90:10. Berdasarkan hasil perbandingan antara tiga skenario tersebut m skenario terbaik untuk model klasifikasi *Naïve Bayes* adalah 90:10 yang menunjukkan nilai akurasi sebesar 83,52%, presisi 71,44%, *recall* 65,84% dan *F1-score* 67,83%. Dan adanya aplikasi ini, diharapkan dapat membantu masyarakat atau pengguna membuat keputusan dalam memilih aplikasi yang lebih terinformasi berdasarkan ulasan – ulasan yang telah dianalisis secara efektif.

**Kata Kunci :** *Analisis Sentimen, Aplikasi Ruang Guru, Google Playstore, Text Mining*

## ***ABSTRACT***

*The rapid development of technology in the field of education, especially in the field of the internet, has changed people looking for information and learning media online with the help of digital products, or commonly known as the Ruang Guru application, on the Google Play Store as a provider platform for users to find applications or digital products. After the innovation of this educational application, people or users debate the advantages and disadvantages of the Ruang Guru application. In each review, the review comments have a rating scale of 1 to 5, but the ratings given by users often do not match the review comments so that they are not sufficient to describe the advantages and disadvantages of the Ruang Guru application. To overcome this problem, sentiment analysis is needed to understand sentiment systematically. Therefore, a sentiment analysis application was designed by implementing the Naïve Bayes method in classifying sentiment into positive, neutral, and negative categories. This application displays the analysis in the form of graphs that provide a visual representation of positive, neutral, and negative sentiments obtained from the information in the reviews and also for sentiment analysis of comments. By comparing three scenarios, namely 70:30, 80:20, and 90:10. Based on the comparison between the three scenarios, the best scenario for the Naïve Bayes classification model is 90:10, which shows an accuracy value of 83,52%, precision of 71,44%, recall of 65,84%, and F1-score of 67,83%. And the existence of this application is expected to help people or users make decisions in choosing a more informed application based on reviews that have been analyzed effectively.*

**Keyword :** *application Ruang Guru, Google Playstore, Sentimen Analisis, Text Mining*

# DAFTAR ISI

|                                                          |      |
|----------------------------------------------------------|------|
| LEMBAR PERSEMBAHAN.....                                  | i    |
| LEMBARAN PERYATAAN KEASLIAN SKRIPSI .....                | ii   |
| LEMBARAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....        | iii  |
| LEMBARAN PENGESAHAN PEMBIMBING .....                     | iv   |
| LEMBARAN PENGESAHAN SKRIPSI .....                        | v    |
| LEMBAR PENGESAHAN TIM PENGUJI .....                      | vi   |
| KATA PENGANTAR .....                                     | vii  |
| ABSTRAK .....                                            | viii |
| <i>ABSTRACT</i> .....                                    | ix   |
| DAFTAR ISI.....                                          | x    |
| DAFTAR TABEL .....                                       | xiii |
| DAFTAR GAMBAR.....                                       | xiv  |
| DAFTAR LAMPIRAN.....                                     | xv   |
| BAB I PENDAHULUAN .....                                  | 1    |
| 1.1    Latar Belakang Penelitian.....                    | 1    |
| 1.2    Identifikasi Masalah .....                        | 4    |
| 1.3    Ruang Lingkup Penelitian .....                    | 4    |
| 1.4    Tujuan dan Manfaat Penelitian .....               | 5    |
| 1.4.1.    Tujuan.....                                    | 5    |
| 1.4.2.    Manfaat.....                                   | 5    |
| 1.5    Sistematika Penulisan.....                        | 6    |
| BAB II LANDASAN TEORI.....                               | 7    |
| 2.1    Teori.....                                        | 7    |
| 2.1.1.    Data.....                                      | 7    |
| 2.1.2.    Informasi.....                                 | 8    |
| 2.1.3.    Ruang Guru .....                               | 9    |
| 2.1.4.    Pelabelan Data Otomatis ( <i>Vader</i> ) ..... | 9    |
| 2.1.5.    Analisis Sentimen.....                         | 10   |
| 2.1.6. <i>Text Mining</i> .....                          | 11   |
| 2.1.7. <i>Python</i> .....                               | 12   |
| 2.1.8.    Klasifikasi.....                               | 13   |

|                                     |                                                                                                                              |    |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1.9.                              | <i>Naïve Bayes</i> .....                                                                                                     | 14 |
| 2.1.10.                             | <i>Naïve Bayes</i> Untuk Klasifikasi .....                                                                                   | 15 |
| 2.1.11.                             | <i>Confusion Matrix</i> .....                                                                                                | 17 |
| 2.1.12.                             | <i>Website</i> .....                                                                                                         | 19 |
| 2.1.13.                             | <i>Laravel</i> .....                                                                                                         | 19 |
| 2.1.14.                             | <i>Mysql</i> .....                                                                                                           | 20 |
| 2.1.15.                             | Rest API .....                                                                                                               | 20 |
| 2.2                                 | Tinjauan Studi .....                                                                                                         | 21 |
| 2.2.1.                              | Penelitian Angelina P. Giovani, Tuti Haryanti dan Windu Gata .....                                                           | 21 |
| 2.2.2.                              | Penelitian M. Yusuf Rismanda, Iqbal Maulana, Oman Kamarudin .....                                                            | 22 |
| 2.2.3.                              | Penelitian Rio Christian, Indah Fenriana .....                                                                               | 23 |
| 2.2.4.                              | Penelitian Alqis Rausanfita, Arip Ramadan, M Dzulfikar Fauzi, Qori Putri Mafidah, Emilia Ramona, Yudha Mahardika Putra ..... | 24 |
| 2.2.5.                              | Penelitian Roopam Srivastaya, P.K Bharti, Parul Verma .....                                                                  | 24 |
| 2.2.6.                              | Penelitian Mohamed Chiny, Marouane Chihab, Omar Bencharf, Younes Chihab .....                                                | 25 |
| 2.2.7.                              | Penelitian Subhasis Sanyal, Mobit Kumar Barai .....                                                                          | 26 |
| 2.2.8.                              | Penelitian Kousik Barik, Sanjay Misra .....                                                                                  | 26 |
| 2.2.9.                              | Penelitian Murahartawaty Arief, Noor Azah Samsudin .....                                                                     | 27 |
| 2.2.10.                             | Penelitian Efrico Fernando Baharsyah, Armanto, Tri H. Bimastara Aviani, Cindi Wilandari .....                                | 28 |
| 2.2.11.                             | Penelitian Rina Afriani Sitorus, Ilka Zufria .....                                                                           | 28 |
| 2.2.12.                             | Penelitian Antoni Carla, Hari Soetanto, Yuliazmi .....                                                                       | 29 |
| 2.2.13.                             | Penelitian Nur Adha Pasaribu, Sriani .....                                                                                   | 30 |
| 2.2.14.                             | Penelitian Syafrizal, M. Afdal, Rice Novita .....                                                                            | 30 |
| 2.2.15.                             | Penelitian M Ikhsan Maulana, Elvia Budianita, Muhammad Fikry, Febi Yanto .....                                               | 31 |
| BAB III METODOLOGI PENELITIAN ..... |                                                                                                                              | 32 |
| 3.1.                                | Analisa Kebutuhan .....                                                                                                      | 32 |
| 3.1.1.                              | <i>Requirement Elicitation</i> Tahapan I .....                                                                               | 33 |
| 3.1.2.                              | <i>Requirement Elicitation</i> Tahap II .....                                                                                | 34 |
| 3.1.3.                              | <i>Requirement Elicitation</i> Tahap III .....                                                                               | 35 |
| 3.1.4.                              | <i>Requirement Elicitation</i> Final .....                                                                                   | 37 |
| 3.2.                                | Pembahasan Metode Yang Digunakan .....                                                                                       | 38 |

|                                  |                                                                          |    |
|----------------------------------|--------------------------------------------------------------------------|----|
| 3.2.1.                           | Pengambilan Data.....                                                    | 38 |
| 3.2.2.                           | <i>Text Preprocessing</i> .....                                          | 40 |
| 3.2.3.                           | Pelabelan Otomatis <i>Lexicon Vader</i> .....                            | 44 |
| 3.2.4.                           | Klasifikasi Manual <i>Naïve Bayes</i> .....                              | 46 |
| 3.2.5.                           | Proses Klasifikasi Data <i>Training</i> Pada <i>Naïve Bayes</i> .....    | 50 |
| 3.2.6.                           | Proses Klasifikasi Data <i>Testing</i> Pada <i>Naïve Bayes</i> .....     | 56 |
| 3.2.7.                           | Klasifikasi Pada <i>Python</i> .....                                     | 59 |
| 3.3.                             | Pembahasan Algoritma Yang Digunakan .....                                | 59 |
| 3.4.                             | Kerangka Pemikiran .....                                                 | 61 |
| 3.5.                             | Perancangan <i>Flowchart</i> .....                                       | 62 |
| 3.5.1.                           | Rancangan <i>Flowchart Text Mining Data Review</i> Komentar Ruang Guru.. | 62 |
| 3.5.2.                           | Rancangan <i>Flowchart</i> Sentimen Analisis Komentar.....               | 63 |
| 3.6.                             | Perancangan Layar, Menu, <i>Database</i> .....                           | 64 |
| 3.6.1.                           | Perancangan <i>Database</i> .....                                        | 64 |
| 3.6.2.                           | Perancangan Layar .....                                                  | 65 |
| 3.7.                             | Jadwal Penelitian .....                                                  | 69 |
| BAB IV HASIL DAN PEMBAHASAN..... |                                                                          | 71 |
| 4.1.                             | Spesifikasi <i>Hardware</i> Dan <i>Software</i> .....                    | 71 |
| 4.1.1.                           | <i>Hardware</i> .....                                                    | 71 |
| 4.1.2.                           | <i>Software</i> .....                                                    | 71 |
| 4.2.                             | Tampilan Program .....                                                   | 72 |
| 4.2.1.                           | Tampilan Program Antarmuka <i>Dataset</i> .....                          | 72 |
| 4.2.2.                           | Tampilan Antarmuka Data Hasil <i>Preprocessing</i> .....                 | 73 |
| 4.2.3.                           | Tampilan Antarmuka Data Klasifikasi .....                                | 73 |
| 4.2.4.                           | Tampilan Antarmuka Data Grafik .....                                     | 74 |
| 4.2.5.                           | Tampilan Antarmuka Analisa Sentimen Komentar .....                       | 75 |
| 4.3.                             | Penerapan Metode dan Algoritma .....                                     | 76 |
| 4.4.                             | <i>Blackbox Testing</i> .....                                            | 85 |
| BAB V SIMPULAN DAN SARAN.....    |                                                                          | 90 |
| 5.1.                             | Simpulan.....                                                            | 90 |
| 5.2.                             | Saran .....                                                              | 91 |
| DAFTAR PUSTAKA .....             |                                                                          | 92 |
| DAFTAR RIWAYAT HIDUP .....       |                                                                          | 96 |
| LAMPIRAN .....                   |                                                                          | 97 |



## DAFTAR TABEL

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| Tabel 3. 1 <i>Requirement Elicitation</i> Tahap I .....                       | 33 |
| Tabel 3. 2 <i>Requirement Elicitation</i> Tahapan II .....                    | 34 |
| Tabel 3. 3 <i>Requirement Elicitation</i> Tahapan III.....                    | 35 |
| Tabel 3. 4 <i>Requirement Elicitation</i> Final .....                         | 37 |
| Tabel 3. 5 Proses Hasil <i>Cleaning</i> .....                                 | 40 |
| Tabel 3. 6 Proses Hasil <i>Case Folding</i> .....                             | 41 |
| Tabel 3. 9 Contoh Hasil <i>Proses Translator</i> .....                        | 44 |
| Tabel 3. 10 Contoh Hasil Data Preprocessing .....                             | 46 |
| Tabel 3. 11 Hasil Tokenizing .....                                            | 47 |
| Tabel 3. 12 Pembobotan Kata.....                                              | 48 |
| Tabel 3. 13 Perhitungan <i>Conditional Probability</i> .....                  | 54 |
| Tabel 3. 14 Klasifikasi Data <i>Testing</i> .....                             | 58 |
| Tabel 3. 15 Keterangan Tabel <i>Dataset</i> .....                             | 64 |
| Tabel 3. 16 Keterangan Tabel databersih ( <i>Proses Preprocessing</i> ) ..... | 64 |
| Tabel 3. 17 Keterangan tabel dataklasifikasi .....                            | 65 |
| Tabel 3. 18 Ganchart Penelitian.....                                          | 69 |
| Tabel 4. 1 Spesifikasi <i>Hardware</i> .....                                  | 71 |
| Tabel 4. 2 Spesifikasi <i>Software</i> .....                                  | 71 |
| Tabel 4. 3 Confusion Matrix 70:30.....                                        | 79 |
| Tabel 4. 4 Confusion Matrix 80:20.....                                        | 80 |
| Tabel 4. 5 Confusion Matrix 90:10.....                                        | 80 |
| Tabel 4. 6 Hasil Perhitungan Setiap Skenario Label Positif.....               | 83 |
| Tabel 4. 7 Hasil Perhitungan Setiap Skenario Label Negatif .....              | 83 |
| Tabel 4. 8 Hasil Perhitungan Setiap Skenario Label Netral .....               | 84 |
| Tabel 4. 9 Hasil Perhitungan Setiap Skenario Semua Label .....                | 84 |
| Tabel 4. 10 <i>Blackbox Functional Testing</i> .....                          | 85 |
| Tabel 4. 11 Uji coba Sentimen analisis.....                                   | 87 |

## DAFTAR GAMBAR

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| Gambar 2. 1 Rumus <i>Naïve Bayes</i> .....                              | 14 |
| Gambar 2. 2 Rumus <i>Teorema Bayes</i> Klasifikasi .....                | 16 |
| Gambar 2. 3 Rumus <i>Teorema Bayes</i> Klasifikasi .....                | 16 |
| Gambar 2. 4 Rumus <i>Teorema Bayes</i> Klasifikasi .....                | 17 |
| Gambar 2. 5 <i>confusion matrix</i> .....                               | 17 |
| <br>                                                                    |    |
| Gambar 3. 1 <i>library scraper</i> .....                                | 38 |
| Gambar 3. 2 Proses <i>Scraping</i> Data Ulasan.....                     | 38 |
| Gambar 3. 3 Hasil <i>Scraping</i> Data Ulasan.....                      | 39 |
| Gambar 3. 4 <i>Proses Cleaning</i> Data .....                           | 40 |
| Gambar 3. 5 Gambar Hasil Pelabelan Otomatis.....                        | 45 |
| Gambar 3. 6 Hasil Klasifikasi <i>Python</i> .....                       | 59 |
| Gambar 3. 7 Flowchart Pembahasan Algoritma.....                         | 60 |
| Gambar 3. 8 Kerangka Pemikiran .....                                    | 61 |
| Gambar 3. 9 Rancang <i>Flowchart Text Mining</i> .....                  | 62 |
| Gambar 3. 10 Rancangan <i>Flowchart</i> Sentimen Analisis Komentar..... | 63 |
| Gambar 3. 11 <i>Entity Relationship Diagram</i> .....                   | 64 |
| Gambar 3. 12 Rancang Tampilan Utama.....                                | 65 |
| Gambar 3. 13 Rancangan Tampilan Data Preprocessing & Pelabelan.....     | 67 |
| Gambar 3. 14 Rancangan Tampilan Klasifikasi <i>Naive Bayes</i> .....    | 67 |
| Gambar 3. 15 Rancangan Tampilan Sentimen Komentar .....                 | 68 |
| Gambar 3. 16 Rancangan Tampilan Grafik sentimen.....                    | 68 |
| Gambar 3. 17 Rancangan Tampilan Hasil Predict Sentimen Komentar.....    | 69 |
| <br>                                                                    |    |
| Gambar 4. 1 Tampilan Antarmuka <i>Dataset</i> .....                     | 72 |
| Gambar 4. 2 Tampilan Antarmuka Hasil Preprocessing.....                 | 73 |
| Gambar 4. 3 Tampilan Antarmuka Hasil Klasifikasi.....                   | 74 |
| Gambar 4. 4 Tampilan Antarmuka Data Grafik.....                         | 75 |
| Gambar 4. 5 Tampilan Analisa Sentimen Komentar .....                    | 75 |
| Gambar 4. 6 Tampilan Antarmuka Hasil Prediksi Sentimen.....             | 76 |
| Gambar 4. 7 Data Rasio 70:30 .....                                      | 77 |
| Gambar 4. 8 Data Rasio 80:20 .....                                      | 78 |
| Gambar 4. 9 Data Rasio 90:10 .....                                      | 79 |



## DAFTAR LAMPIRAN

|                                                                   |    |
|-------------------------------------------------------------------|----|
| Lampiran 1 : Kartu Bimbingan.....                                 | 97 |
| Lampiran 2 : Hasil Kuesioner <i>Requirement Elicitation</i> ..... | 98 |



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Penelitian

Pesatnya teknologi perkembangan dalam bidang pembelajaran, pada saat ini banyak menampakkan dukungannya yang dimanah strategi pembelajaran ke media *online* dimanah media *online* memberikan kenyamanan, mengetahui informasi terbaru dan cepat secara *online*, pemanfaatan media *online* dalam strategi pembelajaran bergantung pada dua aspek penting : metode mengajar dan media pembelajaran (Wulandari, D.P, Kurniawan, Ariyani, & Arif, 2024). Sekarang ini, banyak sekali *website* situs pembelajaran dan aplikasi *online* yang berkembang pesat di Indonesia, dari mulai akses layanan yang secara gratis hingga yang berbayar, dan salah satunya merupakan aplikasi ruang guru (Baharsyah, Armanto, Aviani, & Wulandari, 2024). Dan juga pembelajaran *online* ini menjadi sebuah alternatif dimanah pada masa era pandemi covid-19, yang dimanah aktivitas kegiatan pembelajaran tatap muka yang secara langsung dihentikan sementara waktu dan dilakukan kegiatan pembelajaran *online* secara daring yang dilakukan di rumah, sebagai bentuk untuk pencegahan penyebaran covid-19 (Firman & Rahayu, 2020).

Ruang guru merupakan sebuah perusahaan *startup* pada bidang bisnis teknologi yang terbesar di Indonesia yang bergerak pada layanan pembelajaran secara *online* yang dikenal sebagai aplikasi ruang guru. Dimanah aplikasi dengan pengguna sebanyak 22 juta pengguna di aplikasi ruang guru, dan juga mengawasi 300 ribu mentor pengajar yang menyediakan layanan yang lebih dari seratus bidang ilmu pelajaran. Selain itu, juga ada materi tambahan pembelajaran, platform media ruang kelas virtual, simulasi sistem ujian secara *online*, video pembahasan materi yang berbasis pelanggan dan pasar layanan yang disesuaikan (Giovani, Ardiansyah, Haryanti, Kurniawati, & Gata, 2020). Aplikasi dapat diakses oleh pengguna melalui *Google PlayStore* maupun di *App IOS*.

Setelah bermunculannya inovasi terhadap aplikasi pendidikan ini, masyarakat atau pengguna memperdebatkan manfaat dan kekurangan yang dimiliki aplikasi ruang guru, pada ulasan komentar *review* di *Google PlayStore*.

Setiap ulasan komentar *review* di *Google PlayStore* memiliki skala *rating* 1 sampai 5, namun *rating* yang diberikan oleh pengguna sering kali tidak sesuai dengan ulasan komentar *review*, sehingga tidak cukup untuk menggambarkan kelebihan dan kekurangan pada aplikasi ruang guru. Setiap penerapan *software* atau website memiliki titik keunggulan dan hambatannya, dimanah dapat menyebabkan ulasan yang berbeda-beda setiap pengguna, seperti memberikan ulasan kepuasan atau tidak kepuasan terhadap pengguna aplikasi tersebut.

Untuk penyelesaian studi ini menggunakan analisis sentimen atau *Opinion Mining*, analisis sentimen dipakai untuk menentukan informasi penting yang dapat diberikan pada data yang tidak terstruktur atau *unstruktur*, sehingga bertujuan pada penelitian ini dapat diketahui sentimen pengguna *Google PlayStore* terhadap aplikasi Ruang Guru dan juga data yang akan digunakan untuk analisis sentimen pada penelitian ini. Analisis sentimen adalah proses yang akan menentukan apakah opini yang disampaikan pada ulasan komentar *review* atau dalam sebuah kalimat bersifat positif, negatif atau netral dengan mengelompokkan polaritas teks.

Dengan adanya sistem *Opinion Mining* ini dapat diharapkan untuk membantu pengembangan aplikasi untuk melihat *rating* dan *review* dari suatu ulasan pengguna sebagai acuan dalam pengembangan dan dapat membuat pengambilan keputusan yang lebih baik. Sistem ini juga diharapkan juga berperan untuk mengetahui nilai emosi atau suasana hati yang digambarkan berupa label positif dan negatif suatu produk terhadap pengguna, melihat dari komentar *review* ulasan yang diberikan terhadap suatu produk digital, dan hasil dari

informasi data sentimen dapat digunakan juga sebagai untuk mengklasifikasi analisa sentimen komentar.

Berdasarkan uraian di atas, pemanfaatan *review* ulasan produk digital dalam *Google PlayStore* dapat dilakukan melalui analisis sentimen atau *Opinion Mining*, pada proses analisis untuk menggunakan sebuah metode pengklasifikasi yang merupakan teknik analisa dengan model klasifikasi pada penelitian sebelumnya sebagai berikut :

Pada penelitian (Giovani et al., 2020), pada penelitiannya menggunakan tiga algoritma klasifikasi teks, yaitu *Naïve Bayes* , *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN) sebagai analisa model klasifikasi sentimen dan *dataset* pada label data untuk analisa model tidak diketahui dalam pembahasan penelitian ini, dan hasil dari penelitian pengujian 3 model klasifikasi tersebut model SVM dengan nilai akurasi 78,55%, KNN nilai akurasinya 77,21% dan *Naïve Bayes* nilai akurasi 67,32%.

Pada penelitian (Syafrizal, Afdal, & Novita, 2023), berdasarkan penelitiannya menggunakan dua algoritma klasifikasi teks, yaitu *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN) sebagai perbandingannya analisa model klasifikasi sentimen dan *dataset* pada label data menggunakan pelabelan secara manual yang bantuan ahli pakar bahasa dalam pelabelan datanya untuk analisa model. Hasil dari penelitian pengujian 2 model klasifikasi tersebut model *Naïve Bayes Classifier* (NBC) memiliki performa lebih baik dibandingkan KNN dengan akurasi 77,69% untuk model NBC dan KNN nilai akurasi 76,4%.

Pada penelitian (Taufiqurrahman, Tri Anggraeny, & Muharrom Al Haromainy, 2024), berdasarkan penelitiannya bertujuan menganalisis dan membandingkan performa dua algoritma klasifikasi teks, yakni *Naïve Bayes* (NB) dan *K-Nearest Neighbor* (KNN) sebagai klasifikasi sentimen dan *dataset* pada label datanya menggunakan pelabelan *vader* untuk pelabelan data untuk digunakan analisa model. Hasil dari penelitian perbandingan 2 model

klasifikasi tersebut model *Naïve Bayes* (NB) memiliki performa lebih baik dibandingkan KNN dengan nilai akurasi 75% dan KNN nilai akurasi 73%.

Berdasarkan penelitian sebelumnya model klasifikasi *Naïve Bayes* (NB) menunjukkan memiliki model yang cukup baik dalam klasifikasi sentimen dan berdasarkan pembahasan yang telah didapatkan pada latar belakang maka, penulis menerapkan algoritma klasifikasi *Naïve Bayes* dalam melakukan analisis sentimen terhadap ulasan pengguna di *Google Play Store* dengan judul penelitian “ SENTIMEN ANALISIS REVIEW APLIKASI RUANG GURU PADA GOOGLE PLAY STORE DENGAN METODE NAÏVE BAYES”.

## 1.2 Identifikasi Masalah

Berdasarkan yang ada pada latar belakang tertulis, kami memberikan informasi berikut tentang masalah yang terdapat dan akan digunakan sebagai bahan penelitian :

1. Pada penelitian ini untuk membantu melihat informasi ulasan *review* aplikasi ruang guru, sehingga dapat melihat sisi opini pada pengguna tanpa melihat sekedar dari *rating* yang diberikan oleh pengguna namun juga melihat opini produk digital tersebut dapat diterima baik positif , atau negatif dari opini *review* ulasan komentar dari pengguna *Google PlayStore* ?
2. Dan bagaimana mengimplementasikan algoritma *Naïve Bayes* dan *Complement Naïve Bayes* untuk mengklasifikasi komentar pengguna aplikasi dengan indikasi sentimen komentar aplikasi ?

## 1.3 Ruang Lingkup Penelitian

Untuk ruang lingkup penelitian lebih terfokus, maka diterapkan sejumlah batasan dan beberapa aturan sebagai berikut :

- a. Sumber data berasal dari ulasan *Google PlayStore* ulasan produk aplikasi ruang guru pada tahun 2022 sampai 2024 dengan sebanyak 10000 data.

- b. Pelabelan data otomatis menggunakan metode *lexicon vader* untuk pelabelan otomatis polaritas positif, netral dan negatif yang nanti akan digunakan sebagai data *testing* dan *training*.
- c. Sentimen yang dianalisis menggunakan bahasa Indonesia.
- d. Hasil dari *text mining* berupa sentimen analisa algoritma *Naïve Bayes* berupa positif, netral dan negatif dari ulasan tersebut.
- e. Perancangan sistem informasi yang disebut dengan visual data menggunakan program PHP dengan metode MVC pada *laravel* , *database Mysql* dan *Rest API*.

#### **1.4 Tujuan dan Manfaat Penelitian**

##### **1.4.1. Tujuan**

Adapun tujuan dari pembuatan aplikasi ini adalah sebagai berikut :

- a. Mengatur persepsi pengguna terhadap aplikasi edukasi *online* secara spesifik untuk mengetahui bagaimana pengguna merespons fitur dan fungsionalitas aplikasi ruang guru.
- b. Melihat informasi pengembangan dalam memahami perubahan tren pengguna dari tahun 2022 hingga 2024 terkait pengalaman belajar *online* melalui ulasan.
- c. Mengklasifikasi ulasan komentar pengguna berdasarkan positif, negatif dan netral, untuk menganalisis pola komentar.
- d. Menguji *Naïve Bayes* dengan data komentar ulasan ruang guru

##### **1.4.2. Manfaat**

- a. Menyediakan informasi berbasis data bagi pengguna untuk melihat suatu opini dari suatu produk digital di terima baik positif, negatif atau netral.
- b. Membantu perusahaan ruang guru untuk memahami identifikasi potensi masalah teknis atau fitur yang memerlukan perbaikan berdasarkan sentimen negatif.



- c. Memudahkan proses klasifikasi komentar untuk sentimen analisis dengan pendekatan berbasis *machine learning*.
- d. Mengetahui kinerja model *Naïve Bayes* pada saat model digunakan.

## **1.5 Sistematika Penulisan**

### **BAB I PENDAHULUAN**

Bab ini memaparkan isi tentang latar belakang, identifikasi masalah, ruang lingkup penelitian, tujuan dan manfaat dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Informasi yang di paparkan oleh bab ini tentang teori penelitian yang bersumber dari referensi jurnal – jurnal yang terkait pada penyusunan laporan penelitian skripsi serta beberapa literatur yang berhubungan dengan penelitian ini.

### **BAB III METODOLOGI PENELITIAN**

Bab ini menyajikan uraian informasi mengenai analisa kebutuhan penggunaan, pembahasan terkait algoritma atau metode yang digunakan, perancangan basis data, serta rancangan model visualisasi *dashboard* dan menu.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini menjelaskan dan menguraikan spesifikasi perangkat keras dan perangkat lunak yang digunakan, model tampilan antarmuka aplikasi, proses pengujian sistem menggunakan metode *black box*, proses evaluasi model klasifikasi dan serta pengujian sistem.

### **BAB V KESIMPULAN DAN SARAN**

Berisikan informasi dari kesimpulan dan saran yang berkaitan dengan hasil analisis yang telah di uraikan pada bab – bab sebelumnya.



## BAB II

### LANDASAN TEORI

#### 2.1 Teori

##### 2.1.1. Data

Menurut Gunawan (Gunawan, Ningsih, & Lantana, 2023) “Data adalah sekumpulan fakta mentah, yang digambarkan seperti angka, objek, huruf yang belum memiliki arti atau konteks.”

Data adalah sebuah fakta atau kejadian namun belum memiliki suatu arti dan perlu pengolahan agar menghasilkan sesuatu yang berarti dan berguna. Pada penjelasan di atas, didapatkan kesimpulan bahwa data adalah suatu yang berisikan fakta atau kebenaran suatu kejadian atau objek yang dapat diterapkan dalam bentuk simbol, huruf, angka, maupun penggabungan dari sebuah bentuk yang menyerupai objek.

Data dibedakan menjadi 2 yaitu :

##### 1) Kualitatif

Menurut Mertha Jaya (Mertha Jaya, 2020) “Data kualitatif sederhananya bisa disebut sebagai data hasil kategori (pelabelan kode), untuk identifikasi sebagai data atau isi data yang berupa kata”. Misalnya : status jenis kelamin dan lain sebagainya. Dan data kualitatif memiliki khas ciri data yang tidak dapat melakukan operasi matematika.

##### 2) Kuantitatif

Menurut Mertha Jaya (Mertha Jaya, 2020) “Data kualitatif merupakan data yang isinya data nominal atau angka, dan data kualitatif sering digunakan dalam proses matematika pada sifatnya.”

Menurut Fenti Hikmawati (Hikmawati, 2020) Data juga dibedakan menjadi 2 jenis, dari sumber pengumpulannya yaitu :

1) Data Primer

Data primer adalah data yang langsung yang dikumpulkan oleh peneliti untuk menjawab masalah risetnya secara khusus. Dan data primer diambil langsung pada sumbernya atau narasumber secara langsung.

2) Data Sekunder

Data sekunder adalah data yang dikumpulkan tidak langsung atau data hasil kumpulan dari pihak lain (tangan kedua), peneliti biasanya mengambil secara mengakses, mencatat, meminta data tersebut ke pihak lain yang telah mengumpulkan data.

**2.1.2. Informasi**

Informasi adalah sebuah kumpulan data yang telah diolah menjadi sebuah bentuk yang berarti ataupun konteks bagi penerimanya dan juga bermanfaat untuk mengambil dalam kebijakan atau keputusan saat ini atau yang akan datang (Anjeli, Faulina, & Fakhri, 2022).

Informasi merupakan hasil proses analisis dan sintesis terhadap data, dengan kata lain informasi dapat dikatakan sebagai data yang telah diolah ke dalam bentuk yang disesuaikan dengan kebutuhan seseorang atau penerima (Setyawati, Sarwani, Wijoyo, & Soeharmoko, 2020).

Berdasarkan penjelasan di atas dapat diambil kesimpulan bahwa informasi adalah data yang telah diolah dan dibentuk sesuai kebutuhan penerima yang berfungsi untuk

mewakilkkan suatu pesan yang berisikan pengetahuan tentang sesuatu yang berarti bagi penerimanya.

### 2.1.3. Ruang Guru

Ruang Guru merupakan sebuah perusahaan terbesar dibidang teknologi di Indonesia yang di mana, perusahaan tersebut menyediakan sebuah layanan yang berbasis pendidikan yang secara *online*, di mana perusahaan teknologi tersebut mengelola lebih dari 22 juta pengguna serta mengelola 300 ribu guru dan menawarkan jasa di lebih dari 100 bidang pelajaran (Giovani et al., 2020).

Pada Ruang Guru mengembangkan berbagai layanan belajar yang berbasis secara *online*, yang berupa seperti layanan kelas virtual, platform ujian *online*, akses layanan berlanggan untuk video pembelajaran, dan les *private*. Dan adanya Ruang Guru memudahkan para pelajar untuk mengakses materi pembelajaran yang berkualitas, diakses dengan cepat dan mudah. Karena melalui aplikasi atau *website* dengan bermodalan akses internet (Giovani et al., 2020).

Pada penelitian ini Ruang Guru sebagai sumber data yang akan diambil berupa komentar atau ulasan *review* di *Google PlyaStore*, untuk mengetahui penilaian pengguna atau pembeli dari ulasan atau komentar pada suatu produk atau barang.

### 2.1.4. Pelabelan Data Otomatis (*Vader*)

*Vader* merupakan sebuah pendekatan leksikal yang dipakai sebagai model untuk analisis sebuah emosi dan suasana hati yang dapat digunakan untuk menilai dari berbagai data. C.J. Hutto dan Eric Gilbert dari *Georgia Institute of technology* merupakan yang menciptakan *Lexicon VADER (Valanced Aware Dictionary Sentiment Reasoner)* di tahun 2014 yang untuk membatu secara otomatis melabeli data (Asri, Sulyanti, Kuswardani, & Fajri, 2022).

Berdasarkan pendapat menurut Hutto dan Gilbert, fitur leksikal memiliki nilai rata – rata mendekati nol dan standar deviasi kurang dari 2,5. Ada lebih dari 7500 fitur leksikal yang memiliki skor valensi yang terverifikasi untuk menunjukkan polaritas sensori emosional (positif/negatif), setiap teks akan dicetak oleh *Vader*. Setiap kumpulan dari teks yang dihasilkan akan dievaluasi dan hasilnya dijumlahkan untuk membentuk nilai *compound*. *Compound* merupakan matriks yang menghitung dari skor yang dinormalisasi dari -1 hingga +1 (Asri et al., 2022).

#### **2.1.5. Analisis Sentimen**

Menurut Angelina Puput Giovani (Giovani et al., 2020) “ Analisis sentimen adalah analisis yang digunakan untuk menentukan atau mencari informasi yang berharga dari data tidak terstruktur (*unstructured*), untuk mengetahui penilaian pengguna atau pembeli dari ulasan atau komentar pada suatu produk atau barang.

Menurut Siti Masturoh (Siti Masturoh, 2020) “ Analisis sentimen merupakan analisis yang mempelajari tentang opini, perilaku dan emosi entitas. Dari proses memahami, mengekstraksi dan memproses teks data untuk menghasilkan informasi emosional yang terkandung dalam kalimat opini.

Menurut Lubis (Lubis, Bangsa, & Yudertha, 2024) “ Analisis sentimen adalah sebuah cabang ilmu teknologi pada data *mining* yang dipakai untuk menganalisis, mengolah dan memperoleh data teks dari data sebuah layanan, produk, individu ataupun topik tertentu untuk mendapatkan sebuah informasi dari sebuah himpunan data yang ada.

Pada penjelasan di atas dapat disimpulkan bahwa, analisis sentimen adalah analisis yang mempelajari, dan memproses teks data atau data *unstructured* untuk mendapatkan hasil informasi emosional dari kalimat opini atau komentar ulasan pengguna. Pada analisis sentimen ini terdapat bantuan analisis dari *machine learning* untuk metode

klasifikasinya seperti *Naïve Bayes*, *Lexicon Based*, *Support Vector Machine* dan *Logistic Regression*, yang digunakan untuk membantu mencari hasil yang terbaik.

Menurut Lubis(Lubis et al., 2024) Pada klasifikasinya analisis sentimen terdapat beberapa kelas dalam sentimennya, seperti berikut yaitu :

- 1) Sentimen positif : Menurut KBBI atau dikenal kamus Besar Bahasa Indonesia, sentimen positif adalah sebuah sikap atau reaksi yang meningkatkan sebuah nilai seseorang atau terhadap sesuatu.
- 2) Sentimen Negatif : Menurut KBBI atau dikenal Kamus Besar Bahasa Indonesia, sentimen negatif adalah sebuah sikap atau reaksi yang bisa membuat menurunkan sebuah nilai seseorang atau terhadap sesuatu, demikian sehingga membuat tren menjadi *down*. Kalimat sebuah sentimen negatif dicirikan dengan penggunaan kata negasi. Biasanya negasi dipakai mengarah sebuah pola dari suatu pertanyaan.
- 3) Sentimen Netral : Menurut KBBI atau dikenal Kamus Besar Bahasa Indonesia, bila sentimen negatif merupakan sebuah sikap yang tidak berpihak. Jadi kalimat sentimen netral ini tidak berpihak pada ekspresi kalimat yang memiliki sifat positif maupun negatif.

#### **2.1.6. Text Mining**

Menurut Firdaus (Firdaus et al., 2021) “ *Text Mining* adalah penambangan dan pengolahan data yang bertujuan untuk mencari nilai informasi yang tersembunyi baik dari sumber data informasi semi-struktur ataupun informasi data yang tidak struktur”.

Menurut Rio Christian (Christian & Fenriana, 2023) “ *Text Mining* merupakan sebuah teknik yang dipakai untuk mencari atau menggali sebuah informasi dari data berbentuk teks ataupun dokumen dengan bertujuan mengidentifikasi pola, tren, dan hubungan yang ada dalam pada data teks, pada tekniknya menggunakan eksplorasi dan

analisis yang pada hasilnya yang dapat digunakan untuk pengambilan keputusan yang lebih baik”.

Dari penjelasan di atas dapat disimpulkan *text mining* adalah proses menggali informasi dari sekumpulan data yang berbentuk dokumen teks atau data tidak struktur, yang secara khusus untuk mencari informasi yang ada pada data.

*Text mining* pada tujuannya, secara khusus dapat dibagi menjadi 2 yaitu:

1. Mengkategorikan data teks

Dalam mengkategorikan data teks, *text mining* dipergunakan untuk sebagai alat bantu untuk menemukan kategori yang sesuai dengan kelas yang ditentukan (*supervised learning*).

2. Mengelompokkan data teks

Dalam pengelompokannya, *text mining* yang berfungsi sebagai alat untuk digunakan sebagai mengelompokkan data teks berdasarkan karakteristik yang sama dan *clustering* memberikan kelas sebagai tanda untuk label pada data teks yang belum diketahui kelasnya (*unsupervised learning*).

### 2.1.7. **Python**

*Python* adalah bahasa pemrograman yang berbasis orientasi objek tingkat tinggi dan dibuat oleh Guido Van Rossum. Dan juga *python* merupakan bahasa lintas fungsi yang diartikan secara maksimal yang memiliki banyak fungsi, *python* dikenal sebagai bahasa pemrograman yang sangat mudah dipelajari baik bagi pemula maupun yang sudah menguasai bahasa pemrograman lain. Bahasa pemrograman *python* ini terkenal di kalangan para ilmuwan dan penganalisis data karena fleksibilitasnya, yang di mana memungkinkan untuk mempermudah dalam model data, mengestimasi kumpulan data, dan menerapkan penambahan data untuk penyelesaian tugas dengan waktu yang singkat (Junaidi, Devegi, & Kurniawan, 2023). *Python* juga dikenal sebagai salah satu



bahasa pemrograman yang paling populer sedunia karena memiliki aturan *coding sintaks* yang sederhana yang mudah dipelajari oleh pemula (Hermawan, Suwitno, Junaedi, Kurniawan, & Ayu Kumala, 2024).

Menurut Hermanto (Hermanto, Salim, Wu, Salim, & Gunadi, 2023) “*Python* memiliki berbagai *library* yang dipakai untuk analisa data, berikut beberapa *library* yang banyak digunakan di *python* :

1. *Pandas*

*Pandas* adalah *library* yang digunakan untuk analisa data dan manipulasi struktur data seperti data *frame* yang memungkinkan pengujian data dan pengolahan data secara efisien. Contoh data struktur yang ditampilkan berbentuk tabel atau visual lainnya.

2. *Numpy*

*Numpy* adalah *library* yang digunakan untuk data numerik pada komputasi dalam *python* dan juga *numpy* menyediakan multidimensi *array* dan operasi matematika yang cepat.

3. *Matplotlib*

*Matplotlib* merupakan *library* yang dipakai untuk menampilkan visualisasi dari data, seperti untuk membuat grafik, plot dan visualisasi lainnya. Di mana *matplotlib* ini berfungsi sebagai membantu memahami konteks data dengan tampilan visualisasi.

#### 2.1.8. **Klasifikasi**

Menurut sunata (Sunata, Hariyanto, Dwi Putra, & Wijaya, 2022) “Klasifikasi merupakan sebuah kegiatan yang digunakan untuk proses mengelompokkan data dengan mengacu pada karakteristik tertentu, pada proses pengelompokan di dalam data *mining*”.



Proses klasifikasi berfungsi sebagai untuk mendapatkan informasi membantu dalam membuat keputusan ataupun hasil prediksi kelompok (Barus, 2022).

Pada penjelasan di atas dapat disimpulkan klasifikasi merupakan suatu kegiatan yang dipakai untuk mengelompokkan sebuah data dengan mengacu pada karakteristik tertentu dalam data *mining* ataupun *text mining*.

#### 2.1.9. *Naïve Bayes*

*Naïve Bayes* adalah algoritma yang sering dipakai dengan teknik data *mining* dan juga merupakan teori *Bayes*, algoritma *Naïve Bayes* ini juga sebagai strategi *classifier* probabilitas yang dapat menentukan dengan pertambahan suatu nilai iterasi dari sebuah indeks informasi yang sudah ada dan diberikan, dengan mengasumsikan seluruh atributnya sehingga mereka tidak saling ketergantungan dari nilai yang memiliki variabel kelas (Ramadhani & Wahyudin, 2022).

Penggunaan metode *Naïve Bayes* dapat dirumuskan dengan persamaan umum sebagai berikut :

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

Gambar 2. 1 Rumus *Naïve Bayes*

(Sumber :Christian & Fenriana, 2023)

Keterangan :

X : Data dengan kelas yang belum diketahui.

H : Hipotesis data pada suatu kelas spesifik.

$P(H|X)$  : Probabilitas hipotesis dari kondisi (*posteriori probability*).

$P(H)$  : Probabilitas hipotesis awalan kelas (*prior probability*).

$P(X|H)$  : Probabilitas hipotesis kondisi berdasarkan hipotesis kelasnya(probabilitas *likelihood*).

$P(X)$  : Probabilitas keseluruhan data X (probabilitas *evidence*).

Pada aturan *Bayes* dari ide dasarnya dapat menghasilkan hipotesis atau peristiwa (H) yang dapat diamati dari beberapa bukti (X) dari awal yang diperkirakan. Dan juga ada beberapa hal yang harus diperhatikan penting dalam aturan *Bayes*, sebagai berikut :

1. Sebuah probabilitas awal atau  $P(H)$  merupakan dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau  $P(H|X)$  merupakan dari suatu hipotesis setelah bukti diamati.

Pada situasi yang tepat dan bergantung pada model probabilitas, *Naïve Bayes* dapat berguna sangat efektif dalam *supervised learning*, kelebihan dari *Naïve Bayes* adalah algoritma ini pada metodenya hanya membutuhkan data latih untuk memprediksikan estimasi parameter yang diperlukan untuk proses klasifikasi. Sebagai variabel independen yang diasumsikan, hanya varian dari variabel dalam suatu kelas yang dibutuhkan untuk menentukan pengklasifikasi, bukan dari matriks *konvarias*.

#### **2.1.10. *Naïve Bayes* Untuk Klasifikasi**

Pada kaitan antara *Naïve Bayes* dengan klasifikasi terdapat dua tahapan, di mana untuk tahapan pertama dilakukan yaitu, proses tahapan pembelajaran pada sebuah teks dokumen yang kategorinya sudah diketahui. Pada tahapan kedua yaitu proses pengujian pada klasifikasi teks dokumen yang kategorinya tidak diketahui. *Naïve Bayes Classifier* bekerja lebih unggul dibandingkan model *classifier* lainnya, dan juga algoritma *Naïve Bayes classifier* pada tingkat akurasi lebih baik dari pada model *classifier* lainnya.

Pada formulasinya *Naïve Bayes* untuk klasifikasi algoritma akan mencari probabilitas tinggi data dokumen dari semua kategori yang diuji ( $V_{map}$ ). Dengan diterapkan *teorema Bayes* pada persamaan berikut :

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(X_1, X_2, X_3, \dots, X_n | V_j)}{P(X_1, X_2, X_3, \dots, X_n)}$$

Gambar 2. 2 Rumus Teorema Bayes Klasifikasi

(sumber :Duei Putri, Nama, & Sulistiono, 2022)

Untuk  $P(x_1, x_2, x_3, \dots, x_n)$  nilai konstan pada semua kategori (Duei Putri, Nama, & Sulistiono, 2022). Pada gambar persamaan di atas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j)$$

Gambar 2. 3 Rumus Teorema Bayes Klasifikasi

(sumber :Duei Putri, Nama, & Sulistiono, 2022)

Keterangan :

$V_j$  = Kategori ulasan atau dokumen,  $j=1,2,3,\dots,n$ . diartikan sebagai

$j_1$  = Kategori sentimen negatif,

$j_2$  = Kategori sentimen positif, dan

$j_3$  = Kategori sentimen netral.

$P(x_i | V_j)$  = Probabilitas  $x_i$  pada kategori.

$V_j P(V_j)$  = Probabilitas dari  $V_j$

Untuk  $P(v_j)$  dan  $P(x_i | V_j)$  pada perhitungan di saat pelatihan dirumuskan pada persamaan sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|conto|}$$

$$P(X_i|V_j) = \frac{nk+1}{n+|kosakata|}$$

Gambar 2. 4 Rumus Teorema Bayes Klasifikasi

(sumber: Duei Putri et al., 2022)

Keterangan :

|docs j| = jumlah ulasan atau dokumen setiap kategori j.

|conto| = jumlah ulasan atau dokumen dari semua kategori.

nk = jumlah frekuensi dari setiap kemunculan kata.

n = jumlah frekuensi setiap kategori dari kemunculan kata.

#### 2.1.11. **Confusion Matrix**

*Confusion Matrix* adalah suatu metode yang dipakai untuk mengetahui ukuran atau melakukan perhitungan akurasi pada konsep data mining, *confusion matrix* berfungsi untuk memperoleh nilai *precision*, *recall*, dan *accuracy*. Hasil nilai *confusion matrix* biasanya dilambangkan dalam satuan persen (%) (Rahayu, Prianto, & Novia, 2021). Pada gambar berikut adalah sebuah *confusion matrix* pada analisis model pengklasifikasi pada dua kelas :

|                            | Predicted Positives                      | Predicted Negatives                      |
|----------------------------|------------------------------------------|------------------------------------------|
| Actual Positives instances | Number of True Positives instances (TP)  | Number of False Negatives instances (FN) |
| Actual Negatives instances | Number of False Positives instances (FP) | Number of True Negatives instances (TN)  |

Gambar 2. 5 *confusion matrix*

(sumber : Rahayu et al., 2021)

Pada Gambar di atas berikut keterangannya :

1. *True Positives* (TP) adalah jumlah data dari kelas positif yang dihasilkan dari klasifikasi menjadi nilai positif.
2. *False positives* (FP) adalah jumlah data dari kelas negatif yang dihasilkan dari klasifikasi menjadi nilai positif.
3. *False Negatives* (FN) adalah jumlah data dari kelas positif yang dihasilkan dari klasifikasi menjadi nilai negatif.
4. *True Negatives* (TN) adalah jumlah data dari kelas negatif yang dihasilkan dari klasifikasi menjadi nilai negatif.

Pada penjelasan di atas merupakan nilai yang akan dihasilkan pada metode *confusion matrix* adalah berupa evaluasi sebagai berikut :

1. *Accuracy*

Merupakan sebagai persentase jumlah data yang diprediksikan pada klasifikasi secara benar oleh algoritma.

$$\text{Rumus : } \textit{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

2. *Precision*

Merupakan sebuah nilai dari ketepatan dari analisa model yang digunakan dalam klasifikasi, dimana menunjukkan banyaknya data yang dapat terklasifikasi di kelas ataupun kategori yang benar dalam sebuah pengujian.

$$\text{Rumus : } \textit{precision} = TP/(TP+FP)$$

3. *Recall*

Merupakan sebuah nilai yang mengukur hasil berapa persentase data yang terklasifikasi dengan benar pada kelasnya.

Rumus :  $Recall = TP/(TP+FN)$

#### 2.1.12. *Website*

*Website* merupakan salah satu sebuah aplikasi yang terdapat berisikan dokumen – dokumen multimedia (visual grafik, gambar, teks, video) di mana menggunakan protokol HTTP (*hyperteks transfer protocol*) dan untuk mengakses *website* dengan bantuan perangkat lunak yang disebut browser (Nurlailah & Nova Wardani, 2023) .

Menurut Sonny (Sonny, Sonny, 2021) “*Website* adalah sebuah media yang terdiri dari beberapa halaman yang memiliki saling kaitan satu sama lain, dan berfungsi untuk media yang menampilkan suatu informasi, seperti gambar, video, visual grafik, teks, ataupun gabungan dari semuanya.

#### 2.1.13. *Laravel*

Menurut Aipina (Aipina & Witriyono, 2022) “*Laravel* merupakan sebuah salah satu *framework* yang berbasis open *source* yang dapat membantu developer untuk membuat pengembangan *website*”.

*Laravel* adalah sebuah *framework* web yang berbasis pemrograman PHP dan dikembangkan secara open *source*, *laravel* dikembangkan oleh Taylor Otwell dan *framework* web *laravel* menerapkan sebuah pola yaitu MVC. Di mana struktur MVC yang diterapkan di *laravel* memiliki perbedaan dari MVC yang pada umumnya, pada *laravel* memiliki fitur *routing* yang sebagai menghubungkan antar *request user*, dan sebuah *controller* yang menerimanya. Jadi pada pemodelannya sebuah *controller* tidak bisa secara langsung dapat menerima sebuah *request* tertentu.

Selain dari pemodelan *framework* web *laravel*, *laravel* juga memiliki *dependency manager* yang berfungsi untuk menyimpan *library* yang di *instal*, menggunakan *composer* dan akan disimpan di *direktori* vendor. *Laravel* banyak mendukung dalam



menggunakan pengelolaan datanya di mana *laravel* sangat memudahkan dalam menggunakan pengelolaan datanya baik melalui *database* ataupun *rest API*.

#### 2.1.14. *Mysql*

*Mysql* merupakan sebuah jenis *database* server yang menggunakan bahasa dasar SQL untuk mengakses dan menjalankannya perintah di *database*, *mysql* sendiri termasuk jenis RDBMS (*Relation Database Management System*). Selain itu *mysql* merupakan sebuah aplikasi *software* yang *open source* dan memiliki kinerja sangat cepat, *reliable*, dan mudah dipakai untuk *client server* (Sonny, Sonny, 2021).

Menurut Anjeli (Anjeli et al., 2022) “*Mysql* adalah salah satu aplikasi DBMS (*Database Managemet System*), yang memungkinkan seorang *user* untuk mengelola data seperti mengubah, menghapus, dan menambahkan data yang terdapat pada *database*. *Mysql* juga merupakan aplikasi yang terpopuler para kalangan *programmer* web karena aplikasi yang *open source* dan praktis dalam penggunaannya.

Pada penjelasan teori di atas dapat disimpulkan *mysql* adalah sebuah aplikasi jenis *database* server yang menggunakan bahasa dasar SQL untuk mengelola data pada *database*.

#### 2.1.15. *Rest API*

API atau *Application Program Interface* adalah suatu pengolahan maupun sebuah perintah dan data yang terdiri berbagai fungsi, kelas, antar muka (*interface*), kumpulan struktur serta berbagai protokol untuk merancang, membangun, serta implementasi perangkat lunak (Putra & Susetyo, 2020).

Menurut Fachriza Dian Adhiatma (Adhiatma & Qoiriah, 2022) “API atau REST API merupakan sebuah satu tipe web server yang memungkinkan *clinet* untuk mengakses sumber daya yang dimiliki oleh suatu sistem, sederhananya API dapat memberikan

sebuah data dan fungsinya untuk memfasilitasi interaksi antar program komputer dan juga dapat untuk saling bertukar informasi.”

Rest api memiliki aspek dari HTTP (*Hypertext Transfer Protocol*), di mana termasuk dalam metode *request*, kode *response*, dan *header* pesan. Pada prosesnya *client* dapat menentukan metode yang diinginkan untuk berinteraksi pada API melalui *request* HTTP seperti GET, PUT, dan POST (Adhiatma & Qoiriah, 2022).

## 2.2 Tinjauan Studi

### 2.2.1. Penelitian Angelina P. Giovani, Tuti Haryanti dan Windu Gata

Penelitian Pada tahun 2020 yang dilakukan oleh Angelina P. Giovani, Ardiansyah, Tuti Haryanti, Windu Gata (Giovani et al., 2020) yang berjudul “*Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi*” bertujuan untuk analisis sentimen untuk mendapatkan algoritma yang efektif.

Untuk menganalisa sentimen pengguna *twitter*, dengan menggunakan data komentar pengguna twitter dengan kata kunci Ruang Guru, yang mendapatkan 513 *tweet* berbahasa indonesia, setelah dilakukan proses pembersihan, hasil pembersihan data mendapatkan 338 *tweet* positif dan 175 *tweet* negatif. Penelitian ini memanfaatkan 3 algoritma klasifikasi, yaitu Naive Bayes (NB), Support Vector Machine (SVM), dan K-Nearest Neighbor (KNN), juga dengan perbandingan dengan tanpa *feature selection* menggunakan algoritma *Particle Swarm Optimization* (PSO) ataupun sebaliknya. Evaluasi pada kinerja algoritma menggunakan parameter *Area Under Curve* (AUC) dan akurasi untuk menentukan dari algoritma tersebut manakah yang terbaik.

Hasil penelitian menemukan bahwa tanpa PSO, algoritma SVM mendapatkan nilai performa terbaik dengan akurasi 76,93% dan AUC, diikuti dengan K-NN dengan akurasi 70,42% dan AUC 0,787, sedangkan NB mendapatkan hasil dengan 65,40% dan AUC 0,451. Dengan menggunakan PSO, nilai performa semua algoritma meningkat,

terutama SVM mencapai akurasi 78,55% dan AUC 0,853. Algoritma KNN mendapatkan peningkatan dengan akurasi 77,21% dan AUC 0,853 , sementara NB mendapatkan akurasi dengan 67,32% dan AUC 0,463.

Pada Penelitian ini menyimpulkan bahwa SVM dengan PSO adalah pendekatan paling optimal untuk analisis sentimen terkait komentar pengguna *twitter* pada aplikasi ruang guru. Keuntungan pada penelitian ini mengukur kinerja model menggunakan akurasi dan *Area Under Curve* (AUC) memberikan pandangan yang algoritma klasifikasi yang efektif. Kelemahan Data dikumpulkan tidak terlalu banyak dan rentang waktu sehingga memungkinkan tidak menggambarkan perubahan sentimen pengguna dalam jangka waktu panjang.

#### **2.2.2. Penelitian M. Yusuf Rismanda, Iqbal Maulana, Oman Kamarudin**

Penelitian pada tahun 2023 yang dilakukan oleh M. Yusuf Rismanda , Iqbal Maulana, Oman Kamarudin (Yusuf Rismanda Gaja, Maulana, & Komarudin, 2024) yang berjudul “ *Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan PlayStore Menggunakan Algoritma Naive Bayes*” , yang bertujuan untuk melihat reaksi sentimen pengguna playstore pada ulasan aplikasi Vidio.

Pada Penelitian ini data diambil dengan teknik web scraping, yang memperoleh data sebanyak 1259 ulasan bahasa indonesia dan melalui proses data cleaning. Pada penelitian ini menggunakan dua metode proses pelabelan data yaitu dengan kamus inset dan Vader, untuk mengklasifikasi sentimen menjadi positif dan negatif. Proses data melakukan transformasi data yang dilakukan dengan menggunakan TF-IDF untuk mengubah teks menjadi data numerik yang relevan untuk analisis dengan menggunakan algoritma naive bayes.

Hasil analisis menunjukan bahwa pelabelan dengan Inset lebih optimal dari Vader, baik dari segi akurasi, presisi, maupun recall. Akurasi tertinggi 76 % pada Inset pada

skema data latih dan uji 90:10, sedangkan vader hanya mencapai 73% pada skema 70:30. Selain itu, presisi juga didapatkan pada pelabelan Inset sebesar 73% sementara Vader hanya mencapai 72% pada skema yang sama pada skema 70:30. Recall tertinggi menggunakan pelabelan Inset mencapai 81 % pada skema 60:40%, yang mengungguli Vader yang hanya mencapai 79% pada skema 70:30.

Pada analisisnya dapat disimpulkan hasil pelabelan Inset lebih unggul dari pelabelan Vader untuk analisis sentimen. Keuntungan efisiensi proses pelabelan data yang lebih optimal untuk analisis sentimen teks berbahasa Indonesia. Kelemahan penelitian ini terbatas pada algoritma naive bayes, sehingga belum mencakup eksplorasi algoritma lain yang berpotensi meningkatkan performa klasifikasi.

### **2.2.3. Penelitian Rio Christian, Indah Fenriana**

Penelitian pada tahun 2023 yang dilakukan oleh Rio Christian dan Indah Feriana (Christian & Fenriana, 2023) yang berjudul “*Perancangan Aplikasi Analisis Sentimen Pada Ulasan Maskapai Penerbangan di Tripadvisor Menggunakan Metode Naive Bayes*”. Pada penelitiannya bertujuan untuk membantu masyarakat dalam memahami sentimen ulasan pada maskapai di Tripadvisor.

Pada pembahasannya jurnal ini membahas mengenai perancangan aplikasi analisis sentimen berbasis metode *Naive Bayes* dipakai untuk mengklasifikasi ulasan maskapai penerbangan di Tripadvisor, dengan 2 kelas klasifikasi yaitu positif dan negatif. Proses analisis dengan metode *web scraping*, *preprocessing* teks, klasifikasi, serta evaluasi model menggunakan *confusion matrix* dengan tiga rasio perbandingan data (70:30, 80:20, 90:10).

Hasil dari evaluasi *confusion matrix* menunjukkan bahwa rasio 90:10 memberikan hasil terbaik dengan akurasi 77%, presisi 77%, *recall* 100%, dan F1-score 87,1%. Kelebihannya, bahwa *Naive Bayes* dapat digunakan untuk analisis sentimen

ulasan dengan akurasi yang cukup baik. Kekurangannya, Tidak mempertimbangkan aspek bahasa seperti opini netral dalam analisis sentimennya.

#### **2.2.4. Penelitian Alqis Rausanfita, Arip Ramadan, M Dzulfikar Fauzi, Qori Putri Mafidah, Emilia Ramona, Yudha Mahardika Putra**

Penelitian pada tahun 2023 oleh Alqis Rausanfita, Arip Ramadan, M Dzulfikar Fauzi, Qori Putri Mafidah, Emilia Ramona, Yudha Mahardika Putra (Alqis Rausanfita et al., 2023) yang berjudul “ *Sentiment Analysis Of Comments on Instagram Posts of Indonesia’s 2024 Presidential Candidates Using The Support Vector Machine Method*” yang bertujuan untuk menilai dari reaksi pengguna instagram terhadap calon presiden indonesia 2024 .

Pada proses penelitian digunakan sebuah metode Support Vector Machine (SVM). Dimana analisis sentimen mengklasifikasikan menjadi 3 kelas yaitu positif, negatif dan netral. Tahapan proses melakukan crawling data komentar, preprocessing, pembobotan menggunakan TF-IDF, dan klasifikasi menggunakan SVM. Capaian dari hasil model SVM mencapai tingkat akurasi sebesar 89,41% dengan recall 89% dan presisi 87%.

Bahwa kesimpulan penelitian menunjukkan bahwa SVM memberikan performa terbaik untuk klasifikasi komentar, namun akurasi yang tinggi terutama di didominasi data berlabel positif 97,79% sedangkan akurasi kelas netral dan negatif cukup rendah dibawah 30%.

#### **2.2.5. Penelitian Roopam Srivastaya, P.K Bharti, Parul Verma**

Pada tahun 2022 penelitian yang dilakukan oleh Roopam Srivastava, P.K Bharti, Parul Verma (Srivastava, Bharti, & Verma, 2022). Judul penelitian yaitu “*Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis*” yang bertujuan untuk membahas perbandingan antar metode *machine learning* (ML) dan *lexicon-based* untuk analisis sentimen.



Pada proses penelitiannya menggunakan data ulan hotel dari Trip Advisor dengan data sebanyak 20.000 data. Dengan melakukan proses pembersihan data ,preprocessing dan tokenisasi. Evaluasi pada hasil perbandingan metode bahwa metode machine learning dengan SVM pada fitur TF-IDF mencapai akurasi tertinggi sebesar 96,3% , dan hal lainnya dengan VADER lexicon mencetak akurasi terbaik sebesar 88,7%.

Hasil menunjukan bahwa pendekatan berbasis *machine learning* (SVM dengan TF-IDF) memberikan kinerja terbaik untuk analisis sentimen, namun model berbasis lexicon seperti Vader memberikan relevan untuk kasus tanpa data berlabel.

#### **2.2.6. Penelitian Mohamed Chiny, Marouane Chihab, Omar Bencharf, Younes Chihab**

Penelitian dilakukan pada tahun 2021 oleh Mohamed Chiny, Marouane Chihab, Omar Bencharf, dan Younes Chihab, dengan judul penelitian “*LSTM, Vader and TF-IDF based Hybrid Sentiment Analysis Model*”. Tujuan dari penelitian ini untuk meningkatkan akurasi dan efisiensi dalam menganalisis sentimen dari data teks tidak terstruktur dari ulasan atau komentar *tweet*.

Penelitian ini menggunakan model *hybrid* yang menggabungkan beberapa model individu untuk mengatasi keterbatasan analisis sentimen. Model ini berjalan dalam beberapa tahap , termasuk *preprocessing* teks, pelatihan model (LSTM,VADER,TF-IDF) dan evaluasi skor pada ketiga model dalam klasifikasi. Klasifikasi ini dilakukan menggunakan 5 algoritma yaitu *logistic Regression, K-NN,Random Forest, SVM* dan *NaiveBayes*, dan menggunakan data dari IMDB dan Twitter.

Pada analisis model *hybrid* berhasil mencapai akurasi rata-rata lebih tinggi (+9,51%) dan F1-score (+7,36%) pada saat menggunakan data dari IMDB dibandingkan dengan model individu. Dan pada data set Twitter model *hybrid* juga menunjukkan peningkatan kinerja hingga 11,11% dalam akurasi dan 3,73% dalam F1 *score* pada algoritma K-NN.

Bisa disimpulkan bahwa model *hybrid* mampu meningkatkan akurasi analisis sentimen secara signifikan ini menjadikan bahwa lebih efektif untuk sentimen analisis data.

#### **2.2.7. Penelitian Subhasis Sanyal, Mobit Kumar Barai**

Penelitian dilakukan pada tahun 2021 oleh Subhasis Sanyal dan Mohit Kumar Barai (Sanyal & Barai, 2021), dengan judul penelitian “*Comparative study on Lexicon-Based Sentiment Analysers Over Negative Sentiment*”. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa dua algoritma berbasis lexicon vader dan textblob dalam mengidentifikasi dan klasifikasi sentimen negatif pada ulasan produk di *e-commerce*.

Pada penelitian ini menggunakan sumber data dari ulasan *amazon* dari produk ponsel dengan jumlah data 66.000 data. Dengan teknik teknik *crowdsourcing* untuk mengumpulkan data, dan melakukan *preprocessing* pada data untuk melakukan klasifikasi untuk menggunakan algoritma. Pada hasilnya menunjukkan bahwa *Vader* lebih unggul dibandingkan *TextBlob* dalam hal akurasi dan f1-score untuk mendeteksi sentimen negatif dengan skor sebesar 0,80 dan 0,56. Keunggulan *VADER* sebagian besar terletak pada kemampuan mempertimbangkan konteks linguistik.

#### **2.2.8. Penelitian Kousik Barik, Sanjay Misra**

Pada tahun 2024 penelitian yang dilakukan oleh Kousik Barik dan Sanjay Misra (Barik & Misra, 2024), yang berjudul “*Analysis of Customer Reviews with an Improved Vader Lexicon Classifier*”. Penelitian ini bertujuan untuk memahami sentimen pelanggan terhadap produk atau layanan yang berikan pada platform *e-commerce*.

Pada proses penelitian ini menjadi tantangan utama dalam analisis sentimen adalah menangkap konteks gagasan dalam suatu kalimat memiliki konotasi kata yang positif, netral dan negatif pada satu domain yang sama, jadi untuk membantu dalam penelitian ini memakai *IVADER* sebuah model algoritma *vader* standar dengan menambahkan



kamus dalam pengklasifikasinya. Model ini diuji menggunakan *dataset* platform *e-commerce* dari 4 kategori produk seperti elektronik, buku, *dvd*, dan dapur dengan 8.000 ulasan data pada setiap kategori.

Setelah dilakukan pembersihan data, *preprocessing*, dan TF-IDF, dilakukan klasifikasi sentimen dengan *lexicon* IVADER dengan hasil kinerja model sebesar akurasi 98,64%, presisi 97% ,*recall* 94% dan F1-score 92%. Jadi disimpulkan model IVADER menunjukkan memiliki performa tinggi dalam membedakan sentimen positif dan negatif dalam sentimen analisis.

#### **2.2.9. Penelitian Murahartawaty Arief, Noor Azah Samsudin**

Pada tahun 2023 penelitian yang dilakukan oleh Murahartawaty Arief dan Noor Azah Samsudin (Arief & Samsudin, 2023), yang berjudul “*Hybrid Approach with Vader and Multinomial Logistic Regression for Multiclass Sentiment Analysis in Online Customer Review*”. Penelitian ini bertujuan untuk menganalisis efektivitas gabungan algoritma *lexicon-based* VADER dan model deep learning MLR dalam klasifikasi sentimen menjadi positif , netral dan negatif, dengan menggunakan *dataset* ulasan produk *amazon*.

Pada penelitiannya dengan algoritma VADER mengklasifikasikan dengan polaritas *sentiment* berdasarkan intensitas skor kata, sementara MLR digunakan untuk meningkatkan akurasi klasifikasi. Dataset diolah dengan teknik TF-IDF untuk ekstraksi fitur dalam evaluasi. Hasil dari evaluasi menyatakan akurasi sebesar 92,78%.

Keuntungannya bahwa hasil dari model penggabungan VADER dan MLR mencapai akurasi yang tinggi dibandingkan dengan menggunakan model tunggal VADER.

Kelemahannya pengaruhnya *overfitting* pada rasio data pada data *train* dan data uji coba yang mempengaruhi performa pada model gabungan.

#### **2.2.10. Penelitian Efrico Fernando Baharsyah, Armanto, Tri H. Bimastara Aviani, Cindi Wilandari**

Pada tahun 2024 penelitian yang dilakukan oleh Efrico Fernando Baharsyah, Armanto, Tri Hasanah Bimastara Aviani dan Cindi Wulandari (Baharsyah et al., 2024), yang berjudul “*Analisis Sentimen Pengguna Terhadap Aplikasi Belajar Online Ruang Guru Pada Ulasan Google Play Store Menggunakan Algoritma Naive Bayes dan Support Vector Machine*”. Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma *Naive Bayes* dan SVM dalam mengklasifikasi sentimen analisis.

Pada proses penelitiannya data yang dipakai sebanyak 10.000 data ulasan dari tahun 2020- 2023, pada proses pembersihan data menyisakan 9.580 yang terdiri dari data positif 7.504 data dan data negatif sebanyak 2.076 data negatif. Yang melalui proses pengambilan data dengan bantuan *library python*, *preprocessing* dan pelabelan otomatis yang berdasarkan dengan *rating* (1-3 sebagai negatif dan 4-5 sebagai positif).

Pada hasil penelitiannya dengan uji coba banding *ratio* 90:10 pada algoritma NB dan SVM didapatkan bahwa algoritma SVM lebih unggul dikarenakan akurasi yang didapatkan yaitu 89%, sedangkan NB mendapatkan 84,86%. Keuntungan melakukan *preprocessing* data yang efektif, menghasilkan data yang berkualitas untuk klasifikasi. Kelemahan Model SVM membutuhkan lebih banyak waktu dalam proses dan sumber daya dibandingkan dengan algoritma *Naive Bayes*.

#### **2.2.11. Penelitian Rina Afriani Sitorus, Ilka Zufria**

Pada tahun 2024 penelitian yang dilakukan oleh Rina Afriani dan Ilka Zufria (Sitorus & Zufria, 2024) dengan berjudul “*Application of the Naive Bayes Algorithm in Sentiment Analysis of Using the Shopee Application in the Play Store*”. Penelitian ini bertujuan untuk di analisa sentimen pada produk aplikasi *e- commerce* terhadap ulasan di *playstore*.

Pada proses penelitiannya dengan menggunakan algoritma *naive bayes* untuk mengklasifikasi data sentimen, pada penelitian ini menggunakan data sebanyak 2000 ulasan pengguna shopee di play store, melakukan scraping, proses preprocessing data, TF-IDF dan pelabelan data yang diambil dari *rating* yang diberi pengguna dalam ulasan di mana bila *rating* 1-2 itu diberi label negatif, *rating* 3 dengan label netral dan *rating* 4-5 dengan label positif. Hingga menggunakan klasifikasi dengan algoritma *naivebayes*. Hasil evaluasi yang dihasilkan akurasi 71%, presisi 74%, *recall* 87% dan F1-score sebesar 80%.

Hasil penelitian menunjukkan bahwa sebagian besar ulasan pengguna bersifat negatif. Keuntungan penelitian membantu *shopee* memperbaiki fitur dan mengatasi keluhan pengguna. Kelemahan karena ketimpangan jumlah data yang tidak seimbang pada *labeling* dapat mempengaruhi hasil klasifikasi.

#### **2.2.12. Penelitian Antoni Carla, Hari Soetanto, Yuliazmi**

Pada tahun 2024 penelitian yang dilakukan oleh Antoni Carla, Hari Soetanto, dan Yuliazmi (Carla & Soetanto, 2024), yang berjudul “Implementasi Text Mining untuk Analisis Sentimen Pada Pengguna PLN Mobile Menggunakan Metode Naive Bayes”. Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi PLN *mobile* dengan menggunakan algoritma *naive bayes*.

Pada penelitian ini mengumpulkan 2000 data ulasan dari Google Play Store, pada proses data melalui tahapan *scraping*, *pre-pemrosesan*, pembobotan TF-IDF, hingga klasifikasi menggunakan algoritma *naive bayes*. Pada data ini di *labeling* dengan otomatis dengan mengkategorikan dengan *rating* yang pada pengguna pada pemberian ulasan.

Hasil penelitian ini dengan memberikan evaluasi menggunakan *confusion matrix* menunjukkan akurasi 76%, presisi 76%, *recall* 100% dan F1-score 86%. Keuntungan

pada penelitian ini menawarkan wawasan bagi PLN untuk meningkatkan kualitas aplikasi berdasarkan ulasan pengguna. Kelemahan Hasil evaluasi algoritma tidak menunjukkan performa yang lebih baik menunjukkan ruang untuk pengembangan lebih lanjut.

#### **2.2.13. Penelitian Nur Adha Pasaribu, Sriani**

Penelitian pada tahun 2023 yang dilakukan oleh Nur Adha dan Sriani (Pasaribu & Sriani, 2023) yang berjudul “*The Shopee Application User Reviews Sentiment Analysis Employing Naive Bayes Algorithm*”, dengan tujuan penelitian menganalisis sentimen pada ulasan produk aplikasi pada *google play store*.

Pada penelitian ini menggunakan data awal sebanyak 5000 data ulasan dan setelah dilakukan tahapan pembersihan data mendapatkan data 3946 data. Proses penelitian melakukan langkah – langkah sistematis seperti tahapan pengumpulan data dengan *scraping*, *preprocessing*, pembobotan TF-IDF, dan klasifikasi menggunakan *naive bayes*. Pada pengujiannya dengan rasio data dibagi 80:20 untuk data latih dan data uji, yang menghasilkan 3141 data latih dan data uji 805 data.

Hasil evaluasi dari algoritma *naive bayes* menunjukkan sebesar akurasi 86%, presisi 80,74%, *recall* 78,13% dan F1-score 79%. Hasil klasifikasi menemukan dominasi sentimen ke label positif sebanyak 2050 data dibandingkan data label negatif 1199 data. Metode ini menunjukkan kemampuan klasifikasi yang baik untuk analisis sentimen.

#### **2.2.14. Penelitian Syafrizal, M. Afdal, Rice Novita**

Pada tahun 2023 penelitian yang dilakukan oleh Syafrizal, M. Afdal, dan Rice Novita (Syafrizal et al., 2023) yang berjudul “*Analisa Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor*”. Penelitian bertujuan Melakukan analisis sentimen secara otomatis untuk mengetahui persepsi pengguna terhadap PLN Mobile menggunakan algoritma klasifikasi teks.

Penelitian ini menggunakan 3000 data ulasan yang dikumpulkan dengan teknik *scraping* yang ada pada *Google Play Store*, data dilakukan pelabelan seorang pakar ahli bahasa dan menghasilkan sentimen positif sebanyak 2099 data , netral 368 data dan negatif 533 data. Dan dilakukan pemodelan menggunakan algoritma NBC dan KNN dengan K-Fold Cross Validation sebagai teknik validasi.

Hasilnya evaluasi menunjukkan model NBC lebih baik dibandingkan KNN dengan akurasi sebesar 77,69%, recall 53,14%, precision 59,84% dan F1-Score 54,09%.

#### **2.2.15. Penelitian M Ikhsan Maulana, Elvia Budianita, Muhammad Fikry, Febi Yanto**

Pada tahun 2023 penelitian yang dilakukan oleh M Iksan Maulana, Elvia Budianita, Muhammad Fikry, dan Febi Yanto (Maulana, Budianita, Fikry, & Yanto, 2023), yang berjudul “ *Klasifikasi Sentimen Ulasan Aplikasi Sausage Man Menggunakan Vader Lexicon dan Naive Bayes Classifier*”. Penelitian ini bertujuan untuk menganalisis sentimen ulasan aplikasi *games Sausage Man* di *Google Play Store*.

Penelitian ini menggunakan metode *Lexicon Vader* untuk pelabelan otomatis dan *Naive Bayes Classifier* untuk klasifikasi. Data ulasan sebanyak 3000 sampel diperoleh dengan web *scraping*, diolah melalui tahapan *preprocessing* dan diklasifikasi menjadi sentimen positif, netral dan negatif. Hasil penelitian dengan evaluasi model *confusion matrix* dengan menggunakan data latih dan data uji dengan rasio 90:10 menghasilkan akurasi sebesar 75%, presisi 79% , *recall* 75% dan rasio 80:20 menghasilkan akurasi 73%, presisi 76 % , *recall* 73%. Penerapan penelitian praktis yang dari pelabelan yang secara otomatis dan Hasil analisa dapat menjadi referensi untuk pengembang dalam meningkatkan kualitas.

Berdasarkan tinjauan penelitian yang di dapatkan dapat di analisis pada penelitian sebelumnya yaitu :

1. *Naive Bayes* masih relevan untuk diteliti lebih dalam karena keunggulan komputasional dan interpretasinya.
2. Perlu juga pengembangan lanjut pada representasi fitur dan teknik validasi data sentimen opini emosional.
3. Algoritma *Naive Bayes* yang memiliki model yang ringan dan dapat berjalan baik pada perangkat yang tidak cukup tinggi.



### **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Analisa Kebutuhan**

Pada penelitian ini kebutuhan analisis sentimen, pada suatu produk bukan hanya terjadi pada produk barang ataupun merek perusahaan saja namun pada di era digital ini, produk digital tidak luput dalam suatu analisis. Contohnya pada produk ulasan digital pada *Google PlayStore* yang memiliki produk digital aplikasi yang sering digunakan pada pengguna android ataupun *mobile*.

Pada analisis sentimen membutuhkan teknik yaitu teks *mining* dalam proses untuk memahami informasi pada analisis sentimen pada data ulasan *review* aplikasi Ruang Guru di *Google PlayStore* yang sebagai sumber data. Pada tujuannya untuk memahami informasi sentimen positif, netral dan negatif pada pengguna *Google PlayStore* di aplikasi Ruang Guru yang sebagai subjek penelitian ini.



Untuk memenuhi pada kebutuhan pada sistem dilakukan pendataan secara dengan survei. Hasil jawaban yang di diberikan para responden dalam survei ini dipakai dan menjadi sebagai dasar acuan dalam mendesain aplikasi yang berfungsi untuk melakukan analisis sentimen yang akan dibangun.

### 3.1.1. *Requirement Elicitation Tahapan I*

Pada yang diperoleh hasil responden yang didapatkan dari survei, memaparkan beberapa yang dirancang untuk pengembangan platform model analisis sentimen ini. Berikut ini adalah sebuah mengenai hasil survei yang telah dilakukan :

Tabel 3. 1 *Requirement Elicitation Tahap I*

|                                            |                                                                               |
|--------------------------------------------|-------------------------------------------------------------------------------|
| <b>Fitur Fungsional</b>                    |                                                                               |
| <b>Identifikasi Kebutuhan Sistem</b>       |                                                                               |
| <b>Diharapkan sistem ini mampu untuk :</b> |                                                                               |
| 1                                          | Memberikan informasi hasil klasifikasi sentimen (positif, netral dan negatif) |
| 2                                          | Informasi data yang bentuk grafik dari hasil analisa                          |
| 3                                          | Dapat melakukan filter ulasan                                                 |
| 4                                          | Menampilkan data ulasan berbentuk tabel                                       |
| 5                                          | Menampilkan kata yang sering muncul pada ulasan                               |
| <b>Fitur Non Fungsional</b>                |                                                                               |
| <b>Diharapkan sistem ini mampu untuk :</b> |                                                                               |
| 1                                          | Informasi yang memiliki tampilan yang menarik dan mudah untuk dipahami        |

|   |                                                                    |
|---|--------------------------------------------------------------------|
| 2 | Memberikan tampilan desain yang menarik                            |
| 3 | Layanan Tampilan yang <i>responsive</i>                            |
| 4 | Menampilkan data ulasan dan visualisasi grafik yang mudah dipahami |

### 3.1.2. Requirement Elicitation Tahap II

Dari hasil yang diperoleh dari *requirement Elicitation* Tahap I, pada selanjutnya akan partisipasikan dengan klasifikasi dengan memakai teknik MDI yang dipakai untuk mengambil antara sistem perancangan yang dapat dipenuhi dan yang dilaksanakan. Pada metodenya dipakai dengan bantuan tanda yaitu : “M” yang melambangkan *mandatory* (hal yang dipertahankan untuk pada rancangan), “D” yang melambangkan *Desirable* (hal yang boleh dihilangkan), dan “I” melambangkan *Inessential* (hal kebutuhan yang bukan bagian dari hal sistem yang dibahas).

Tabel 3. 2 *Requirement Elicitation* Tahapan II

| Fitur Fungsional                    |                                                                               | M | D | I |
|-------------------------------------|-------------------------------------------------------------------------------|---|---|---|
| Identifikasi Kebutuhan Sistem       |                                                                               |   |   |   |
| Diharapkan sistem ini mampu untuk : |                                                                               |   |   |   |
| 1                                   | Memberikan informasi hasil klasifikasi sentimen (positif, netral dan negatif) | ✓ |   |   |
| 2                                   | Informasi data yang bentuk grafik dari hasil analisa                          | ✓ |   |   |
| 3                                   | Dapat melakukan filter ulasan                                                 |   | ✓ |   |
| 4                                   | Menampilkan data ulasan berbentuk tabel                                       | ✓ |   |   |

|                                            |                                                                    |   |   |   |
|--------------------------------------------|--------------------------------------------------------------------|---|---|---|
| 5                                          | Menampilkan kata yang sering muncul pada ulasan                    |   |   | ✓ |
| <b>Fitur Non Fungsional</b>                |                                                                    |   |   |   |
| <b>Diharapkan sistem ini mampu untuk :</b> |                                                                    |   |   |   |
| 1                                          | Memberikan tampilan yang menarik dan mudah untuk dipahami          | ✓ |   |   |
| 2                                          | Memberikan tampilan desain yang menarik                            | ✓ |   |   |
| 3                                          | Layanan Tampilan yang <i>responsive</i>                            |   | ✓ |   |
| 4                                          | Menampilkan data ulasan dan visualisasi grafik yang mudah dipahami | ✓ |   |   |

### 3.1.3. Requirement Elicitation Tahap III

Pada di tahapan ini, melakukan pengurangan jumlah dengan mengeliminasi semua analisa kebutuhan pada *requirement elicitation* tahapan II di atas. Dan untuk digunakan pada klasifikasinya dengan mengadopsi pendekatan manfaat TOE. Pada Dalam metode yang digunakan ini, menggunakan tanda sebagai berikut :

“T” melambangkan *Technical* merupakan langkah proses sebuah yang diterapkan untuk kebutuhan dalam sistem yang diusulkan.

“O” melambangkan *Operational* merupakan sebuah persyaratan sistem diterapkan selama tahapan pengembang.

“E” melambangkan *Economy* merupakan aspek yang meliputi biaya yang diperlukan.

Tabel 3. 3*Requirement Elicitation* Tahapan III

| Fitur Fungsional                    |                                                                               |   |   |   |   |   |   |   |   |   |
|-------------------------------------|-------------------------------------------------------------------------------|---|---|---|---|---|---|---|---|---|
| Identifikasi Kebutuhan Sistem       |                                                                               |   |   |   |   |   |   |   |   |   |
| Diharapkan sistem ini mampu untuk : |                                                                               |   |   |   |   |   |   |   |   |   |
| <i>Feasibility</i>                  |                                                                               | T |   |   | O |   |   | E |   |   |
|                                     |                                                                               | L | M | H | L | M | H | L | M | H |
| <i>Risk</i>                         |                                                                               |   |   |   |   |   |   |   |   |   |
| 1                                   | Memberikan informasi hasil klasifikasi sentimen (positif, netral dan negatif) | ✓ |   |   | ✓ |   |   | ✓ |   |   |
| 2                                   | Informasi data yang bentuk grafik dari hasil analisa                          |   | ✓ |   |   | ✓ |   |   | ✓ |   |
| 3                                   | Dapat melakukan filter ulasan                                                 |   | ✓ |   |   | ✓ |   |   | ✓ |   |
| 4                                   | Menampilkan data ulasan berbentuk tabel                                       |   | ✓ |   |   | ✓ |   |   | ✓ |   |

|                                     |                                                           |  |  |  |   |   |   |   |   |   |
|-------------------------------------|-----------------------------------------------------------|--|--|--|---|---|---|---|---|---|
| Fitur Non Fungsional                |                                                           |  |  |  |   |   |   |   |   |   |
| Identifikasi Kebutuhan Sistem       |                                                           |  |  |  |   |   |   |   |   |   |
| Diharapkan sistem ini mampu untuk : |                                                           |  |  |  |   |   |   |   |   |   |
| Feasibility                         |                                                           |  |  |  | T |   | O |   | E |   |
| Risk                                |                                                           |  |  |  | L | M | H | L | M | H |
| 1                                   | Memberikan tampilan yang menarik dan mudah untuk dipahami |  |  |  |   | ✓ |   |   | ✓ |   |

|   |                                                                    |   |   |  |   |   |  |   |  |  |
|---|--------------------------------------------------------------------|---|---|--|---|---|--|---|--|--|
| 2 | Memberikan tampilan desain yang menarik                            | ✓ |   |  | ✓ |   |  | ✓ |  |  |
| 3 | Layanan Tampilan yang <i>responsive</i>                            |   | ✓ |  |   | ✓ |  | ✓ |  |  |
| 4 | Menampilkan data ulasan dan visualisasi grafik yang mudah dipahami |   | ✓ |  |   | ✓ |  | ✓ |  |  |

#### 3.1.4. Requirement Elicitation Final

Setelah yang dilakukan pada tahapan III, pada selanjutnya tahapan ini merupakan dari tahapan akhir dari proses *elicitation* yang dipakai dalam rancangan pembuatan sistem yang melihat kebutuhan yang diperlukan berikut tabel hasil akhirnya :

Tabel 3. 4 Requirement Elicitation Final

|                                            |                                                                               |
|--------------------------------------------|-------------------------------------------------------------------------------|
| <b>Fitur Fungsional</b>                    |                                                                               |
| <b>Identifikasi Kebutuhan Sistem</b>       |                                                                               |
| <b>Diharapkan sistem ini mampu untuk :</b> |                                                                               |
| 1                                          | Memberikan informasi hasil klasifikasi sentimen (positif, netral dan negatif) |
| 2                                          | Informasi data yang bentuk grafik dari hasil analisa                          |
| 3                                          | Dapat melakukan filter ulasan                                                 |
| 4                                          | Menampilkan data ulasan berbentuk tabel                                       |
| <b>Fitur Non Fungsional</b>                |                                                                               |
| <b>Diharapkan sistem ini mampu untuk :</b> |                                                                               |
| 1                                          | Memberikan tampilan yang menarik dan mudah untuk dipahami                     |

|   |                                                                    |
|---|--------------------------------------------------------------------|
| 2 | Memberikan tampilan desain yang menarik                            |
| 3 | Layanan Tampilan yang <i>responsive</i>                            |
| 4 | Menampilkan data ulasan dan visualisasi grafik yang mudah dipahami |

### 3.2. Pembahasan Metode Yang Digunakan

#### 3.2.1. Pengambilan Data

Pada tahapan ini, sumber data yang diambil melalui sumber data dari *Google PlayStore* yang berupa data dari komentar *review* yang didapatkan melalui proses *scraping* dengan menggunakan bantuan *jupyter lab* ataupun menggunakan aplikasi dari *google colab* dengan menggunakan bantuan *library scraper*, tujuan dari menggunakan *dataset* komentar *review* untuk melihat sentimen dan opini pengguna terhadap aplikasi ruang guru. Berikut gambar dari pemanggilan *library scraper* di *google colab* :

```
from google_play_scraper import app
import pandas as pd
import numpy as np
```

Gambar 3. 1 *library scraper*

```
#Scrape desired number of reviews
#Run kode ini jika ingin scrape data dengan jumlah tertentu. Ganti (misal, ingin scrape sejumlah 1000, maka ganti kode , count = 1000 )

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.ruangguru.livestudents', # URL sumber data
    lang='id', # defaults bahasa
    country='id', # defaults negara'
    sort=Sort.NEWEST, # defaults to data yang terbaru pada ulasan review
    count=10000, # defaults jumlah data yang ingin diambil
    filter_score_with=None # defaults to None(means all score) Use 1 or 2 or 3 or 4 or 5 to select certain score
)
```

Selanjutnya setelah menyiapkan *library* untuk menggunakan *scraper*, pada

Gambar 3. 2 Proses *Scraping* Data Ulasan

penelitian ini mengambil banyaknya data ulasan di *Google PlayStore* sebanyak sepuluh



ribu data pada aplikasi ruang guru, berikut gambar pengambilan data melalui *google colab* :

Pada proses pengambilan data yang bersumber secara langsung pengambilan data atau konten ulasan dari *Google PlayStore*. Hasil dari pengambilan data sebanyak 10.000 data,



|      | reviewId                             | userName        | userImage                            | content                                           | score | thumbUpCount | reviewCreatedVersion | at                  | replyContent                                       | repliedAt           | appVersion |
|------|--------------------------------------|-----------------|--------------------------------------|---------------------------------------------------|-------|--------------|----------------------|---------------------|----------------------------------------------------|---------------------|------------|
| 0    | 73e75c12-cbb1-4557-8737-96b1a8cc016  | Pengguna Google | lh.googleusercontent.com/EGemoI2N... | Mantap sekali                                     | 5     | 0            | 6.89.2               | 2024-09-19 01:47:53 | NaN                                                | NaN                 | 6.89.2     |
| 1    | 78bd59dc-e6f8-480e-976a-c39d563d276  | Pengguna Google | lh.googleusercontent.com/EGemoI2N... | Bagus banget seru asik dan mantap                 | 5     | 0            | 6.89.2               | 2024-09-18 23:02:43 | NaN                                                | NaN                 | 6.89.2     |
| 2    | e06960a9-150b-47db-a04b-8369dee064d0 | Pengguna Google | lh.googleusercontent.com/EGemoI2N... | Karena bagus                                      | 5     | 0            | 6.89.2               | 2024-09-18 13:30:55 | NaN                                                | NaN                 | 6.89.2     |
| 3    | 240abdc-c543-4ba9-b64a-0beaf73a666d  | Pengguna Google | lh.googleusercontent.com/EGemoI2N... | Terima kasih ruang guru saya bisa belajar mate... | 5     | 0            | NaN                  | 2024-09-18 13:21:01 | NaN                                                | NaN                 | NaN        |
| 4    | d6499eb6-546a-4485-95c3-1698bdad9755 | Pengguna Google | lh.googleusercontent.com/EGemoI2N... | Aplikasinya bagus 🤝                               | 5     | 0            | NaN                  | 2024-09-18 12:53:27 | NaN                                                | NaN                 | NaN        |
| ...  | ...                                  | ...             | ...                                  | ...                                               | ...   | ...          | ...                  | ...                 | ...                                                | ...                 | ...        |
| 9995 | 9561cd3-36a2-4c89-a431-9c3bb3d9aa78  | Rahma Bina      | lh.googleusercontent.com/a/ACgBoc... | Aplikasi ruang guru sangat bagus                  | 5     | 0            | 6.38.1               | 2022-07-18 06:10:03 | Hai Rahma, terima kasih atas bintang 5-nya ya...   | 2022-07-18 05:23:03 | 6.38.1     |
| 9996 | ccafb2cc-fe7b-48fa-8d5f-f165e9da1d44 | Syahrin Indilah | lh.googleusercontent.com/a/ALVU...   | Sangat suka sama ruang guru                       | 5     | 0            | NaN                  | 2022-07-18 03:47:12 | Wah terima kasih yaa Syahrin sudah memberi rev...  | 2022-07-18 04:06:46 | NaN        |
| 9997 | c9583ec-0b6b-40b3-3c30-3cd4d7c0b6f1  | rasidi umar     | lh.googleusercontent.com/a/ACgBoc... | sangat membantu                                   | 5     | 0            | 6.38.1               | 2022-07-18 03:43:45 | Halo Rasidi, terima kasih telah memberikan bi...   | 2022-07-18 04:04:44 | 6.38.1     |
| 9998 | 87fc26cc-d327-4c4e-9e47-6b3d78ce451  | Ana Fitria      | lh.googleusercontent.com/a/ACgBoc... | aku suka game ini                                 | 5     | 0            | NaN                  | 2022-07-18 03:42:49 | Yoy! kami senang kalau Ana suka belajar menggu...  | 2022-07-18 04:02:18 | NaN        |
| 9999 | 5622bbf-7d43-4b75-9378-b68ba45d4110  | Suci Suci       | lh.googleusercontent.com/a/ACgBoc... | Bgus                                              | 5     | 0            | NaN                  | 2022-07-18 02:24:39 | Waaah, terima kasih ya Suci sudah memberikan bi... | 2022-07-18 02:41:26 | NaN        |

10000 rows × 11 columns

dan data yang didapatkan dari tahun 2022-2024. Data *scraper* yang didapatkan bisa kita simpan dalam format yang berbentuk seperti csv ataupun xls(yang berbentuk table) untuk dilakukan proses untuk pembersihan data atau *text preprocessing*, data yang di simpan berbentuk csv dilakukan pemilihan data untuk data di gunakan sebagai analisa komentar seperti ulasan, pengguna akun dan *rating* pengguna gambar . Gambar proses pengambilan dan hasil data pada gambar 3.2 dan gambar 3.3.

Gambar 3. 3 Hasil *Scraping* Data Ulasan

### 3.2.2. Text Preprocessing

Setelah dilakukan pengumpulan data, tahapan selanjutnya melakukan penanganan pengelolaan informasi berbasis teks dengan bertujuan untuk menghilangkan simbol, *emoji* gambar ataupun penghapusan kata – kata yang dianggap tidak mendukung dengan konteks ataupun relevansi, dan maupun tidak yang mengandung atau memiliki dampak nilai yang cukup besar secara makna. Dilakukan tahapan urutan proses dari *text preprocessing* yang dimulai dari *cleaning*, *case folding*, *tokenizing*, *stopword* dan *stemming* (Mehta & Pandya, 2020).

#### a. Cleaning

Setelah proses pengumpulan data dan pemisahan data yang ingin digunakan, bahwa data pada ulasan komentar masih memiliki simbol, tanda baca, angka, dan *emoticons*. Pada proses *cleaning* data menggunakan bantuan *library* yang ada pada python dengan code berikut :

```
import re
emoji_pattern = re.compile("[
    u'\U0001F600-\U0001F64F' # emoticons
    u'\U0001F300-\U0001F5FF' # symbols & pictographs
    u'\U0001F680-\U0001F6FF' # transport & map symbols
    u'\U0001F1E0-\U0001F1FF' # flags (iOS)
    "]+", flags=re.UNICODE)

def clean_text(text):
    #menghilangkan angka
    text = re.sub('[0-9]', '', text)
    #menghilangkan karakter
    text = re.sub('<.*>', '', text)
    text = re.sub('!', '', text)
    text = re.sub('%', '', text)
    text = re.sub('~', '', text)
    text = re.sub('-', '', text)
    #menghilangkan emoji
    text = re.sub(emoji_pattern, '', text)
    return text

[ ] data['data_clean'] = [clean_text(x) for x in data['ulasan']]
#tampung data clean dari emoji untuk di normalisasi
data_clean = data['data_clean']
print(data_clean)
```

Gambar 3. 4 Proses *Cleaning* Data

Contoh hasil setelah proses *cleaning* ditampilkan berikut ini :

Tabel 3. 5 Proses Hasil *Cleaning*

| No. | Awal                                                                                                                                                                                                                                                                                                                                 | Hasil                                                                                                                                                                                                                                                                                               |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Aplikasinya bagus 👍                                                                                                                                                                                                                                                                                                                  | Aplikasinya bagus                                                                                                                                                                                                                                                                                   |
| 2   | Bisa gampang buat memahami materi yang disekolah                                                                                                                                                                                                                                                                                     | Bisa gampang buat memahami materi yang disekolah                                                                                                                                                                                                                                                    |
| 3   | makasih bgt ruangguru, aku dpt akses gratis dan belajar nya gak ngebosenin. kiw kiw akses 1 bulannya ruangguru 😊😊😊. bljrnya jdi lebih fun ges, wajib download ni. yaallah semoga aku masuk smk 27 🙏🙏🙏. sukses selalu ruangguru ku, ruangguru teman belajarku ❤️❤️❤️. semoga ada rejeki trs bisa beli akses buat 1 bulan, aamiiinnn.. | makasih bgt ruangguru aku dpt akses gratis dan belajar nya gak ngebosenin kiw kiw akses bulannya ruangguru bljrnya jdi lebih fun ges wajib download ni yaallah semoga aku masuk smk sukses selalu ruangguru ku ruangguru teman belajarku semoga ada rejeki trs bisa beli akses buat bulan aamiiinnn |

b. *Case Folding*

*Case Folding* merupakan hal yang proses transformasi data teks untuk mengubah semua data teks menjadi huruf kecil. Tujuannya untuk membuat sama atau meragamkan kata teks, sehingga bila ada terdapat data yang berbeda seperti huruf tebal kapital, tidak terjadi kesalahan yang dimanah saat membandingkan dengan data latih yang sudah diseragami dengan bentuk huruf kecil. Prosesnya menggunakan fungsi dari bawaan seperti *strlower* dalam python. Berikut merupakan contoh hasil *case folding* :

Tabel 3. 6 Proses Hasil *Case Folding*

| No. | Awal                                             | Hasil                                            |
|-----|--------------------------------------------------|--------------------------------------------------|
| 1   | Aplikasinya bagus                                | aplikasinya bagus                                |
| 2   | Bisa gampang buat memahami materi yang disekolah | bisa gampang buat memahami materi yang disekolah |

c. *Tokenizing*

*Tokenizing* merupakan sebuah pengolahan data yang memisahkan data yang berupa kalimat teks diubah menjadi per kata atau sederhananya dikenal sebagai *token*. Pada kata atau *token* mewakili satu per kata atau bagian penting lainnya di dalam teks.

*Tokenizing* dilakukan untuk mengubah kalimat teks sebagian yang ada pada komentar atau ulasan pada aplikasi Ruang Guru dapat dibagi per kata atau per *tokenizing*. Berikut merupakan contoh hasil proses *tokenizing* :

Tabel 3. 7 Contoh Hasil Proses *Tokenizing*

| No. | Awal                                             | Hasil                                                                    |
|-----|--------------------------------------------------|--------------------------------------------------------------------------|
| 1   | aplikasinya bagus                                | ['aplikasinya', 'bagus']                                                 |
| 2   | bisa gampang buat memahami materi yang disekolah | ['bisa', 'gampang', 'buat', 'memahami', 'materi', 'materi', 'disekolah'] |

d. *Stopword*

Pada tahapan setelah melakukan proses *cleaning*, *case folding* dan *tokenizing*, pada selanjutnya melakukan proses penghapusan kata hubung atau *stopword* yang bertujuan untuk menghilangkan kata yang tidak memiliki nilai ataupun makna, dan sehingga data berubah menjadi lebih berkualitas dan proses menjadi lebih cepat. Agar dapat penghilangan kata *stopword* dalam teks berbahasa Indonesia menggunakan *library* fitur pada bahasa pemrograman *python* yaitu *nlTK stopwords*. Dimana dari sumber <https://www.kaggle.com/code/raniaprillyap/nltk-tokenisasi-stopwords-stemming-sastrawi#Stopwords> . Daftar *stopword* yang digunakan berjumlah 758 kata dan juga bisa menambahkan kata yang tidak mengandung makna dari data yang sudah ada. Proses pada *stopword* menggunakan kata – kata yang merupakan sebuah teks, dari setiap kata akan di cek pada setiap suku kata – perkata untuk memastikan apakah terdapat kata – kata tersebut ada pada daftar isi list *stopword* atau tidak. Proses penghilangan kata bila

tidak ada kata yang di dalam daftar *stopword* maka kata tersebut dikembali dan digabungkan menjadi teks utuh. Berikut list daftar kata *stopword* :

Tabel 3. 8 Daftar kata *Stopword*

| Daftar kata <i>Stopword</i> |             |
|-----------------------------|-------------|
| Agaknya                     | Adanya      |
| Agakpun                     | Agak        |
| Adalah                      | Amat        |
| Amatlah                     | Apakah      |
| Bagi                        | Balik       |
| Banyak                      | Bahwasanya  |
| Belakangan                  | Berkerja    |
| Belum                       | Berjumlah   |
| Berada                      | Berakhirlah |
| Berakhirnya                 | Berapa      |
| Berapakah                   | Dekat       |
| Demi                        | Demikian    |

e. *Stemming*

Proses setelah dilakukan menghapus *stopword* tahap selanjutnya melakukan *stemming* yang bertujuan untuk menguiah transformasi kata – kata dalam teks ke bentuk

dasar dilakukan dengan menghilangkan sebuah imbuhan pada setiap kata yang terdapat di awal ataupun di akhir kata. Pada prosesnya menggunakan *library* sastrawi, yang merupakan *library* yang merupakan *library* yang mendukung Bahasa Indonesia dengan bantuan pemrosesan *Natural Language Processing* (NLP) yang dikembangkan oleh Andy Librian pada *library* ini bersumber pada <https://github.com/sastrawi/sastrawi/tree/master/src/Sastrawi/Stemmer>

Hasil pada tahapan *preprocessing* ini dari sepuluh ribu data ulasan yang di ambil dari proses *scraper*, dan setelah melakukan proses *preprocessing* menjadi data 9700 data bersih yang di hasilkan untuk proses pelabelan otomatis dan juga analisa sentimen menggunakan *NaiveBayes*.

### 3.2.3. Pelabelan Otomatis *Lexicon Vader*

Pada penggunaan untuk pelabelan data otomatis menggunakan bantuan *lexicon vader*. Setelah tahapan *preprocessing*, dan sebelum melakukan proses klasifikasi maka memerlukan data label untuk pada data *training* yang akan digunakan pada klasifikasi. Proses pelabelan data dengan *library lexicon vader* pada pemrograman bahasa *python*, data dari hasil *preprocessing* melakukan proses *translating* data dikarenakan *lexicon vader* mendukung bahasa dari bahasa Inggris atau *corpus*. Tujuan dari proses *translator* ini untuk membantu proses pelabelan otomatis. Berikut ini adalah contoh hasil dari proses *translator* :

Tabel 3. 7 Contoh Hasil Proses *Translator*

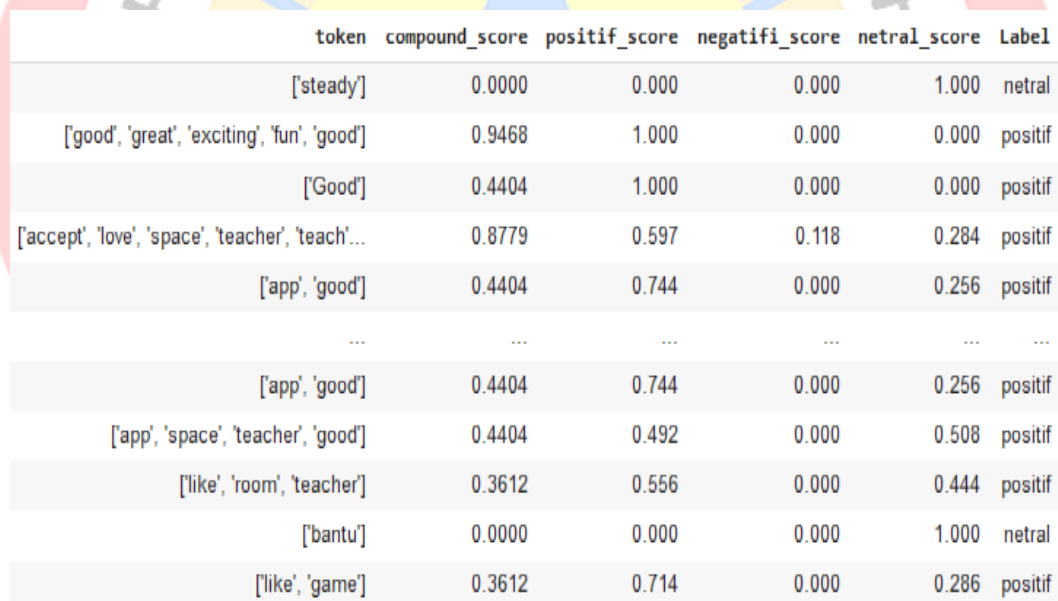
| No. | Awal                                         | Hasil                                |
|-----|----------------------------------------------|--------------------------------------|
| 1   | ['aplikasi', 'bagus']                        | [app, good]                          |
| 2   | ['gampang', 'memahami', 'materi', 'sekolah'] | [easy, understand, material, school] |



Setelah dilakukan pada tahapan *translator* pada data ulasan yang sudah bersih, selanjutnya adalah tahapan untuk melakukan proses pelabelan otomatis dengan *lexicon vader*. Pada *library lexicon vader* ini menggunakan bantuan kamus atau *lexicon* sebagai pokok bahasanya. Untuk menentukan hasil pelabelan otomatis ini pada *library lexicon* dengan melihat hasil *score polarity* yang dihasilkan dari banyaknya nilai di hasilkan pada *compound\_score*. Hasil dari *score polarity* ini dibagi menjadi 3 bagian yaitu :

1. *Score polarity* < 0 yang melabelkan sebagai data label negatif.
2. *Score polarity* = 0 yang melabelkan sebagai data label netral.
3. *Score polarity* > 0 yang melabelkan sebagai data label positif.

Berikut gambar dari hasil pelabelan otomatis dengan bantuan *library lexicon vader* :



| token                                          | compound_score | positif_score | negatifi_score | netral_score | Label   |
|------------------------------------------------|----------------|---------------|----------------|--------------|---------|
| [steady]                                       | 0.0000         | 0.000         | 0.000          | 1.000        | netral  |
| [good, 'great', 'exciting', 'fun', 'good']     | 0.9468         | 1.000         | 0.000          | 0.000        | positif |
| [Good]                                         | 0.4404         | 1.000         | 0.000          | 0.000        | positif |
| [accept, 'love', 'space', 'teacher', 'teach... | 0.8779         | 0.597         | 0.118          | 0.284        | positif |
| [app, 'good']                                  | 0.4404         | 0.744         | 0.000          | 0.256        | positif |
| ...                                            | ...            | ...           | ...            | ...          | ...     |
| [app, 'good']                                  | 0.4404         | 0.744         | 0.000          | 0.256        | positif |
| [app, 'space', 'teacher', 'good']              | 0.4404         | 0.492         | 0.000          | 0.508        | positif |
| [like, 'room', 'teacher']                      | 0.3612         | 0.556         | 0.000          | 0.444        | positif |
| [bantu]                                        | 0.0000         | 0.000         | 0.000          | 1.000        | netral  |
| [like, 'game']                                 | 0.3612         | 0.714         | 0.000          | 0.286        | positif |

Gambar 3. 5 Gambar Hasil Pelabelan Otomatis

Jadi pada hasil dari pelabelan otomatis ini mendapatkan total data yang berlabel positif sebanyak 6726 data, 2214 data berlabel netral dan sebanyak 760 data berlabel negatif. Dan yang di mana data tersebut akan dibagi menjadi data latih dan data uji. Pemisahan proses data latih dan data uji bertujuan untuk melatih model menggunakan data latih dan data uji sebagai untuk menguji kinerja model.

### 3.2.4. Klasifikasi Manual *Naïve Bayes*

#### a. Perhitungan Manual *Naïve Bayes*

Pada tahapan ini, akan melakukan proses perhitungan secara manual dengan menggunakan *Naïve Bayes*, sebelumnya untuk dilakukan perhitungan secara manual, data yang akan digunakan sudah melalui tahapan *Text Preprocessing* pada data yang ingin dijadikan sebagai contoh manual ini, untuk datanya yang digunakan menjadi bersih dan bernilai. Di bawah ini adalah langkah hasil proses *preprocessing* dan perhitungan secara manual dengan *Naïve Bayes*.

Tabel 3. 8 Contoh Hasil Data *Preprocessing*

| Sebelum <i>Text Preprocessing</i>                                                                                                                                                                                                         | Sesudah <i>Text Preprocessing</i>                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Aplikasi Ruangguru ini menyenangkan bagi saya<br><br>mempermudah proses belajar, meningkatkan daya belajar, membuat suasana belajar, mengasah pengetahuan, membangun motivasi bagi guru lain<br><br>untuk menggunakan aplikasi Ruangguru. | aplikasi ruangguru senang mudah proses ajar tingkat daya ajar suasana ajar asah tahu bangun<br><br>motivasi guru aplikasi ruangguru            |
| Khusus belajar menambah pengetahuan dan ketangkasan berfikir                                                                                                                                                                              | khusus ajar tambah tahu tangkas berfikir                                                                                                       |
| Buruk... selalu ada gangguan, baik saat melihat profil, melihat ruang kelas maupun mengerjakan soal, gangguannya yaitu Bug seperti tidak bisa                                                                                             | buruk selalu gangguan lihat profil lihat ruang kelas<br><br>kerja soal gangguan bug tekan muncul jadi bisa<br><br>ubah kelas begini mohon baik |

|                                                |  |
|------------------------------------------------|--|
| ditekan dan tidak muncul, saya jadi tidak bisa |  |
| merubah kelas kalau begini. Mohon segera       |  |
| diperbaiki                                     |  |

Selanjutnya setelah melakukan tahapan *Text Preprocessing*, dilakukan tahapan memisahkan teks menjadi perkata – kata (*tokenizing* /tokenisasi) untuk mencari nilai bobot dari setiap kata – kata agar dapat di proses ketahap klasifikasi dengan *Naïve Bayes*.

Tabel 3. 9 Hasil *Tokenizing*

| <i>Term</i> |           |          |        |         |          |
|-------------|-----------|----------|--------|---------|----------|
| Aplikasi    | ruangguru | senang   | mudah  | proses  | Ajar     |
| Tingkat     | daya      | suasana  | asah   | tahu    | bangun   |
| Motivasi    | guru      | khusus   | tambah | tangkas | berfikir |
| Buruk       | selalu    | gangguan | lihat  | profil  | Ruang    |
| Kelas       | kerja     | soal     | bug    | tekan   | muncul   |
| Jadi        | bisa      | ubah     | begini | mohon   | baik     |

Setelah dilakukan pemisahan kalimat teks yang diubah menjadi per kata pada tahapan berikutnya akan melakukan pembobotan kata atau *Term Frequency* disetiap masing – masing komentar pada kategori atau dokumen.

Tabel 3. 10 Pembobotan Kata

| <i>Term</i> | <b>D1(positif)</b> | <b>D2(netral)</b> | <b>D3(negatif)</b> |
|-------------|--------------------|-------------------|--------------------|
| Ajar        | 3                  | 1                 | 0                  |
| Aplikasi    | 2                  | 0                 | 0                  |
| Asah        | 1                  | 0                 | 0                  |
| Bangun      | 1                  | 0                 | 0                  |
| Begini      | 0                  | 0                 | 1                  |
| Bisa        | 0                  | 0                 | 1                  |
| Berpikir    | 0                  | 1                 | 0                  |
| Buruk       | 0                  | 0                 | 1                  |
| Bug         | 0                  | 0                 | 1                  |
| Daya        | 1                  | 0                 | 0                  |
| Gangguan    | 0                  | 0                 | 2                  |
| Guru        | 1                  | 0                 | 0                  |
| Jadi        | 0                  | 0                 | 1                  |
| Kelas       | 0                  | 0                 | 2                  |
| Kerja       | 0                  | 0                 | 1                  |
| Khusus      | 0                  | 1                 | 0                  |

|           |   |   |   |
|-----------|---|---|---|
| Lihat     | 0 | 0 | 2 |
| Motivasi  | 1 | 0 | 0 |
| Mohon     | 0 | 0 | 1 |
| Profil    | 0 | 0 | 1 |
| Proses    | 1 | 0 | 0 |
| Ruang     | 0 | 0 | 1 |
| Ruangguru | 2 | 0 | 0 |
| Senang    | 1 | 0 | 0 |
| Soal      | 0 | 0 | 1 |
| Selalu    | 0 | 0 | 1 |
| Suasana   | 1 | 0 | 0 |
| Tambah    | 0 | 1 | 0 |
| Tangkas   | 0 | 1 | 0 |
| Tahu      | 1 | 1 | 0 |
| Tekan     | 0 | 0 | 1 |
| Tingkat   | 1 | 0 | 0 |
| Ubah      | 0 | 0 | 1 |
| Muncul    | 0 | 0 | 1 |

|                    |           |   |    |
|--------------------|-----------|---|----|
| Mudah              | 1         | 0 | 0  |
| Baik               | 0         | 0 | 1  |
| <b>Total Term</b>  | 18        | 6 | 21 |
| <b>Total Bobot</b> | <b>45</b> |   |    |

### 3.2.5. Proses Klasifikasi Data *Training* Pada *Naïve Bayes*

Pada tahapan ini setelah melakukan pembobotan perhitungan pada setiap *term*, selanjutnya melakukan perhitungan nilai probabilitas pada setiap *term*.

#### a. Perhitungan Nilai Probabilitas *Priori*

Pada perhitungan tahapan dilakukanlah, perhitungan pada kategori atau atribut label, terdapat tiga buah kategori merupakan label positif, label netral dan label negatif. Ketiga label tersebut akan dihitung nilai probabilitasnya dengan memanfaatkan 3 komentar atau dokumen dalam proses pelatihan.

$$P(d) = \frac{Nd}{N}$$

$Nd$  = adalah jumlah dokumen pada data latih yang tergolong label  $d$

$N$  = adalah banyak dokumen pada data latih

$$P(\text{positif}) = \frac{1}{3} = 0,3333$$

$$P(\text{netral}) = \frac{1}{3} = 0,3333$$

$$P(\text{negatif}) = \frac{1}{3} = 0,3333$$

#### b. Menghitung *Conditional Probability*



Pada di tahapan ini akan melakukan perhitungan *conditional probability* yang merupakan suatu perhitungan untuk mencari nilai probabilitas dari masing – masing kata atau *term*. Estimasi proses penghitungan peluang ini dengan memakai model *Multinomial*.

$$P(k|d) = \frac{\text{Count}(k,d)+1}{\text{count}(c)+|V|}$$

Pada perhitungan untuk menghindari adanya nilai probabilitas nol, maka dilakukan penambahan +1 pada pengambilan dan |V| pada penyebutan.

Count(k,d) = adalah total kata yang muncul di label d.

Count(d) = adalah total setiap kata di label d.

|V| = adalah total unit teks yang dikelan sebagai kata atau token.

Berikut merupakan contoh dari hitungan menggunakan model *multinomial* :

Dokumen atau komentar 1 = “aplikasi ruangguru senang mudah proses ajar tingkat daya ajar asah tahu bangun motivasi guru aplikasi ruangguru”.

#### 1. Probabilitas kata “aplikasi”

$$P(\text{aplikasi}|\text{positif}) = \frac{2+1}{18+45} = 0,047619$$

$$P(\text{aplikasi}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{aplikasi}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

#### 2. Probabilitas kata “ruangguru”

$$P(\text{ruangguru}|\text{positif}) = \frac{2+1}{18+45} = 0,047619$$

$$P(\text{ruangguru}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{ruangguru}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

3. Probabilitas kata “senang”

$$P(\text{senang}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{senang}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{senang}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

4. Probabilitas kata “mudah”

$$P(\text{mudah}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{mudah}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{mudah}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

5. Probabilitas kata “proses”

$$P(\text{proses}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{proses}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{proses}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

6. Probabilitas kata “ajar”

$$P(\text{ajar}|\text{positif}) = \frac{3+1}{18+45} = 0,063492$$

$$P(\text{ajar}|\text{netral}) = \frac{1+1}{6+45} = 0,039219$$

$$P(\text{ajar}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

7. Probabilitas kata “tingkat”

$$P(\text{tingkat}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{tingkat}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{tingkat}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

8. Probabilitas kata “daya”

$$P(\text{daya}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{daya}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{daya}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

9. Probabilitas kata “suasana”

$$P(\text{suasana}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{suasana}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{suasana}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

10. Probabilitas kata “asah”

$$P(\text{asah}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{asah}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{asah}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

11. Probabilitas kata “tahu”

$$P(\text{tahu}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{tahu}|\text{netral}) = \frac{1+1}{6+45} = 0,039219$$

$$P(\text{tahu}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

12. Probabilitas kata “bangun”

$$P(\text{bangun}|\text{positif}) = \frac{1+1}{18+45} = 0,03174$$

$$P(\text{bangun}|\text{netral}) = \frac{0+1}{6+45} = 0,01960$$

$$P(\text{bangun}|\text{negatif}) = \frac{0+1}{21+45} = 0,01515$$

### 13. Probabilitas kata “motivasi”

$$P(\text{motivasi}|\text{positif}) = \frac{1+1}{18+45} = 0,03175$$

$$P(\text{motivasi}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{motivasi}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

### 14. Probabilitas kata “guru”

$$P(\text{guru}|\text{positif}) = \frac{1+1}{18+45} = 0,031746$$

$$P(\text{guru}|\text{netral}) = \frac{0+1}{6+45} = 0,019608$$

$$P(\text{guru}|\text{negatif}) = \frac{0+1}{21+45} = 0,015152$$

Setelah melakukan perhitungan nilai probabilitas, pada setiap *term* atau kata maka selanjutnya melakukan tahap untuk pengklasifikasi dengan data uji. Berikut merupakan hasil setiap perhitungan *conditional probability* :

Tabel 3. 11 Perhitungan *Conditional Probability*

| <i>Term</i> | <b>D1(positif)</b> | <b>D2(netral)</b> | <b>D3(negatif)</b> |
|-------------|--------------------|-------------------|--------------------|
| Ajar        | 0,063492           | 0,039216          | 0,015152           |
| aplikasi    | 0,047619           | 0,019608          | 0,015152           |
| Asah        | 0,031746           | 0,019608          | 0,015152           |
| bangun      | 0,031746           | 0,019608          | 0,015152           |

|           |          |          |          |
|-----------|----------|----------|----------|
| Begini    | 0,015873 | 0,019608 | 0,030303 |
| Bisa      | 0,015873 | 0,019608 | 0,030303 |
| berfikir  | 0,015873 | 0,019608 | 0,015152 |
| Buruk     | 0,015873 | 0,019608 | 0,030303 |
| Bug       | 0,015873 | 0,019608 | 0,030303 |
| Daya      | 0,031746 | 0,019608 | 0,015152 |
| gangguan  | 0,015873 | 0,019608 | 0,045455 |
| Guru      | 0,031746 | 0,019608 | 0,015152 |
| Jadi      | 0,015873 | 0,019608 | 0,030303 |
| Kelas     | 0,015873 | 0,019608 | 0,045455 |
| Kerja     | 0,015873 | 0,019608 | 0,030303 |
| Khusus    | 0,015873 | 0,039216 | 0,015152 |
| Lihat     | 0,015873 | 0,019608 | 0,045455 |
| motivasi  | 0,031746 | 0,019608 | 0,015152 |
| Mohon     | 0,015873 | 0,019608 | 0,030303 |
| Profil    | 0,015873 | 0,019608 | 0,030303 |
| Proses    | 0,031746 | 0,019608 | 0,015152 |
| Ruang     | 0,015873 | 0,019608 | 0,030303 |
| ruangguru | 0,047619 | 0,019608 | 0,015152 |

|         |          |          |          |
|---------|----------|----------|----------|
| Senang  | 0,031746 | 0,019608 | 0,015152 |
| Soal    | 0,015873 | 0,019608 | 0,030303 |
| Selalu  | 0,015873 | 0,019608 | 0,030303 |
| suasana | 0,031746 | 0,019608 | 0,015152 |
| tambah  | 0,015873 | 0,039216 | 0,015152 |
| tangkas | 0,015873 | 0,039216 | 0,015152 |
| Tahu    | 0,031746 | 0,039216 | 0,015152 |
| Tekan   | 0,015873 | 0,019608 | 0,030303 |
| tingkat | 0,031746 | 0,019608 | 0,015152 |
| Ubah    | 0,015873 | 0,019608 | 0,030303 |
| muncul  | 0,015873 | 0,019608 | 0,030303 |
| Mudah   | 0,031746 | 0,019608 | 0,015152 |
| Baik    | 0,015873 | 0,019608 | 0,030303 |

### 3.2.6. Proses Klasifikasi Data *Testing* Pada *Naïve Bayes*

Pada tahapan perhitungan untuk menghitung nilai Posterior pada setiap - setiap kelas ataupun kategori label. Hasil dari nilai yang didapatkan nilai posterior tertinggi akan menjadi kelas baru dari data uji tersebut.

$$P(k|d) = P(k) * \prod P(w|k)$$

$P(k|d)$  adalah nilai probabilitas *posterior* dari kelas k untuk dokumen d.

$P(k)$  adalah nilai probabilitas dasar atas kemunculan kategori  $k$ .

$\prod P(w|k)$  merupakan dari perhitungan yang didapatkan *conditional probability* tiap – tiap *term* atau kata yang ada pada dokumen  $k$ . Di bawah ini merupakan perhitungan pengelompokan atau klasifikasi data uji pada tiga dokumen sebelumnya.

“aplikasi ruangguru senang mudah proses ajar tingkat daya ajar suasana ajar asah tahu bangun motivasi guru aplikasi ruangguru”

Positif :

$$P(\text{positif}|D1) = P(\text{positif}) * P(\text{aplikasi}|\text{positif}) * P(\text{ruangguru}|\text{positif}) * P(\text{positif}|\text{senang}) * P(\text{mudah}|\text{positif}) * P(\text{proses}|\text{positif}) * P(\text{ajar}|\text{positif}) * P(\text{tingkat}|\text{positif}) * P(\text{daya}|\text{positif}) * P(\text{suasana}|\text{positif}) * P(\text{asah}|\text{positif}) * P(\text{tahu}|\text{positif}) * P(\text{bangun}|\text{positif}) * P(\text{motivasi}|\text{positif}) * P(\text{guru}|\text{positif})$$

$$P(\text{positif}|D1) = 0,3333 * 0,047619 * 0,047619 * 0,031746 * 0,031746 * 0,031746 * 0,063492 * 0,031746 * 0,031746 * 0,031746 * 0,031746 * 0,031746 * 0,031746 * 0,031746 * 0,031746$$

$$P(\text{positif}|D1) = 1,58378E-21$$

Netral :

$$P(\text{netral}|D1) = P(\text{netral}) * P(\text{aplikasi}|\text{netral}) * P(\text{ruangguru}|\text{netral}) * P(\text{senang}|\text{netral}) * P(\text{mudah}|\text{netral}) * P(\text{proses}|\text{netral}) * P(\text{ajar}|\text{netral}) * P(\text{tingkat}|\text{netral}) * P(\text{daya}|\text{netral}) * P(\text{suasana}|\text{netral}) * P(\text{asah}|\text{netral}) * P(\text{tahu}|\text{netral}) * P(\text{bangun}|\text{netral}) * P(\text{motivasi}|\text{netral}) * P(\text{guru}|\text{netral})$$

$$P(\text{netral}|D1) = 0,3333 * 0,019608 * 0,019608 * 0,019608 * 0,019608 * 0,019608 * 0,039219 * 0,019608 * 0,019608 * 0,019608 * 0,019608 * 0,039219 * 0,019608 * 0,019608 * 0,019608$$

$$P(\text{netral}|D1) = 1,65588E-24$$

Negatif :



$$P(\text{negatif}|D1) = P(\text{negatif}) * P(\text{aplikasi}|\text{negatif}) * P(\text{ruangguru}|\text{negatif}) * \\ P(\text{senang}|\text{negatif}) * P(\text{mudah}|\text{negatif}) * P(\text{proses}|\text{negatif}) * P(\text{ajar}|\text{negatif}) * \\ P(\text{tingkat}|\text{negatif}) * P(\text{daya}|\text{negatif}) * P(\text{suasana}|\text{negatif}) * P(\text{asah}|\text{negatif}) * \\ P(\text{tahu}|\text{negatif}) * P(\text{bangun}|\text{negatif}) * P(\text{motivasi}|\text{negatif}) * P(\text{guru}|\text{negatif})$$

$$P(\text{negatif}|D1) = 0,3333 * 0,015152 * 0,015152 * 0,015152 * 0,015152 * 0,015152 * \\ 0,015152 * 0,015152 * 0,015152 * 0,015152 * 0,015152 * 0,015152 * 0,015152 * \\ 0,015152 * 0,015152$$

$$P(\text{negatif}|D1) = 1,12051\text{E-}26$$

Berikut table hasil dari perhitungan klasifikasi data *testing* :

Tabel 3. 12 Klasifikasi Data *Testing*

| Hasil Komentar Positif                                                                                                           |               |                      |            |
|----------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------|------------|
| Teks                                                                                                                             | Peluang kelas | Probabilitas Peluang | Kelas baru |
| aplikasi ruangguru senang mudah proses ajar tingkat daya ajar suasana ajar asah tahu bangun motivasi guru aplikasi ruangguru     | P(positif)    | 1,58378E-21          | Positif    |
|                                                                                                                                  | P(netral)     | 1,65588E-24          |            |
|                                                                                                                                  | P(negatif)    | 1,12051E-26          |            |
| Hasil Komentar Netral                                                                                                            |               |                      |            |
| khusus ajar tambah tahu tangkas berfikir                                                                                         | P(positif)    | 4,26464E-11          | Netral     |
|                                                                                                                                  | P(netral)     | 1,21226E-09          |            |
|                                                                                                                                  | P(negatif)    | 4,03248E-12          |            |
| Hasil Komentar Negatif                                                                                                           |               |                      |            |
| buruk selalu gangguan lihat profil lihat ruang kelas kerja soal gangguan bug tekan muncul jadi bisa ubah kelas begini mohon baik | P(positif)    | 1,36366E-33          | Negatif    |
|                                                                                                                                  | P(netral)     | 6,11747E-32          |            |
|                                                                                                                                  | P(negative)   | 5,22226E-28          |            |

Berdasarkan hasil dari perhitungan di atas nilai yang dimiliki dari nilai probabilitas *posterior* terbesar, maka akan menentukan hasil pada label pada nilai yang terbesar. Contohnya pada dokumen 1, yang menghasilkan nilai peluang positif lebih tinggi dari pada nilai netral ataupun negatif, maka dokumen 1 dikategorikan pada kelas positif.

### 3.2.7. Klasifikasi Pada *Python*

Pada klasifikasinya proses dengan *python* dibantu dengan *library textblob* untuk proses klasifikasinya. Berikut merupakan proses klasifikasi *naive bayes* pada *python* :



```
!pip install textblob
from textblob import TextBlob
from textblob.classifiers import NaiveBayesClassifier
cl = NaiveBayesClassifier(dataset)

datauji = 'ruang guru aplikasi terbaik'
analysis = TextBlob(datauji, classifier=cl)
print(datauji)
print('Hasil Klasifikasi : '+analysis.classify())
```

Requirement already satisfied: textblob in /usr/local/lib/python3.11/dist-packages (0.19.0)  
Requirement already satisfied: nltk>=3.9 in /usr/local/lib/python3.11/dist-packages (from textblob) (3.9.1)  
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (8.1.8)  
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (1.4.2)  
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (4.6.0)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (4.66.1)  
ruang guru aplikasi terbaik  
Hasil Klasifikasi : positif

Gambar 3. 6 Hasil Klasifikasi *Python*

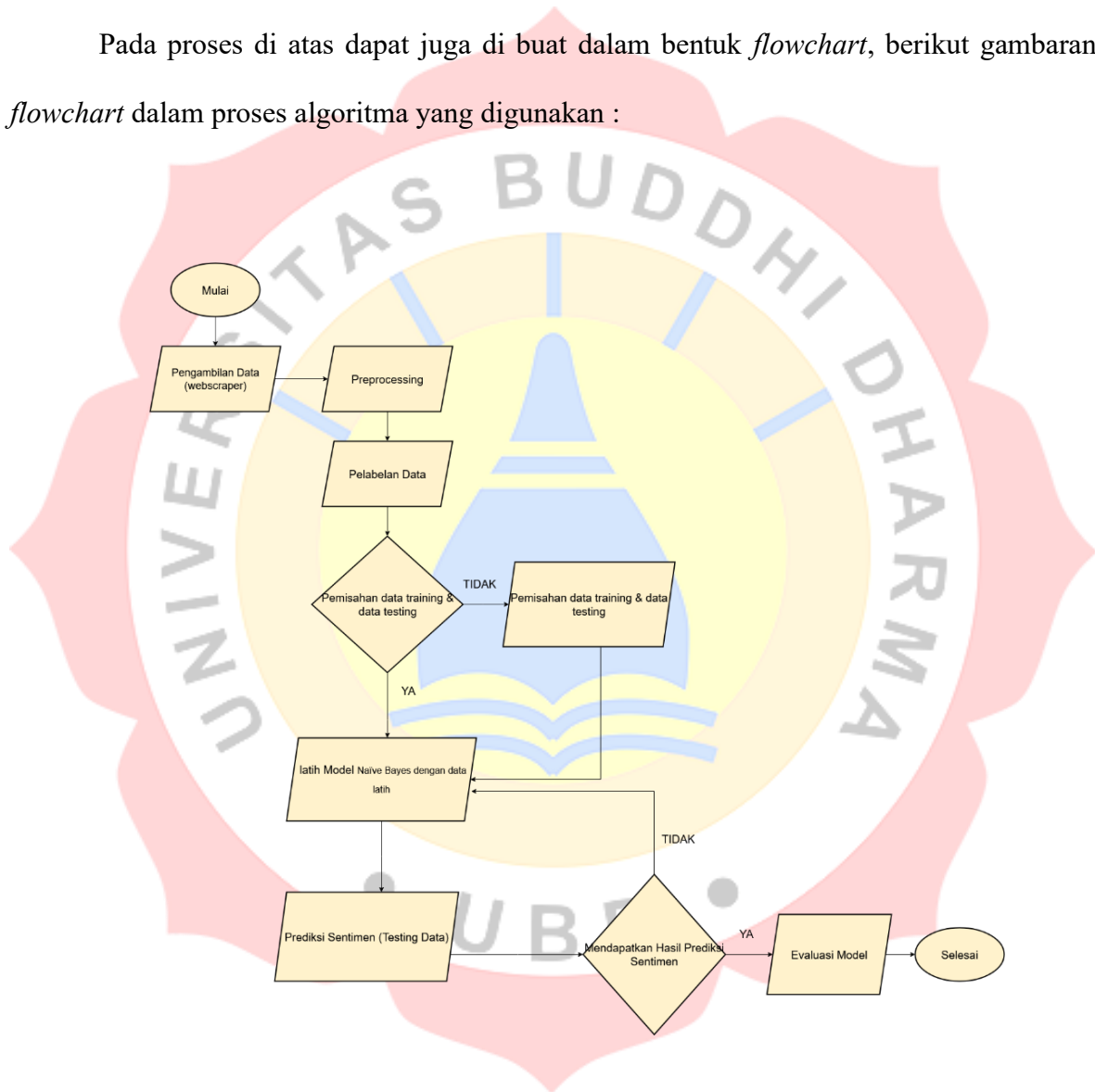
### 3.3. Pembahasan Algoritma Yang Digunakan

Pada pembahasan ini, algoritma *Naïve Bayes* yang digunakan untuk proses permasalahan pada sentimen analisis ini. Berikut langkah – langkah proses penyelesaian algoritma yang digunakan :

1. Mulai (*start*).
2. Pengambilan data (*webscraper*) Pada Aplikasi ruangguru.
3. *Preprocessing* Data.
4. Pelabelan data.

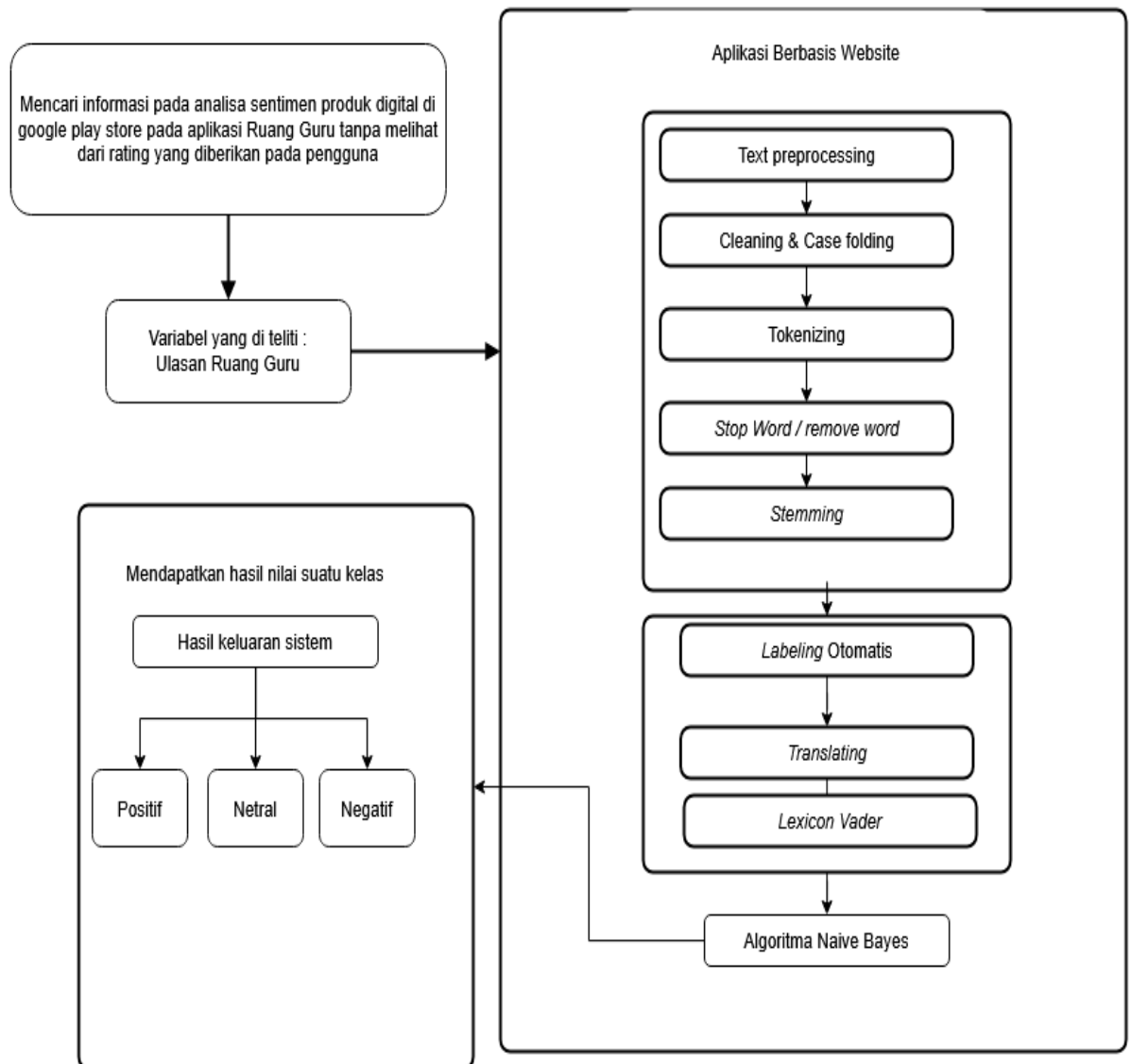
5. Pemisahan data testing dan data *training*.
6. Latih Model *Naïve Bayes* dengan data *training*.
7. Prediksi sentimen untuk data baru dengan data testing.
8. Evaluasi model.
9. Selesai (*end*).

Pada proses di atas dapat juga di buat dalam bentuk *flowchart*, berikut gambaran *flowchart* dalam proses algoritma yang digunakan :



Gambar 3. 7 *Flowchart* Pembahasan Algoritma

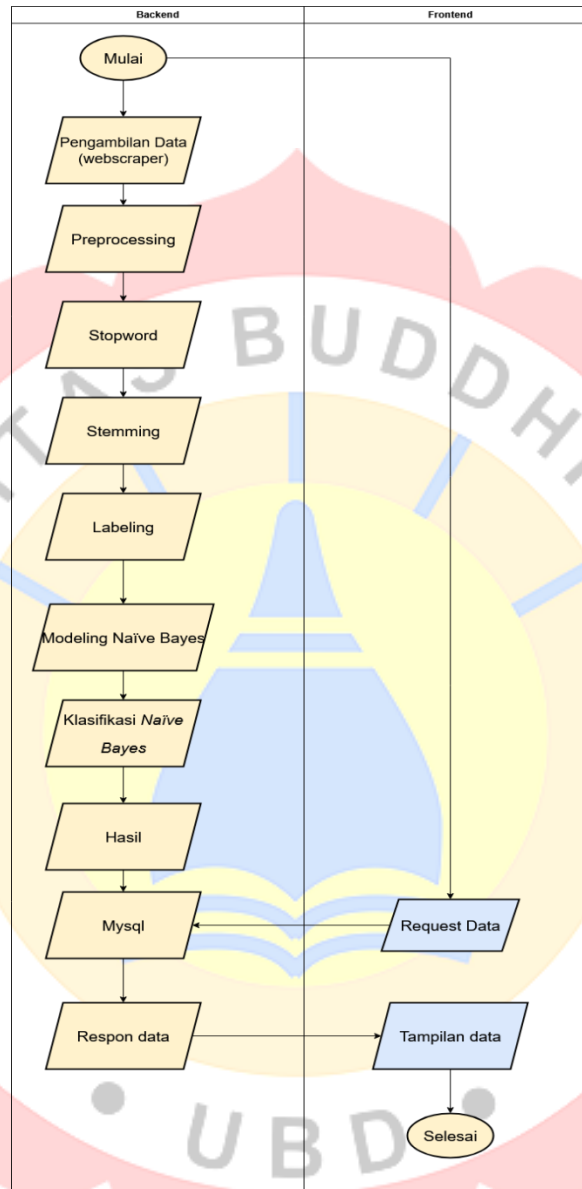
### 3.4. Kerangka Pemikiran



Gambar 3. 8 Kerangka Pemikiran

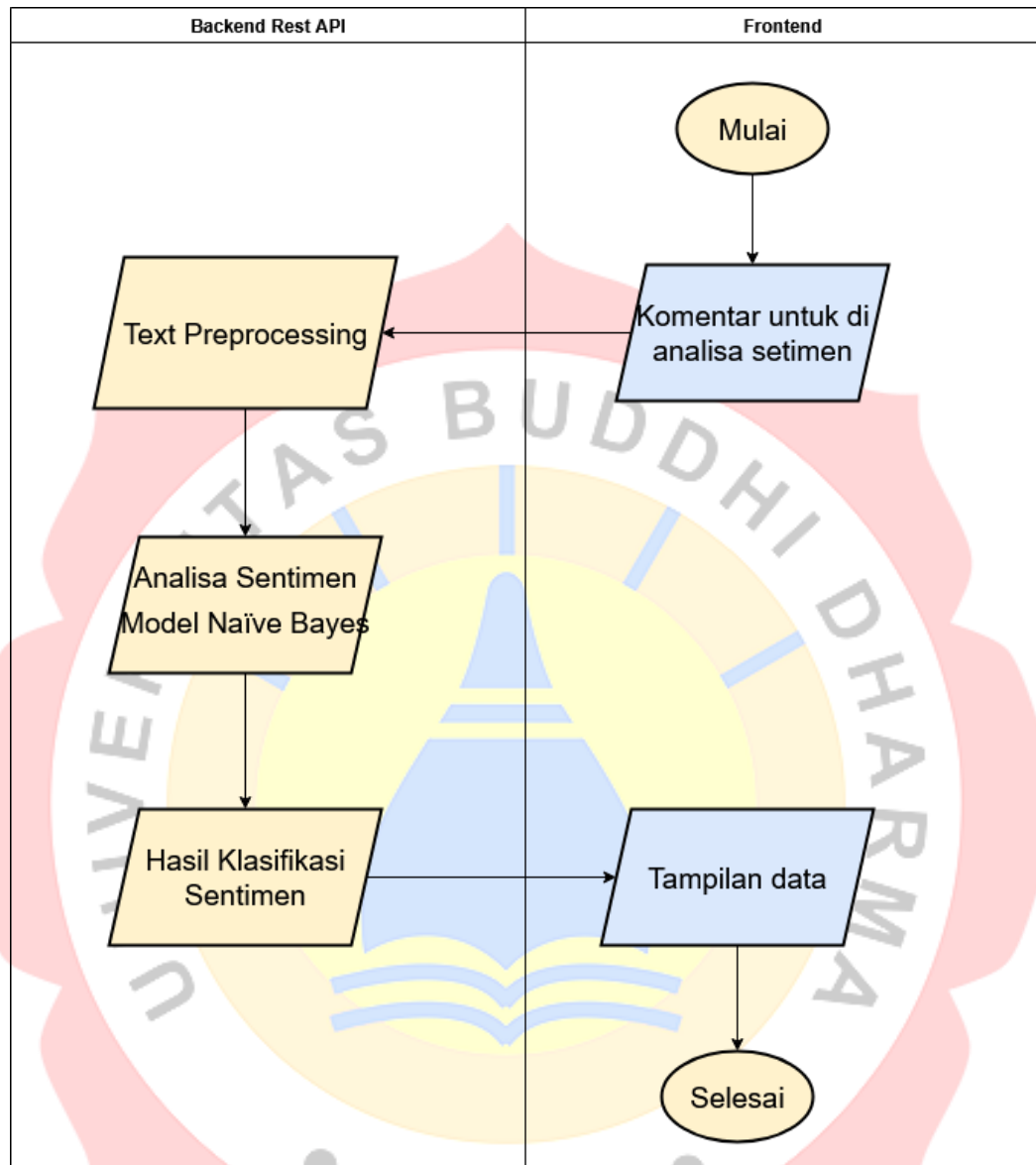
### 3.5. Perancangan *Flowchart*

#### 3.5.1. Rancangan *Flowchart Text Mining Data Review Komentor Ruang Guru*



Gambar 3. 9 Rancang *Flowchart Text Mining*

### 3.5.2. Rancangan *Flowchart* Sentimen Analisis Komentar

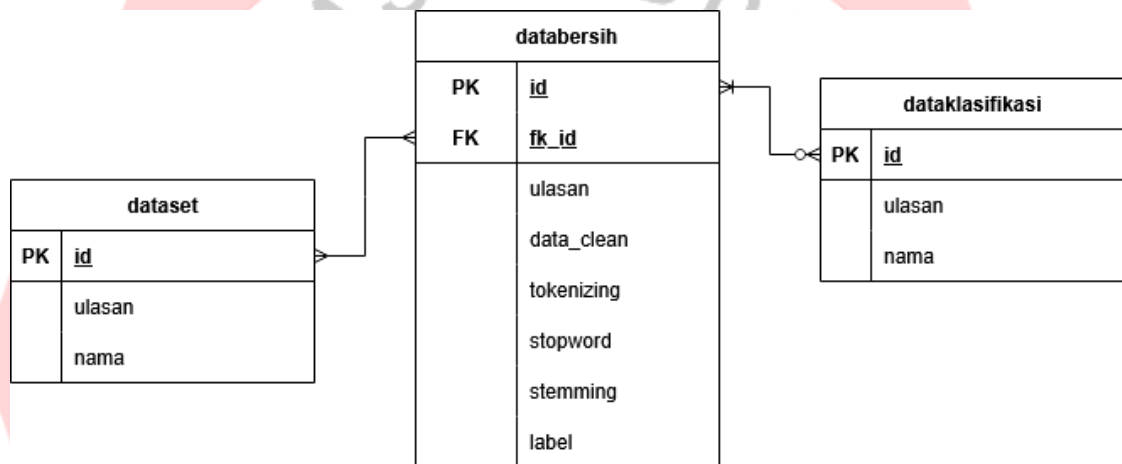


Gambar 3. 10 Rancangan *Flowchart* Sentimen Analisis Komentar

### 3.6. Perancangan Layar, Menu, Database

#### 3.6.1. Perancangan Database

Pada tahapan perancangan akan membahas pada tentang basis data yang dikembangkan dan merancang struktur basis data yang akan dipakai sebagai saranan tempat simpan data dan mengelola data pada aplikasi yang akan dikembangkan dengan program *database mysql*, berikut diagram perancangan basis data atau dikenal *Entity Relationship Diagram (ERD)*:



Gambar 3. 11 Entity Relationship Diagram

Tabel 3. 13 Keterangan Tabel Dataset

| No. | Attribut | Tipe Data | Uraian                              |
|-----|----------|-----------|-------------------------------------|
| 1   | Id       | Int(11)   | primary key dataset                 |
| 2   | ulasan   | text      | ulasan kotor hasil web scrapping    |
| 3   | nama     | text      | nama dari pengguna google playstore |

Tabel 3. 14 Keterangan Tabel databersih (Proses Preprocessing)

| No. | Attribut | Tipe Data | Uraian |
|-----|----------|-----------|--------|
|-----|----------|-----------|--------|



|   |            |         |                                             |
|---|------------|---------|---------------------------------------------|
| 1 | Id         | Int(11) | <i>primary key</i> databersih               |
| 2 | fk_id      | Int(11) | <i>Foreign key</i> id pada table dataset    |
| 3 | ulasan     | text    | ulasan kotor hasil web scrapping            |
| 4 | data_clean | text    | Hasil ulasan data dari proses data cleaning |
| 5 | tokenizing | text    | Hasil ulasan data dari proses tokenizing    |
| 6 | stopword   | text    | Hasil ulasan data dari proses stopwords     |
| 7 | stemming   | text    | Hasil ulasan data dari proses stemming      |
| 8 | label      | text    | Hasil pelabelan otomatis dari lexicon Vader |

Tabel 3. 15 Keterangan tabel dataklasifikasi

| No. | Attribut          | Jenis Data | Keterangan                                                   |
|-----|-------------------|------------|--------------------------------------------------------------|
| 1   | fk_id             | Int(11)    | <i>Foreign key</i> databersih (data <i>preprocessing</i> )   |
| 2   | ulasan            | text       | ulasan data bersih (hasil dari proses <i>preprocessing</i> ) |
| 3   | klasifikasi_bayes | text       | Hasil prediksi klasifikasi dengan Naïve Bayes                |
| 4   | kata_kunci        | Int(11)    | Data rating dari setiap ulasan                               |

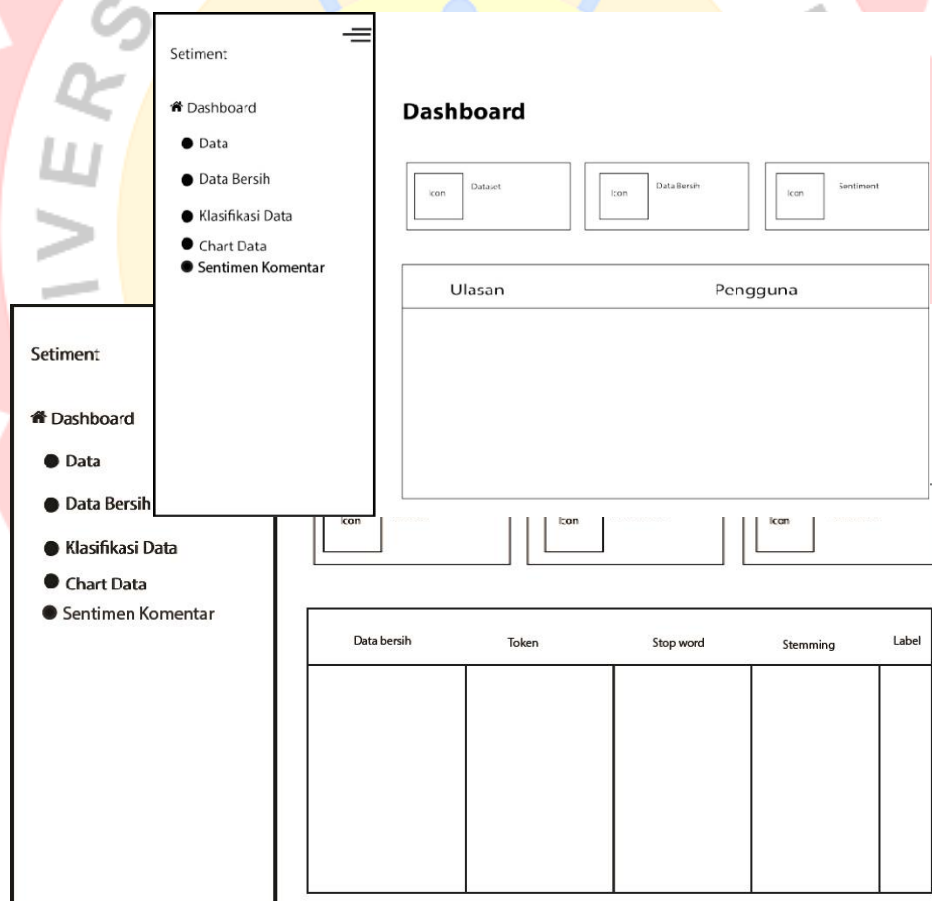
### 3.6.2. Perancangan Layar

Pada perancangan layer aplikasi yang berbentuk website ini akan ada empat tampilan halaman, pada tampilan halaman pertama akan menampilkan data ulasan yang

Gambar 3. 12 Rancang Tampilan Utama

sudah di kumpulkan dari *google play store* dengan bantuan scrapping dan di simpan pada *database*, halaman kedua akan menampilkan data hasil dari proses *text preprocessing* dan pelabelan otomatis, selanjutnya pada tampilan halaman ketiga berisi hasil dari pengklasifikasian data dengan algoritma naïve bayes dengan menghasilkan sentimen komentar berupa positif, netral dan negatif. Pada tampilan halaman keempat akan menampilkan sebuah grafik. Dan tampilan terakhir akan menampilkan analisis sentimen komentar untuk memprediksikan komentar berupa label sentimen positif, netral ataupun negatif.

Berikut gambaran desain tampilan antarmuka pada platform aplikasi yang akan dibuat:

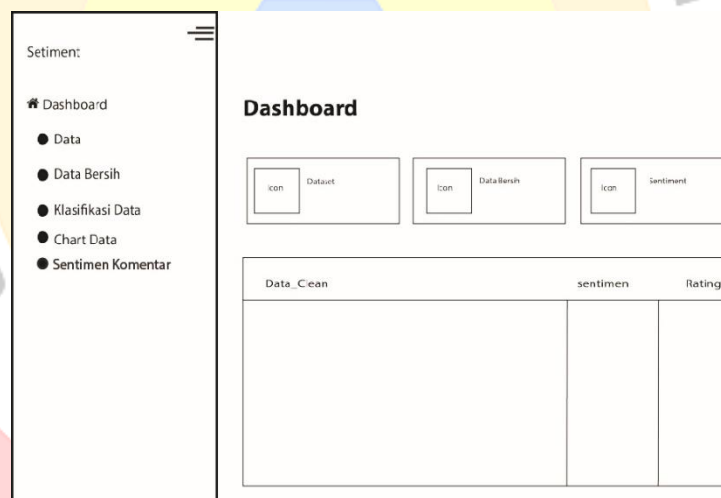


Pada Rancangan Tampilan gambar 3.12 yang akan menampilkan data ulasan komentar yang di dapatkan dari ruang guru dan menjadikan halaman *home page* pada saat pertama kali membuka pada website

Pada gambar 3.13 menampilkan hasil data dari proses *text preprocessing* dan juga label dengan *lexicon vader*.

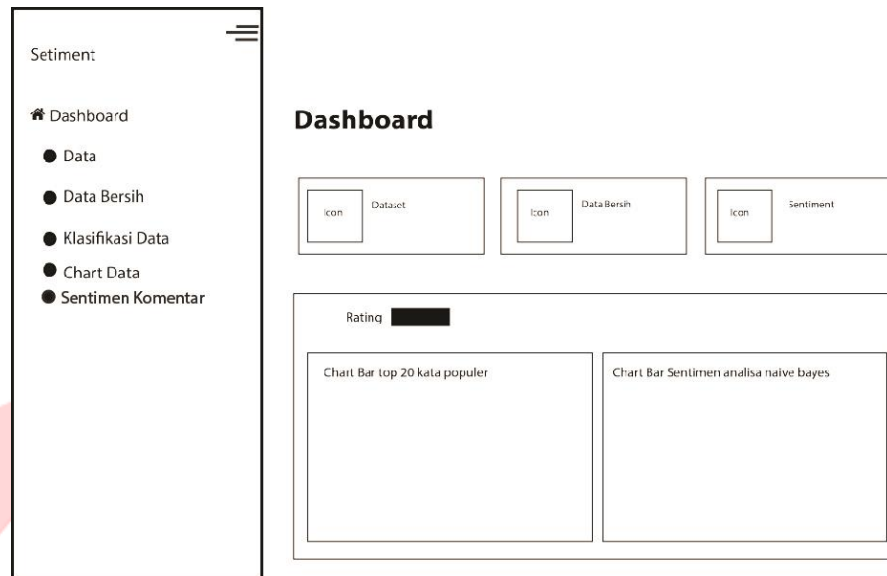
Gambar 3. 13 Rancangan Tampilan Data Preprocessing & Pelabelan

Pada gambar 3.14 Rancangan tampilan data hasil komentar yang di klasifikasi dengan model



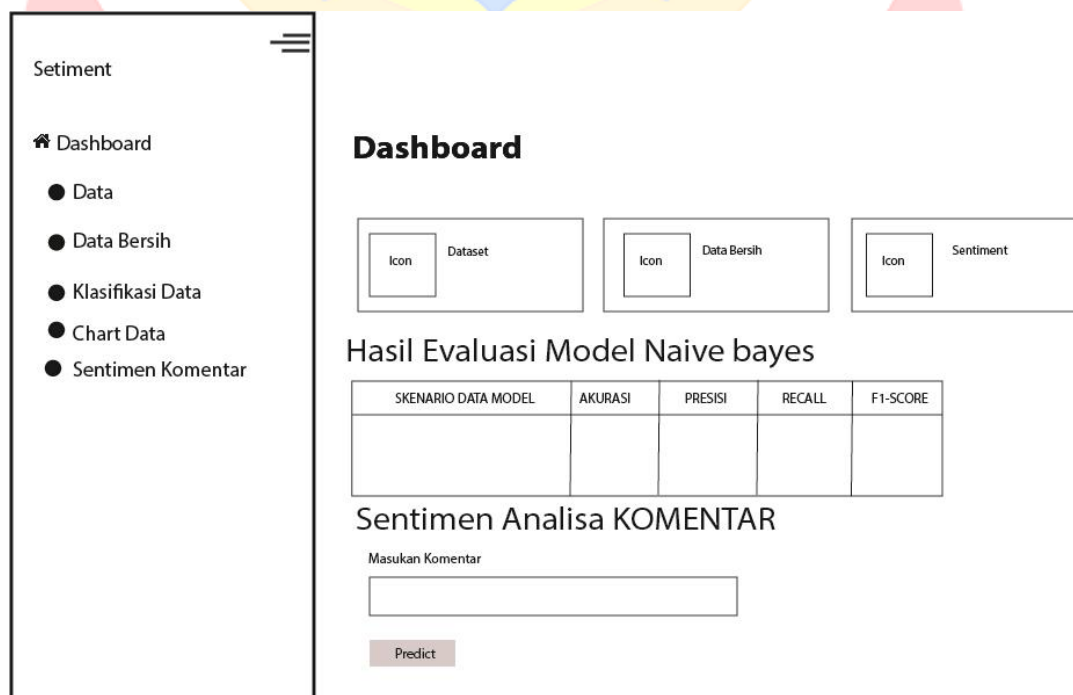
Gambar 3. 14 Rancangan Tampilan Klasifikasi *Naive Bayes*

*Naive Bayes*.

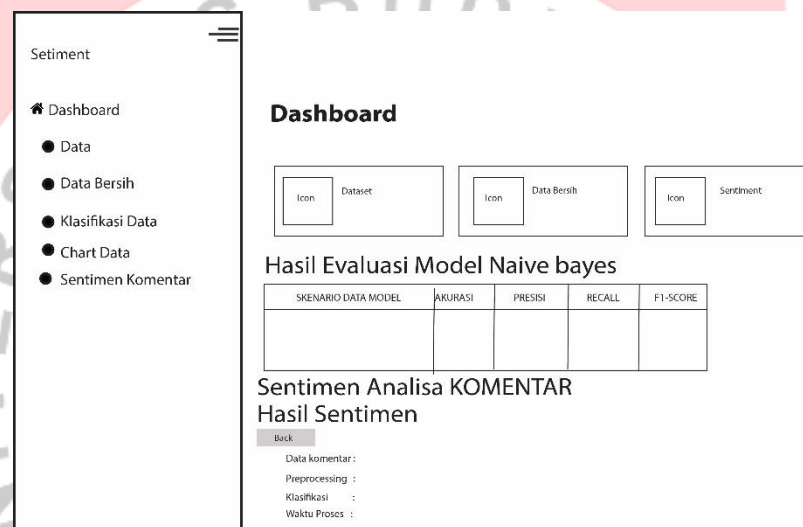


Gambar 3. 15 Rancangan Tampilan Grafik Sentimen

Pada gambar 3.15 Rancangan tampilan visual data dari banyaknya kata pada data komentar ulasan pengguna aplikasi ruang guru dan hasil grafik dari data hasil prediksi model Naive Bayes.



Pada gambar 3.16 menampilkan rancangan untuk hasil dari evaluasi model Naive Bayes dan juga analisa sentimen komentar yang akan di prediksi dengan model Naive Bayes.



Gambar 3. 17 Rancangan Tampilan Hasil *Predict* Sentimen Komentar

Pada gambar 3.17 rancangan yang menampilkan hasil dari prediksi analisa sentimen komentar dengan model Naive Bayes.

### 3.7. Jadwal Penelitian

Tabel 3. 16 *Ganchart* Penelitian

| No. | Kegiatan | Bulan 2025 |     |       |     |     |     |
|-----|----------|------------|-----|-------|-----|-----|-----|
|     |          | Jan        | Feb | Maret | Apr | Mei | Jun |

|   |                              |  |  |  |  |  |  |
|---|------------------------------|--|--|--|--|--|--|
| 1 | Tahapan Persiapan Penelitian |  |  |  |  |  |  |
|   | Persiapan Data               |  |  |  |  |  |  |
| 2 | Tahapan Pelaksanaan          |  |  |  |  |  |  |
|   | Model                        |  |  |  |  |  |  |
|   | Rancangan                    |  |  |  |  |  |  |
| 3 | Tahapan Penyusunan laporan   |  |  |  |  |  |  |