

**USULAN PERBAIKAN KUALITAS APLIKASI *REDDOORZ* BERDASARKAN
ULASAN PENGGUNA PADA *GOOGLE PLAYSTORE* MENGGUNAKAN *TEXT
MINING***

Dosen Pengampu

Dr. Ir. Agus Mansur, S.T., M.Eng.Sc., IPU



Nama : Wulan Purnamasari Hasanah

No. Mahasiswa : 21 916 035

PROGAM STUDI TEKNIK INDUSTRI

PROGAM MAGISTER

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2024

LEMBAR PERNYATAAN

Dengan ini saya mengakui bahwa karya ini adalah hasil kerja saya sendiri kecuali kutipan dan ringkasan yang setiap satunya telah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak kekayaan intelektual, maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Islam Indonesia.

Temanggung, 24 Juni 2024



Wulan Purnamasari Hasanah

NIM. 21916035

LEMBAR PENGESAHAN DOSEN PEMBIMBING

**USULAN PERBAIKAN KUALITAS APLIKASI *REDDOORZ*
BERDASARKAN ULASAN PENGGUNA PADA GOOGLE PLAYSTORE
MENGUNAKAN *TEXT MINING***

TESIS

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Magister Strata-
2 Pada Jurusan Teknik Industri Fakultas Teknologi Industri
Universitas Islam Indonesia**

Disusun Oleh:

Wulan Purnamasari Hasanah

21 916 035

Yogyakarta, 24 Juni 2024

Menyetujui,

Dosen Pembimbing



Dr. Ir. Agus Mansur, S.T., M.Eng.Sc., IPU

LEMBAR PENGESAHAN DOSEN PENGUJI

**USULAN PERBAIKAN KUALITAS APLIKASI *REDDOORZ*
BERDASARKAN ULASAN PENGGUNA PADA GOOGLE PLAYSTORE
MENGUNAKAN *TEXT MINING***

TESIS

Disusun Oleh:

Nama : Wulan Purnamasari Hasanah

No. Mahasiswa : 21916035

Fakultas/Jurusan : FTI/Teknik Industri

**Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk
memperoleh gelar Sarjana Strata-1 Teknik Industri**

Dr. Ir. Agus Mansur, S.T., M.Eng.Sc., IPU

Ketua

Dr. Drs. Imam Djati Widodo, M.Eng.Sc

Anggota 1

Dr. Harwati, S.T., M.T.

Anggota 2

Mengetahui,

Ketua Program Studi Teknik Industri

Universitas Islam Indonesia

Ir. Winda Nur Cahyo, S.T., M.Sc., Ph.D., IPM.

HALAMAN PERSEMBAHAN

Rasa syukur saya panjatkan kepada Allah SWT, atas limpahan karunia-Nya serta menjadi pribadi yang berilmu dan berakhlak sehingga dapat mencapai cita-cita. Tidak lupa karya ini saya persembahkan kepada kedua orang tua, kakak-kakak, sahabat, teman, dan seluruh pihak yang telah membantu dan memotivasi saya dalam menyelesaikan tesis ini.

Tesis ini saya persembahkan kepada diri saya sendiri, yang telah berjuang dan bertahan untuk menyelesaikannya dalam segala kondisi yang saya lewati. Untuk saya yang tidak pernah patah semangat dalam mengerjakan tesis ini, dan untuk saya yang terus bersyukur dan bersabar dalam tiap langkah yang dilewati untuk pengerjaan tesis hingga selesai.

MOTTO

“Hidup yang tidak dipertaruhkan tidak akan pernah dimenangkan” – **Sutan Sjahrir**

“Dan janganlah kamu (merasa) lemah, dan jangan (pula) bersedih hati, sebab kamu paling tinggi (derajatnya), jika kamu orang beriman” – **QS. Ali-Imran: 139**

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya” – **QS. Al-Baqarah : 286**

ABSTRAK

Internet memiliki pengaruh yang besar hampir dalam segala aspek, salah satunya yaitu *hospitality*. Reddoorz merupakan salah satu dampak perkembangan dari internet yang pesat. Penelitian ini bertujuan untuk mengusulkan perbaikan aplikasi reddoorz berdasarkan ulasan pengguna di Google Playstore menggunakan teknik *text mining*. Hal ini bertujuan untuk meningkatkan rating yang aplikasi tersebut, tudy ini mencakup terkait perbandingan akurasi antara *Naive Bayes Classifier* dan *Support Vector Machine* dalam mengklasifikasikan ulasan pengguna, serta penerapan algoritma *Association Rules* untuk mengidentifikasi akar masalah yang ditemukan dan dianalisis menggunakan *fishbone diagram*. Data yang digunakan mencakup 612 ulasan dengan kategori positif, negatif, dan netral dengan perbandingan akurasi *Naive Bayes Classifier* sebesar 85.41% dan *Support Vector Machine* sebesar 91.89%, menggunakan pembagian data uji sebesar 80% dan data latih sebesar 20%. Selain itu, algoritma *Association Rules* digunakan khususnya untuk analisis sentimen negatif dengan menghasilkan kata kunci seperti "aplikasi", "hotel", "RedDoorz", "pesan", "bayar", "uang", "refund", "kamar", "batal", dan "kecewa". Algoritma *Association Rules*, dengan fokus pada kata kunci yang telah ditetapkan, digunakan untuk mengungkapkan hubungan antara faktor-faktor yang ditemukan dalam ulasan dengan masalah-masalah yang teridentifikasi. Ini memberikan wawasan yang lebih dalam tentang penyebab akar masalah yang mempengaruhi pengalaman pengguna terhadap aplikasi RedDoorz. analisis hasil dilakukan menggunakan diagram fishbone untuk memvisualisasikan faktor-faktor utama yang menyebabkan masalah-masalah yang diidentifikasi. Pendekatan ini membantu pengembang dan manajemen aplikasi untuk memprioritaskan perbaikan berdasarkan pada temuan yang didukung oleh analisis data yang sistematis dan mendalam.

Kata kunci: Analisis sentiemn, *text mining*, *Naïve Bayes Classifier*, *Support Vector Machine*, *Association Rules*, *Fishbone Diagram*, RedDoorz

KATA PENGANTAR

Assalamu 'alaikum Warohmatullahi Wabarokatuh

Alhamdulillah, segala puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, serta shalawat dan salam yang selalu tercurah kepada Nabi Muhammad SAW sehingga penulis dapat menyusun dan menyelesaikan tesis yang berjudul **“USULAN PERBAIKAN KUALITAS APLIKASI REDDOORZ BERDASARKAN ULASAN PENGGUNA PADA GOOGLE PLAYSTORE MENGGUNAKAN *TEXT MINING*”** ini dengan baik dan lancar.

Tesis ini dilakukan sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan jenjang Strata-2 pada Jurusan Teknik Industri Universitas Islam Indonesia. Dalam pelaksanaan dan penyusunan Tesis, penulis banyak mendapatkan bantuan, dukungan, dan kesempatan dari berbagai pihak. Untuk itu penulis ingin mengucapkan terimakasih kepada :

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Winda Nur Cahyo, S.T., M.T., Ph.D. selaku Ketua Program Studi Magister Teknik Industri Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Dr. Ir. Agus Mansur, S.T., M.Eng.Sc., IPU selaku dosen pembimbing yang telah memberikan bimbingan kepada penulis dan memberikan motivasi yang sangat besar serta selalu bersabar kepada penulis.
4. Kepada kedua orang tua penulis, Nana Iriyana dan Suyami serta kakak-kakak, Jupa Hasanah, Lutfi Kumala, Yoko Yodi, dan Sofia Khasanah yang selalu memberikan doa, motivasi, dan dukungan moril maupun material kepada penulis.
5. Kepada teman-teman Bikini Buttom yang memberikan dukungan, saran, motivasi, keyakinan dan waktu untuk penulis dalam menyelesaikan tesis ini.

6. Kepada rekan kerja *E-commerce Reddoorz* yang memberikan motivasi untuk menyelesaikan tesis ini.
7. Kepada teman-teman Magister Teknik Industri angkatan 2021 Universitas Islam Indonesia yang menemani dan mendukung penulis.
8. Semua pihak yang telah membantu dalam penyelesaian tesis ini yang tidak dapat disebutkan satu-satu persatu.

Wassalamu'alaikum Warohmatullahi Wabarokatuh

Yogyakarta, 7 Juni 2023

(Wulan Purnamasari Hasanah)

DAFTAR ISI

LEMBAR PERNYATAAN.....	ii
LEMBAR PENGESAHAN DOSEN PEMBIMBING	iii
LEMBAR PENGESAHAN DOSEN PENGUJI.....	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	iv
MOTTO	v
ABSTRAK.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiii
BAB I.....	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah	17
1.3 Tujuan Penelitian	17
1.4 Batasan Penelitian	18
1.5 Manfaat Penelitian	18
1.6 Sistematika Penulisan.....	18
BAB II.....	20
2.1 Penelitian Terdahulu	20
2.2 Landasan Teori.....	29
2.2.1 <i>Reddoorz</i>	29
2.2.2 <i>Google Playstore Crawling</i>	29
2.2.3 Analisis Sentimen	30
2.2.4 <i>Text Mining</i>	30

2.2.5	<i>Pre-Processing Data</i>	30
2.2.6	<i>Sentimen Scoring</i>	31
2.2.7	<i>Stopword Removal</i>	32
2.2.8	<i>Stemming</i>	32
2.2.9	Pembobotan Kata TF-IDF	33
2.2.10	<i>Naïve Bayes Classifier</i>	34
2.2.11	<i>Support Vector Machine</i>	35
2.2.12	Evaluasi Kinerja Klasifikasi	36
2.2.13	<i>Wordcloud</i>	37
2.2.14	<i>Association Rules</i>	38
2.2.15	<i>Fishbone Diagram</i>	38
BAB III		40
3.1	Jenis dan Sumber Data	40
3.2	Variabel Penelitian	40
3.3	Prosedur Penelitian.....	40
BAB IV		44
4.1	Ekstrasi Ulasan.....	44
4.2	<i>Pre-Processing Data</i>	45
4.2.1	<i>Case Folding</i>	45
4.2.2	<i>Remove Number</i>	46
4.2.3	<i>Remove Punctuation</i>	47
4.2.4	Normalisasi kata	48
4.3	<i>Sentimen Scoring</i>	49
4.4	<i>Stopwords Removal</i>	51
4.5	<i>Stemming dan Tokenizing</i>	53
4.6	Pembobotan kata TF-IDF.....	54
4.7	Klasifikasi Sentimen	55

4.7.1	Data Latih dan Data Uji	55
4.7.2	Algoritma <i>Naïve Bayes Classifier</i>	55
4.7.3	Algoritma <i>Support Vector Machine</i>	56
4.7.4	Perbandingan Hasil Kinerja Klasifikasi	58
4.8	Visualisasi <i>Wordcloud</i>	59
4.9	<i>Association Rules</i>	60
4.10	Analisis Diagram <i>Fishbone</i>	69
4.10.1	Rekomendasi dari Analisis <i>Fishbone Diagram</i>	72
BAB V	74
5.1	Kesimpulan	74
5.2	Saran.....	75
DAFTAR PUSTAKA	76
LAMPIRAN	82

DAFTAR TABEL

Tabel 2.1 Kajian Induktif.....	23
Tabel 2.2 Kombinasi Aturan Awalan-Akhiran.....	33
Tabel 2.3 Indikator confusion matrix	36
Tabel 4.1 Sampel Casefolding	46
Tabel 4.2 Sample Remove Number	47
Tabel 4.3 Sample Remove Punctuation	48
Tabel 4.4 Sampel Normalisasi Kata	49
Tabel 4.5 Contoh Perhitungan Sentimen Scoring.....	50
Tabel 4.6 Sample Stopwords Removal.....	52
Tabel 4.7 Sample Stemming	53
Tabel 4.8 Sample Tokenizing	54
Tabel 4.9 Hasil Pembobotan Kata TF-IDF	55
Tabel 4.10 Hasil Confusion Matrix Model Naïve Bayes Classifier	56
Tabel 4.11 Hasil Confusion Matrix Model Support Vector Machine	57
Tabel 4.12 Hasil Akhir Nilai Akurasi	58
Tabel 4.13 Common Words.....	61
Tabel 4.14 Association Rules Aplikasi.....	62
Tabel 4.15 Association Rules Hotel	62
Tabel 4.16 Association Rules Reddoorz.....	63
Tabel 4.17 Association Rules Pesan	64
Tabel 4.18 Association Rules Bayar.....	65
Tabel 4.19 Association Rules Uang.....	65
Tabel 4.20 Association Rules Refund.....	66
Tabel 4.21 Association Rules Kamar	67
Tabel 4.22 Association Rules Batal	69
Tabel 4.23 Premises	70
Tabel 4.24 Rekomendasi Perbaikan.....	72

DAFTAR GAMBAR

Gambar 1.1 Ulasan Pengguna Pada Google Playstore	16
Gambar 2.1 Contoh Wordcloud.....	37
Gambar 2.2 Fishbone	39
Gambar 3.1 Alur Penelitian	41
Gambar 4.1 Scraping Ulasan Aplikasi Reddoorz	44
Gambar 4.2 Scrapping Ulasan Aplikasi Reddoorz Tahun 2024.....	45
Gambar 4.3 Sintaks Pre-processing data	45
Gambar 4.4 Sentimen Polarity.....	51
Gambar 4.5 Parameter Perfomansi	56
Gambar 4.6 Parameter Perfomansi Support Vector Machine.....	57
Gambar 4.7 Wordcloud Ulasan Sentimen Positif.....	59
Gambar 4.8 Wordcloud Ulasan Sentimen Negatif	60
Gambar 4.9 Fishbone Diagram Masalah Aplikasi.....	71

DAFTAR LAMPIRAN

Lampiran 1 Data Scrapping	82
Lampiran 2 Kamus Normalisasi	83
Lampiran 3 Kamus Indonesian Sentiment Lexicon Positif	84
Lampiran 4 Kamus Indonesian Sentiment Lexicon Negatif.....	85
Lampiran 5 Kamus Stopwords Indonesia	86
Lampiran 6 Sintaks Python Webscrapping.....	87
Lampiran 7 Sintaks Python Pre-Processing Data	88
Lampiran 8 TF-IDF	90
Lampiran 9 Naive Bayes Classifier	91
Lampiran 10 Support Vector Machine	92
Lampiran 11 Apriori	92

BAB I

PENDAHULUAN

1.1 Latar Belakang

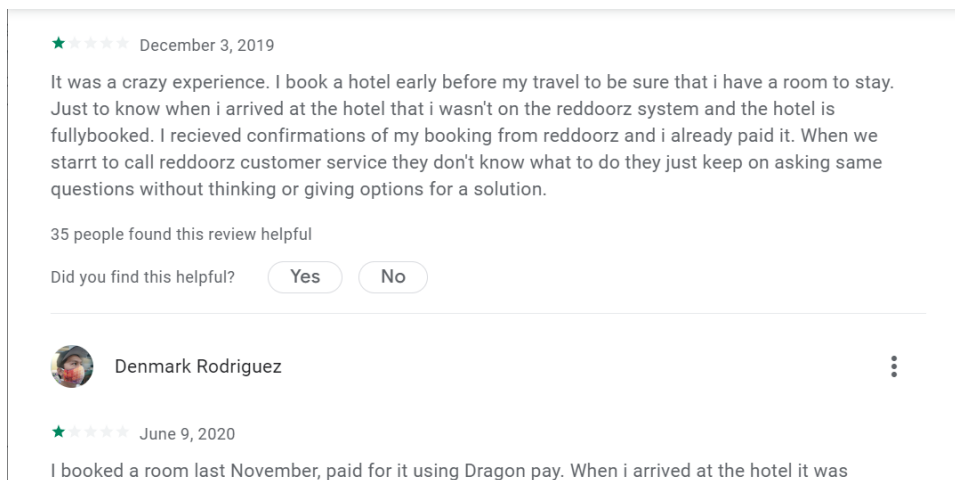
Aplikasi *mobile* yang telah berkembang pesat menjadi bagian penting untuk industri perhotelan, dimana para pengguna menggunakan hal tersebut untuk mencari, menemukan, dan memesan akomodasi yang mereka butuhkan. Pada era ini, pemakaian perangkat seluler semakin meluas, sehingga penting untuk melakukan perubahan salah satunya di sektor perhotelan. Penyedia pelayanan hotel juga harus memahami dan mengevaluasi sentimen pengguna terkait pengalaman mereka dengan aplikasi perhotelan.

Reddoorz adalah salah satu aplikasi atau platform untuk memesan hotel *budget* di berbagai negara, salah satunya Indonesia, dimana aplikasi ini menyediakan lebih dari 1000 properti yang bisa dipesan oleh *traveler* secara *online* maupun *offline*. Bisnis *Reddoorz* merupakan bentuk kerja sama dengan hotel-hotel kecil yang diintegrasikan kedalam satu aplikasi serta memberikan pelatihan teknologi, pelajaran dasar dalam melayani pelanggan dan perawatan ruangan agar seragam kepada staf. (Ginting, et al, 2023).

Kemudahan yang diberikan aplikasi ini memberikan kemudahan kepada pengguna dengan menyederhanakan proses pencarian, pemesanan, dan pembayaran bagi pengguna melalui aplikasi *mobile* ataupun *website Reddoorz*. Tidak hanya hal tersebut, kepuasan pengguna juga ditingkatkan melalui pemberian diskon kepada pengguna baru ataupun pengguna setia aplikasi *Reddoorz*.

Berbagai hal telah diberikan kepada konsumen, namun tetap diperlukan pemeliharaan untuk mempertahankan atau meningkatkan kepuasan konsumen. Salah satu cara dalam pemeliharaan tersebut adalah mengevaluasi kinerja dengan melihat keluhan dan opini konsumen. *Google Playstore* merupakan salah satu sarana untuk mengunduh aplikasi *mobile* seperti *Reddoorz*, pada layanan ini terdapat fitur untuk melihat dan memberikan ulasan terhadap aplikasi yang akan atau telah diunduh. Pengguna *google playstore* hanya dapat memberikan ulasan sekali pada setiap aplikasi

yang mereka unduh namun dapat merubah ulasan tersebut kapan saja (Google Play, 2023).



Gambar 1.1 Ulasan Pengguna Pada Google Playstore

Analisis sentimen (*opinion mining*) merupakan proses secara otomatis untuk mengontrol, mengekstrak, dan mengolah data teks yang bertujuan untuk memperoleh informasi sentimen dalam sebuah kalimat, sehingga dapat diketahui kecenderungan dalam polaritas data (Rozy et al., 2012). Analisis ini dapat membantu dalam meneliti opini dari pengguna aplikasi *Reddoorz*, untuk melihat kecenderungan opini berupa opini positif, negatif atau netral. Salah satu algoritma yang dapat membantu dalam analisis sentimen adalah *Naïve Bayes Classifier* dan *Support Vector Machine*. Algoritma *Naïve Bayes Classifier* menggunakan pendekatan yang didasari oleh nilai peluang (Buololo, 2020). Terdapat beberapa model dalam algoritma ini, namun dua model yang sering digunakan dalam pengolahan teks adalah *Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes*, dimana pada *Multinomial Naïve Bayes* model memperhatikan jumlah kata yang muncul dalam dokumen, sedangkan pada *Bernoulli Naïve Bayes* model hanya memperhatikan kemunculan kata pada dokumen tetapi mengabaikan jumlah kemunculannya (McCallun & Nigam, 1998). Algoritma *Naïve Bayes Classifier* banyak digunakan karena kemudahan dalam penggunaannya, waktu pemrosesan yang cenderung singkat, dan kemudahan dalam implementasi karena strukturnya yang sederhana, dan efisien (Taheri dan Mammadov, 2013). Algoritma *Support Vector Machine* bekerja dengan cara menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak antar kelasnya. Algoritma ini digunakan dalam pemrosesan bahasa yang alami karena memiliki kinerja yang baik dan mampu diterapkan dalam

tugas-tugas seperti pengenalan tulisan tangan, pengenalan suatu objek, dan klasifikasi teks (Cortes dan Vapnik, 1995). Sehingga aplikasi algoritma ini dapat digunakan pada kepentingan pekerjaan pengenalan pola seperti gambar tidak hanya klasifikasi teks.

Penelitian terkait analisis sentimen telah banyak dilakukan, salah satunya adalah penelitian yang meneliti tentang klasifikasi sentimen pemerintah terhadap penanganan COVID-19 menggunakan algoritma *Multinomial Naïve Bayes*, yang menghasilkan tingkat akurasi sebesar 74% (Yuyun et al, 1995). Penelitian lainnya merupakan implementasi algoritma *Support Vector Machine* pada penelitian analisis sentimen mengenai pelayanan Gojek dengan ulasan yang terdapat pada media sosial Twitter dan menghasilkan tingkat akurasi sebesar 79.19%.

Pada penelitian ini akan membahas analisis sentimen terhadap ulasan pengguna aplikasi *Reddoorz* pada *Google Playstore* dengan menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine* dan penggunaan algoritma *Association Rules* untuk mencari aturan atau kombinasi item yang sering terjadi pada suatu data. Tujuan penggunaan algoritma ini untuk mencari kata sebagai akar masalah yang akan digunakan untuk *root cause analysis* dengan *fishbone diagram*, sehingga dapat menjadi rujukan untuk rekomendasi perbaikan pada aplikasi *Reddoorz*.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, dapat diketahui bahwa rumusan masalah pada penelitian ini adalah :

1. Bagaimana perbandingan kinerja *Naïve Bayes Classifier* dan *Support Vector Machine* untuk hasil klasifikasi sentimen pada ulasan aplikasi *Reddoorz* di *Google Playstore*?
2. Bagaimana hasil dari algoritma *Association Rules* pada ulasan aplikasi *Reddoorz* di *Google Playstore*?
3. Apa rekomendasi yang diberikan berdasarkan analisis akar masalah sebagai rujukan perbaikan aplikasi *Reddoorz*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, berikut merupakan tujuan pada penelitian ini:

1. Mengetahui metode dengan nilai terbaik diantara *Naïve Bayes Classifier* dan *Support Vector Machine* dalam analisis sentimen ulasan aplikasi *Reddoorz* di *Google Playstore* berdasarkan nilai akurasi.
2. Mengetahui akar masalah yang dijadikan sebagai referensi rujukan perbaikan aplikasi *Reddoorz*.
3. Membuat rekomendasi yang tepat untuk perbaikan aplikasi *Reddoorz* berdasarkan masalah yang ditemukan.

1.4 Batasan Penelitian

Pada penelitian ini terdapat batasan penelitian yang bertujuan membatasi ruang lingkup pembahasan, berikut merupakan batasan pada penelitian ini :

1. Data yang digunakan berupa ulasan aplikasi *Reddoorz* di *Google Playstore* tanpa ada pengulangan.
2. Data yang digunakan hanya ulasan aplikasi yang dilakukan dari bulan Januari 2024 hingga Juni 2024.
3. Analisis sentimen yang dilakukan menggunakan metode metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* dengan menggunakan bahasa pemrograman *python*.
4. Klasifikasi sentimen menggunakan dua kategori, yaitu sentimen positif dan sentimen negatif

1.5 Manfaat Penelitian

Adapun manfaat yang dihasilkan dari penelitian ini adalah :

1. Penelitian ini bermanfaat bagi keilmuan teknik industri.
2. Penelitian ini dapat menjadi referensi guna penelitian selanjutnya terkait metode *Naïve Bayes Classifier* (NBC) ,*Support Vector Machine* (SVM), dan algoritma *Association Rules*.
3. Penelitian ini dapat digunakan sebagai rujukan untuk aplikasi *reddoorz* dalam melakukan perbaikan berdasarkan aspek-aspek keluhan mayoritas

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan agar mempermudah penulis dalam penyusunan Thesis :

BAB I : PENDAHULUAN

Bab ini berisikan uraian-uraian terkait latar belakang permasalahan, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan laporan tugas akhir.

BAB II : TINJAUAN PUSTAKA

Bab ini berisikan penjelasan tentang kajian pustakan dan landasan teori yang berkaitan dengan prinsip dasar materi yang diangkat sehingga dapat dijadikan dasaran dalam penelitian ini

BAB III : METODOLOGI PENELITIAN

Bab ini berisikan uraian terkait alur penelitian dan metode yang digunakan dalam penelitian secara detail.

BAB IV : PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini berisikan data-data yang didapat dan diperlukan dalam penelitian ini serta cara pengolahan data tersebut sehingga menghasilkan output yang diinginkan.

BAB V : PEMBAHASAN

Bab ini berisikan pembahasan dan penjelasan secara detail mengenai hasil pengolahan data yang telah didapatkan sehingga dapat digunakan untuk usulan perbaikan.

BAB VI : KESIMPULAN DAN SARAN

Bab ini berisikan simpulan dari hasil analisis yang dilakukan dan jawaban atas rumusan yang telah dibuat dan saran yang merupakan masukan bagi pihak terkait.

DAFTAR PUSTAKA

LAMPIRAN

BAB II

KAJIAN LITERATUR

2.1 Penelitian Terdahulu

Penelitian ini membahas terkait analisis data sentimen pada ulasan aplikasi *Reddoorz* di *google playstore* menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Oleh karena itu, dibutuhkan penelitian terdahulu yang dapat berbentuk buku dan jurnal sebagai pedoman atau informasi terkait topik yang diangkat. Pada penelitian terkait kinerja *machine learning* menggunakan metode *Naïve Bayes* dengan model Bernoulli dan Multinomial pada data komentar mengenai pemindahan ibu kota negara melalui aplikasi Youtube, penelitian menghasilkan performa sebesar 94,45% terhadap model Multinomial sehingga dapat dikatakan baik dalam melakukan klasifikasi teks untuk kelas sentimen (Wardani et al., 2020).

Selain penelitian tersebut, terdapat penelitian lain yang menggunakan klasifikasi *Naïve Bayes* dan *Support Vector Machine*, dimana kedua metode ini digunakan untuk membandingkan klasifikasi sinyal Sinyal elektromiografi permukaan (sEMG) untuk enam identifikasi Gerakan tangan yang berbeda menggunakan fitur berbasis transformasi wavelet diskrit (DWT) dari hasil penelitian ini diketahui bahwa *Support Vector Machine* memiliki tingkat akurasi lebih tinggi dengan nilai 95,8% dibandingkan *Naïve Bayes Classifier* (Narayan, 2021). Selanjutnya adalah penelitian analisis sentimen terhadap opini publik pada aplikasi twitter menggunakan *Support Vector Machine* dengan skenario pemisahan data 80:20 menghasilkan presisi 85%, recall 86%, dan f1-score 84% (Agustina et al., 2024)

Penelitian selanjutnya merupakan analisis sentimen menggunakan *Naïve Bayes Classifier* terhadap pendapat atau opini masyarakat pada aplikasi twitter terkait kampanye Anti-LGBT di Indonesia dengan hasil perolehan akurasi sebesar 86.43% (Fitri et al., 2019). Kemudian penelitian berikutnya adalah Analisis Sentimen Pada Perusahaan Penyedia Jasa Logistik J&T Menggunakan Algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine* oleh (Rahman, 2022) diketahui bahwa nilai akurasi

klasifikasi menggunakan *Multinomial Naïve Bayes* sebesar 72,80% dan *Support Vector Machine* sebesar 82,40%

Naïve Bayes Classifier juga dapat digunakan untuk pengklasifikasian dalam pembuatan sistem seperti penelitian yang berjudul *Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier* yang menghasilkan tingkat akurasi model sebesar 94,31% (Hubert et al., 2021). Penelitian selanjutnya adalah penggunaan metode *Support Vector Machine* dan *Multilayer Perceptron* untuk analisis sentimen pada tweet terkait keuangan dalam bahasa turki terhadap 17189 tweet dengan pembagian data 80:20 dengan memperoleh tingkat akurasi sebesar 89% dan 88% (Cam et al., 2024) Selanjutnya merupakan penelitian yang menganalisis terkait dampak buruk COVID-19 terhadap perekonomian Thailand yang bergantung pada komoditi pariwisata memberikan hasil akurasi maksimum sebesar 77,4% menggunakan algoritma *Support Vector Machine* (Leelawat, et al., 2022)

Penelitian selanjutnya merupakan penggunaan metode *Naïve Bayes Classifier* dalam mengklasifikasikan kepuasan pelanggan melalui ulasan daring pada *hospitality service* dengan jumlah ulasan 47,172 yang menghasilkan akurasi lebih akurat dan *reliable* sehingga dapat membantu untuk improvisasi dalam memberikan pelayanan (Sanchez-Franco et al., 2019). Kemudian penelitian dengan memperluas pengklasifikasi label menggunakan *Naïve Bayes Classifier* yang bertujuan untuk memanfaatkan hubungan yang berguna antar label untuk mencapai akurasi yang lebih baik dengan mempertimbangkan hubungan atau ketergantungan antar label. (Kim et al., 2020). Selanjutnya penelitian terkait analisis *data mining* untuk menganalisis artikel riset dengan topik *mobile learning* dengan data artikel yang diambil dari berbagai situs seperti wiley, Sciencedirect, SAGE, Cambridge, dan IEEE lalu diekstraksi dengan beberapa teknik yaitu *clustering*, *association rules*, *visualization*, dan *term frequency* yang menghasilkan kata yang paling sering muncul adalah “*learning*” dan “*patients*” (Salloum et al., 2019). Penelitian terakhir merupakan penelitian terkait riset analisis *data mining* menggunakan algoritma *association rules* mengenai pola kecelakaan lalu lintas dengan sumber data dari data kecelakaan lalu lintas yang terdapat di Kota Tasikmalaya dan Ciamis menghasilkan korelasi antara profesi dan usia pelaku dengan atribut profesi siswa dengan usia 16-30 tahun di Kota Tasikmalaya dan korelasi antara usia dan tingkat pendidikan dengan atribut tingkat pendidikan SMA dengan usia 16-30 tahun

untuk Kota Ciamis (Ruswati, Guffroni, & Rianto, 2018) Dari pemaparan penelitian terdahulu, dapat diketahui bahwa penelitian yang akan dilakukan berfokus pada analisis sentimen menggunakan *Naïve Bayes Classifier*, *Support Vector Machine*, dan *Association Rules*.

Tabel 2.1 Kajian Induktif

No	Judul	Peneliti	Hasil	Analisis Sentimen	<i>Naïve Bayes Classifier</i>	<i>Support Vector Machine</i>	<i>Association Rules</i>
1	Analisi Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi <i>Naive Bayes</i> Untuk Model Bernoulli dan Multinomial	Wardani, Nabila Surya, Prahutama A, Kartikasari (2020)	Penelitian ini mengenai pengklasifikasian teks menggunakan <i>Naïve Bayes Classifier</i> dengan model Multinomial menghasilkan tingkat akurasi sebesar 94,45%	✓	✓		
2	<i>Comparative analysis of SVM and Naive Bayes classifier for the SEMG signal classification</i>	Yogendra Narayan (2021)	Pada penelitian ini terkait klasifikasi sinyal elektromografi menggunakan metode <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i> menghasilkan tingkat akurasi <i>Support Vector Machine</i> lebih tinggi dengan nilai akurasi sebesar 95,8%			✓	✓

3	<i>The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentimen Analysis Using Support Vector Machine Algorithm</i>	C.A Nurhaliza Agustina, Rice Novita, Mustakim, Nesdi Evrilyan Rozanda (2024)	Pada penelitian ini mengenai analisis sentimen terhadap opini publik pada aplikasi twitter menggunakan <i>Support Vector Machine</i> yang menghasilkan presisi sebesar 85%, <i>recall</i> 86%, dan <i>Fi-score</i> 84%	✓	✓
4	<i>Sentimen Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm</i>	Veny Amalia Fitri, Rachmadita Andreswari, Hasibuan, Muhammad Azani Hasibuan (2019)	Penelitian ini terkait analisis sentimen menggunakan <i>Naïve Bayes Classifier</i> terhadap opini publik pada aplikasi twitter mengenai kampanye Anti-LGBT di Indonesia memberikan hasil tingkat akurasi sebesar 86.43%	✓	✓

5	<p>Analisis Sentimen Pada Perusahaan Penyedia Jasa Logistik J&T Menggunakan Algoritma Multinomial Naive Bayes dan Support Vector Machine</p>	<p>Helmi Aulia Rahman (2022)</p>	<p>Penelitian ini mengenai analisis sentimen terhadap opini masyarakat mengenai perusahaan penyedia jasa logistik J&T menggunakan Algoritma Multinomial <i>Naive Bayes</i> dan <i>Support Vector Machine</i> yang menghasilkan tingkat akurasi <i>Support Vector Machine</i> lebih tinggi dibandingkan <i>Naive Bayes</i> dengan tingkat akurasi sebesar 82.40%</p>	✓	✓	✓
6	<p><i>Classifying Promotion Images Using Optical Character Recognition and Naive Bayes Classifier</i></p>	<p>Hubert, Peter Phoenix, Richard Sudaryono, Derwin Suhartono (2021)</p>	<p>Penelitian ini menjelaskan terkait <i>Naive Bayes Classifier</i> yang digunakan dalam pembuatan sistem klasifikasi foto dengan hasil akursi model sebesar 94.31%</p>	✓	✓	

- | | | | | | |
|---|--|--|--|---|---|
| 7 | <i>Sentimen analysis of financial Twitter posts on Twitter with the machine learning classifiers</i> | Handan Cam, Alper Veli Cam, Ugur Demirel, Sana Ahmed (2024) | <p>Penelitian ini menjelaskan terkait penggunaan <i>Support Vector Machine</i> dalam melakukan analisis sentimen terhadap keuangan dalam bahasa turki pada aplikasi twitter dengan jumlah tweet sebanyak 17189 menghasilkan tingkat akurasi sebesar 89%.</p> | ✓ | ✓ |
| 8 | <i>Twitter data sentimen analysis of tourism in Thailand during the COVID-19 pandemic using machine learning</i> | Natt Leelawat, Sirawit Jariyapongpaiboon, Arnon Promjun, Samit Boonyarak, Kumpol Saengtabtum, Ampan Laosunthara, Alfian Kurnia Yudha, Jing Tang (2022) | <p>Penelitian ini menjelaskan penggunaan algoritma <i>Support Vector Machine</i> dalam melakukan analisis sentimen dampak buruk COVID-19 terhadap perekonomian Thailand melalui aplikasi twitter yang menghasilkan tingkat akurasi sebesar 77,4%</p> | ✓ | ✓ |

9	<p><i>A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services</i></p>	<p>Manuel J.Sanchez-Franco, Antonio Navarro-Garcia, Rondan-Cataluna, Francisco Javier (2019)</p>	<p>Penelitian ini mengenai penggunaan metode <i>Naïve Bayes Classifier</i> dalam mengklasifikasikan kepuasan pelanggan melalui ulasan pada <i>hospitality service</i> yang menghasilkan akurasi lebih akurat dan <i>reliable</i> sehingga dapat membantu improvisasi dalam pelayanan</p>	✓	✓
10	<p><i>Multilabel naïve Bayes classification considering label dependence</i></p>	<p>Hae-Cheon Kim, Jin-Hyeong Park, Dae-Won Kim, Jaesung Lee (2020)</p>	<p>Penelitian ini memanfaatkan hubungan yang berguna antar label dalam pengklasifikasian label menggunakan <i>Naïve Bayes Classifier</i> untuk mencapai akurasi yang lebih baik dengan mempertimbangkan hubungan atau ketergantungan antar label.</p>		✓

- 11 *Text Mining dan Asscoiation Rules* S.A. Salloum, M. Al-Emran, A.A. Monem, K. Shaalan (2019) Penelitian ini bertujuan untuk menganalisis kata yang sering muncul menggunakan algoritma *Association Rules* dengan topik *mobile learning*, dimana pada topik tersebut diketahui bahwa kata yang sering "*learning*" dan "*patients*". ✓ ✓
- 12 *Association Rules dan Algoritma Apriori* Ruswati, A. Guffroni, R. Rianto (2018) Penelitian ini dilakukan pada pola kecelakaan lalu lintas yang diketahui bahwa kota Tasikmalaya memiliki relasi antara profesi dan usia dari pelaku. Sedangkan relasi pada kota Ciamis adalah antara usia dan tingkat pendidikan ✓
-

2.2 Landasan Teori

2.2.1 Reddoorz

Reddoorz merupakan jaringan penginapan *budget* online yang tersebar di berbagai negara, salah satunya Indonesia. *Reddoorz* berfokus pada pengembangan penginapan serta distribusi penjualan, jaringan ini memilih untuk bekerja sama dengan hotel yang memiliki potensi untuk berkembang dan dapat terikat komitmen dengan *Reddoorz*. Setelah terikat kontrak dengan brand *Reddoorz*, kualitas hotel akan ditingkatkan secara permanen dengan memenuhi standar utama brand beserta kelengkapan fasilitas. *Reddoorz* bertanggung jawab dalam membantu penjualan kamar secara *online* dan *offline* dengan skala kecil atau besar seperti *corporate sales program* (Kresmanto et al., 2020).

Aplikasi *Reddoorz* telah digunakan oleh banyak masyarakat karena kemudahan dan *range* harga yang diberikan. Jaringan aplikasi *Reddoorz* semakin meluas dengan lebih dari 1000 properti di Indonesia. *Reddoorz* juga mengembangkan jaringannya tidak hanya pada hotel budget, namun terdapat hotel bintang 3, apartemen, villa, dan kos.

2.2.2 Google Playstore Crawling

Google Playstore adalah layanan distribusi aplikasi resmi oleh *Google*, dimana *google playstore* menyediakan berbagai aplikasi yang bisa diunduh oleh pengguna dan menyediakan layanan bagi pengembang untuk mengunggah aplikasi. (Latif, et al., 2019). *Crawling* merupakan proses pengambilan data pada penyimpanan lokal dengan menggunakan kata kunci. Cara kerja *crawling* secara otomatis berdasarkan kata kunci yang diberikan oleh pengguna (Saputra, 2019).

Google Playstore crawling adalah teknik *crawling* yang bertujuan untuk mendapatkan informasi dari *playstore*, melalui penyerapan konten ataupun ulasan yang akan diukur dan dianalisis (Bagnasco, et al., 2015), dimana pada *google playstore* terdapat 196.651 ulasan dari pengguna yang dapat dianalisis. Proses *crawling* pada *google playstore* dapat dilakukan dengan *python*.

2.2.3 Analisis Sentimen

Analisis sentimen merupakan bidang studi yang bertujuan untuk menganalisis opini, sentimen, penilaian, sikap, dan emosi seseorang terhadap sebuah sentitas dan atribut yang diekspresikan kedalam bentuk tulisan atau teks. Entitas yang dimaksud dapat berupa produk, layanan, organisasi, individu, acara, isu, atau topik. Hasil analisis yang didapatkan akan digunakan untuk mengidentifikasi pendapat positif dan atau pendapat negatif yang secara tersurat atau tersirat dalam sebuah teks (liu, 2015)

Tujuan utama dari analisis ini adalah membedakan antara kalimat subjektif dan objektif. Klasifikasi subjektivitas dilakukan dengan cara membedakan kalimat yang mengungkapkan informasi objektif (kalimat objektif), dimana kalimat tersebut mengungkapkan pandangan dan pendapat subjektif (kalimat subjektif). Salah satu contoh kalimat objektif adalah “Reddoorz adalah layanan pemesanan hotel” sedangkan kalimat subjektifnya “Reddoorz itu murah” (Pozzi et al, 2017)

2.2.4 Text Mining

Text mining merupakan proses dalam mengekstraksi informasi dari sebuah sumber berupa data yang tidak terstruktur, dimana tujuan dari proses ini adalah untuk menangani masalah *classification*, *clustering*, *information extraction*, dan *information retrieval* (Feldman & Sanger, 2007)

Langkah ini dibagi menjadi tiga tahapan, yaitu proses awal terhadap text (*text preprocessing*), *text transformation/feature generation*, dan penemuan pola (*pattern discovery*). Langkah pertama dalam *text mining* adalah memasukan data text yang akan menghasilkan luaran berupa pola sebagai hasil interpretasi.

2.2.5 Pre-Processing Data

Pre-processing data adalah suatu proses dalam mempersiapkan data mentah sebelum diolah untuk proses selanjutnya dengan cara mengeminilasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah untuk dipahami sistem. Proses ini sangat penting untuk analisis sentimen, terutama pada media sosial yang memiliki kata-kata tidak formal, tidak terstruktur, dan mempunyai *noise* yang besar (Mujilahwati, 2016).

Tahapan *pre-processing* data adalah tahapan awal dimana data akan diekstraksi berdasarkan keadaan data yang dimiliki. Berikut merupakan tahapan *pre-processing data* yang dilakukan pada penelitian *text mining* :

1. *Case Folding*

Case Folding adalah tahapan yang memiliki tujuan untuk merubah keseluruhan teks dari huruf kecil menjadi besar ataupun sebaliknya agar seragam pada suatu dokumen (Mujilahwati, 2016)

2. *Cleaning*

Cleaning merupakan tahapan dimana karakter-karakter yang tidak dibutuhkan diluar alphabet a-z, tanda baca, angka, dan *emoticon* akan dihapuskan. Proses ini bertujuan untuk membersihkan data dari karakter tidak penting agar tidak mempengaruhi proses klasifikasi (fatmawati & Affandes, 2017)

3. Normalisasi Kata

Tahapan ini merupakan langkah untuk mengubah kata yang tidak baku menjadi baku sehingga sesuai pedoman Kamus Besar Bahasa Indonesia (Rozy et al., 2019) Salah satu contohnya adalah penggunaan bahasa gaul atau sehari-hari oleh pengguna saat memberi ulasan seperti “murce” , pada proses normalisasi kata akan diubah menjadi bahasa baku yaitu murah.

2.2.6 Sentimen Scoring

Sentimen *Scoring* adalah teknik dalam pelabelan suatu pernyataan, atau dapat dikatakan pernyataan tersebut dapat digolongkan menjadi sentimen positif atau negative. Tahapan ini membutuhkan bantuan dari kamus *boosterwords* dan kamus negasi. *Boosterwords* merupakan kata yang bertujuan untuk meningkatkan atau menurunkan intensitas sentimen kata disebelahnya sedangkan negasi merupakan kata yang terdapat dalam suatu kalimat yang bisa mengubah sebuah opini contohnya kata “nyaman” yang memiliki sentimen positif namun jika didahului kata negasi “tidak” maka susunan kata menjadi “tidak nyaman” yang bermakna negative (Wahid & Azhari, 2016)

Selain dua kamus yang telah disebutkan, dibutuhkan kamus sentimen dalam proses penentuan kelas sentimen pada sebuah dokumen, dimana kamus ini berisi kumpulan kata yang telah diberi bobot dengan kekuatan sentimen 1 (tidak mempunyai sentimen positif) hingga 5 (mempunyai sentimen positif sangat kuat) dan -1 (tidak mempunyai

sentimen negative) hingga -5 (mempunyai sentimen negative yang sangat kuat) (Wahid & Azhari, 2016) Kamus sentimen ini didapatkan dari hasil translasi kamus sentimen berbahasa Inggris yang mengalami penambahan dan pengurangan kata berdasarkan pengamatan dalam proses pengembangan sistem.

2.2.7 Stopword Removal

Pada sebuah text terdapat banyak kemungkinan yang terjadi, salah satunya adalah adanya kata umum yang sering muncul namun tidak memiliki arti sehingga tidak memiliki pengaruh atau tidak memberikan informasi pada kalimat tersebut. Hal tersebut perlu dihilangkan karena dapat menurunkan tingkat akurasi dalam sistem klasifikasi (Ratiasasadara, 2022) *Stopword Removal* merupakan proses menghilangkan kata-kata irelevan dalam proses analisis sentimen. Kata yang dimaksud adalah kata yang sering digunakan namun tidak memiliki arti (Khosmah & Ariwibowo, 2020). Berikut merupakan contoh kata yang masuk dalam daftar *stopword* :

Table 2.1 Contoh Kata *Stopwords*

<i>Stopword</i> Indonesia	<i>Stopwords</i> Manual
Aku	Deh
Di	Halo
dan	Tuh
Sejak	Wow
belum	loh

2.2.8 Stemming

Stemming adalah proses mengubah kata berimbuhan menjadi kata dasar (Alhanin & Juzaidinn, 2011). Salah satu dalam *stemming* adalah *Dictory Base Stemming* dengan mendekati struktur morfologi dalam suatu bahasa yang terdapat aturan imbuhan untuk mengestrak kata dasar. Pola aturam aturan imbuhan dalam Bahasa Indonesia memiliki beberapa variative seperti prefix, sufiks, konfiks, dan infiks dengan segmentasi sebagai berikut ;

Tabel 2.2 Kombinasi Aturan Awalan-Akhiran

Awalan	Akhiran yang tidak diizinkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan

2.2.9 Pembobotan Kata TF-IDF

Pembobotan kata digunakan untuk memberikan bobot pada setiap kata (*term*) yang bertujuan untuk memudahkan proses klasifikasi (Nurjanah & Astusti, 2013) Terdapat berbagai metode untuk pembobotan kata salah satunya adalah *Term Frequency Inverse Document Frequency* (TF-IDF) dengan menggunakan modul *scikit-learn*. Metode ini dilakukan dengan cara melakukan perhitungan TF dan nilai IF dari sebuah kata dalam suatu tesk, dimana *Term Frequency* merupakan perhitungan pembobotan kata dengan menghitung frekuensi kata tertentu yang muncul pada teks, perhitungan ini menggunakan modul *scikit-learn* yang dibantu *software python* (Developers, 2022). Sedangkan *Inverse Document Frequency* (IDF) bertujuan untuk mengurangi bobot dari kata yang frekuensinya terlalu tinggi dalam teks. Kata yang jarang muncul pada sebuah teks memiliki nilai IDF yang lebih besar dibandingkan dengan kata yang sering muncul (Deolika & Taufik Lutfhi, 2019). Pada penelitian ini perhitungan *Inverse Document Frequency*) menggunakan *smooth_idf = True*, sehingga perhitungan bobot *Inverse Document Frequency* menggunakan persamaan :

$$idf_{t,d} = tf_{t,d} \times \left[\ln \left(\frac{N+1}{df_t+1} \right) + 1 \right] \dots\dots\dots(2.1)$$

Dengan :

$idf_{t,d}$: *Inverse Document Frequency*

$tf_{t,d}$: *Term Frequency*

N : jumlah seluruh dokumen

df_t : jumlah dokumen yang mengandung kata

Sehingga dari persamaan (x), perhitungan *Term Frequency Inverse Document Frequency* (TF-IDF) dengan menggunakan modul *scikit-learn* dapat dilakukan dengan persamaan berikut :

$$W_{t,d} = \frac{idf_{t,d}}{\sqrt{\sum_{d=1}^n idf_{t,d}^2}} \dots\dots\dots(2.2)$$

Dengan :

$W_{t,d}$: *Term Frequency Inverse Document Frequency*

$idf_{t,d}$: nilai *Inverse Document Frequency*

$\sum_{d=1}^n idf_{t,d}^2$: jumlah nilai IDF pada dokumen ke d

2.2.10 *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah salah satu teknik dalam *data mining*, dimana metode ini menggunakan prinsip probabilitas untuk memecahkan masalah. Metode ini digunakan dalam *data mining* karena mudah digunakan, waktu pemrosesan singkat, mudah diimplemtasikan dengan struktur yang sederhana, dan efisien (Taheri & Mammadov, 2013).

Naïve Bayes Classifier muncul karena teorema *bayes* yang mengacu pada konsep probabilitas bersyarat (Tan et al., 2006), dimana maksud dari probabilitas bersyarat adalah probabilitas yang variabel acaknya akan mengambil nilai tertentu mengingat hasil dari variabel acak lainnya diketahui (Adam et al., 2002). Berikut merupakan teorema bayes secara matematis :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \dots\dots\dots(2.3)$$

Dengan $P(A \cap B)$ merupakan probabilitas interaksi A, sedangkan $P(B)$ adalah probabilitas B. Sehingga nilai $P(A \cap B) = P(A|B)P(B)$. Selanjutnya nilai $P(A \cap B)$ disubstitusikan sehingga diperoleh persamaan baru sebagai berikut :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \dots\dots\dots(2.4)$$

Dengan :

A : Sampel data dengan label kelas tidak diketahui

B : Kelas hasil klasifikasi

$P(A|B)$: Probabilitas terjadinya A jika B diketahui

$P(B|A)$: Probabilitas terjadinya B jika A diketahui

$P(A)$: Probabilitas prior A yang mendahului terjadinya B

$P(B)$: Probabilitas prior B dan bertindak sebagai *normalizing constant*

Naïve Bayes Classifier memiliki beberapa model, namun model yang digunakan untuk mengoptimalkan klasifikasi dalam bentuk teks dengan memperhitungkan jumlah kemunculan setiap kata pada dokumen adalah *Multinomial Naïve Bayes* (Hidayah & Sahibu, 2021).

2.2.11 Support Vector Machine

Support Vector Machine merupakan pendekatan pembelajaran mesin modern untuk mengintegrasikan data neuroimaging, dengan vektor pendukung berupa model pembelajaran terkait yang memeriksa data yang digunakan untuk investigasi regresi dan klasifikasi. Metode ini secara efektif melakukan klasifikasi non-linier selain klasifikasi linier secara implisit menerjemahkan input data ke dalam ruang fitur berdimensi tinggi dengan cara menarik margin yang bertujuan agar memiliki jarak terpendek antara garis dan kelas sehingga dapat meminimalkan kesalahan saat klasifikasi (Mahesh, 2019). Hasil dari metode ini seimbang, bahkan untuk studi dengan jumlah sampel yang dibatasi namun tetap dapat menangani klasifikasi dengan baik (Pisner & Schnyer, 2019)

Metode ini merumakan *machine learning* dengan *supervised learning*, artinya *Support Vector Machine* bekerja dengan mendapatkan sampel yang telah memiliki kategori untuk melakukan pengujian data, sehingga metode ini dapat mengatasi permasalahan pada klasifikasi regresi, namun pengembangannya saat ini dapat melakukan pengujian dengan sebagian data yang terindetifikasi tidak memiliki label yang dikategorikan *semi-supervised* (B et al., 2023).

2.2.12 Evaluasi Kinerja Klasifikasi

Evaluasi Kinerja Klasifikasi merupakan pengukuran performa dan klasifikasi yang telah dibuat. Pengukuran ini dilakukan dengan cara melihat tingkat akurasi dari prediksi klasifikasi yang telah dihasilkan. Proses ini dibantu menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang berfungsi untuk menganalisis hasil klasifikasi dan mengenali *tuple* dari kelas yang berbeda (Han et al., 2012), dimana matriks ini terdiri dari beberapa indikator yaitu *true positive* (TP) untuk data positif yang diprediksi positif, *false negative* (FN) untuk data positif yang diprediksi negatif, *true negative* (TN) untuk data negatif yang diprediksi negatif, dan *False positive* (FP) untuk data negatif yang diprediksi positif. Berikut merupakan tabel x yang menunjukkan indikator pada *confusion matrix* :

Tabel 2.3 Indikator *confusion matrix*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	TP	FN
	Negatif	FP	TN

Evaluasi Kinerja Klasifikasi memiliki parameter yaitu *accuracy*, *precision*, dan *recall*. *Accuracy* merupakan nilai proporsi dari jumlah kata tiap kelas yang dikelompokkan benar dari total semua data, parameter ini menjadi bagian penting pada akhir pengolahan untuk mendapatkan hasil yang mendekati sebenarnya. *Precision* adalah nilai proporsi dari observasi positif yang diprediksi benar dari seluruh observasi positif yang diprediksi. *Recall* (sensitifitas) adalah nilai proporsi dari pengamatan positif yang diprediksi benar untuk semua pengamatan di kelas sebenarnya. Berikut merupakan tabel x yang menjelaskan terkait ukuran evaluasi sistem klasifikasi :

Table 2.2 Ukuran Evaluasi Sistem Klasifikasi

Ukuran	Rumus
<i>Overall accuracy</i>	$\frac{TP + TN}{TP + FP + TN + FN}$
<i>Precision</i>	$\frac{TP}{TP + FP}$
<i>Recall</i>	$\frac{TP + TN}{TN + FN}$

2.2.13 Wordcloud

Wordcloud adalah suatu metode yang bertujuan untuk menampilkan data teks secara visual., dimana semakin besar ukuran kata maka menunjukkan semakin besar frekuensi kata tersebut muncul dalam sebuah teks (Ramasubramanian & Singh, 2017). Visualisasi *wordcloud* akan mempermudah untuk mengamati gagasan dan pendirian dari penulis teks sehingga dapat menjadi alat bantu untuk menganalisis sebuah teks atau wacana tertulis.



Gambar 2.1 Contoh Wordcloud

2.2.14 Association Rules

Association rules dalam data mining dilakukan dengan tujuan untuk mengekstraksi korelasi yang unik, pola yang sering terjadi, asosiasi, atau struktur sederhana dari data set dalam *database* atau *repository* (Kotsiantis & Kanellpoulos, 2006). Hal ini digunakan untuk memenuhi ketentuan minimum *support* dan ketentuan minimum *confidence* dalam waktu yang bersama (I. A. Khan et al., 2015). Dalam *association rules* terbagi atas tiga perhitungan dasar yaitu penentuan nilai *support*, *confidence*, dan *lift ratio*. Dengan rumus sebagai berikut :

1. Support

Nilai *support* dari sebuah item :

$$\text{Support (A)} = \frac{\text{Jumlah transaksi yang mengandung A}}{\text{Total transaksi}} \times 100$$

Nilai *support* yang merupakan kombinasi dari dua item :

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Total transaksi}} \times 100$$

2. Confidence

Nilai *confidence* dari $A \rightarrow B$:

$$\text{Confidence} = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah transaksi yang mengandung A}} \times 100$$

3. Lift ratio

Sebelum menghitung *lift ratio*, dibutuhkan nilai ekspektasi *confidence* dengan formulasi sebagai berikut :

$$\text{Ekspetasi Confidence} = \frac{\text{Jumlah transaksi yang mengandung B}}{\text{Total transaksi}} \times 100$$

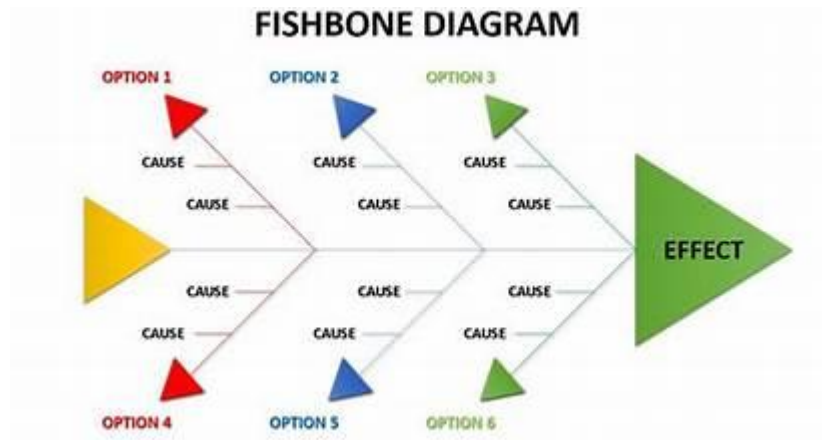
Hasil perhitungan dari *Ekspetasi confidence* tersebut digunakan sebagai perbandingan dengan nilai *confidence* :

$$\text{Lift ratio} = \frac{\text{Confidence}}{\text{Ekspetasi Confidence}}$$

2.2.15 Fishbone Diagram

Fishbone diagram merupakan sebuah alat yang digunakan untuk mengidentifikasi akar penyebab dari sebuah masalah yang berhubungan dengan pengendalian kualitas (Ilie & Ciocoiu, 2010). Metode ini ditemukan oleh Kaoru Ishikawa pada tahun 1990 sehingga sering disebut *Ishikawa diagram*, dimana tujuan dari diagram ini adalah untuk meningkatkan performa tim dalam menganalisis sebab dan akibat dari suatu masalah. Diagram ini memiliki bentuk seperti tulang ikan, dengan bagian kepala

diinterpretasikan sebagai *effect* atau masalah, sedangkan bagian tulang ikan digambarkan sebagai *cause* atau kemungkinan penyebab dari suatu masalah.



Gambar 2.2 Fishbone

BAB III

METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data

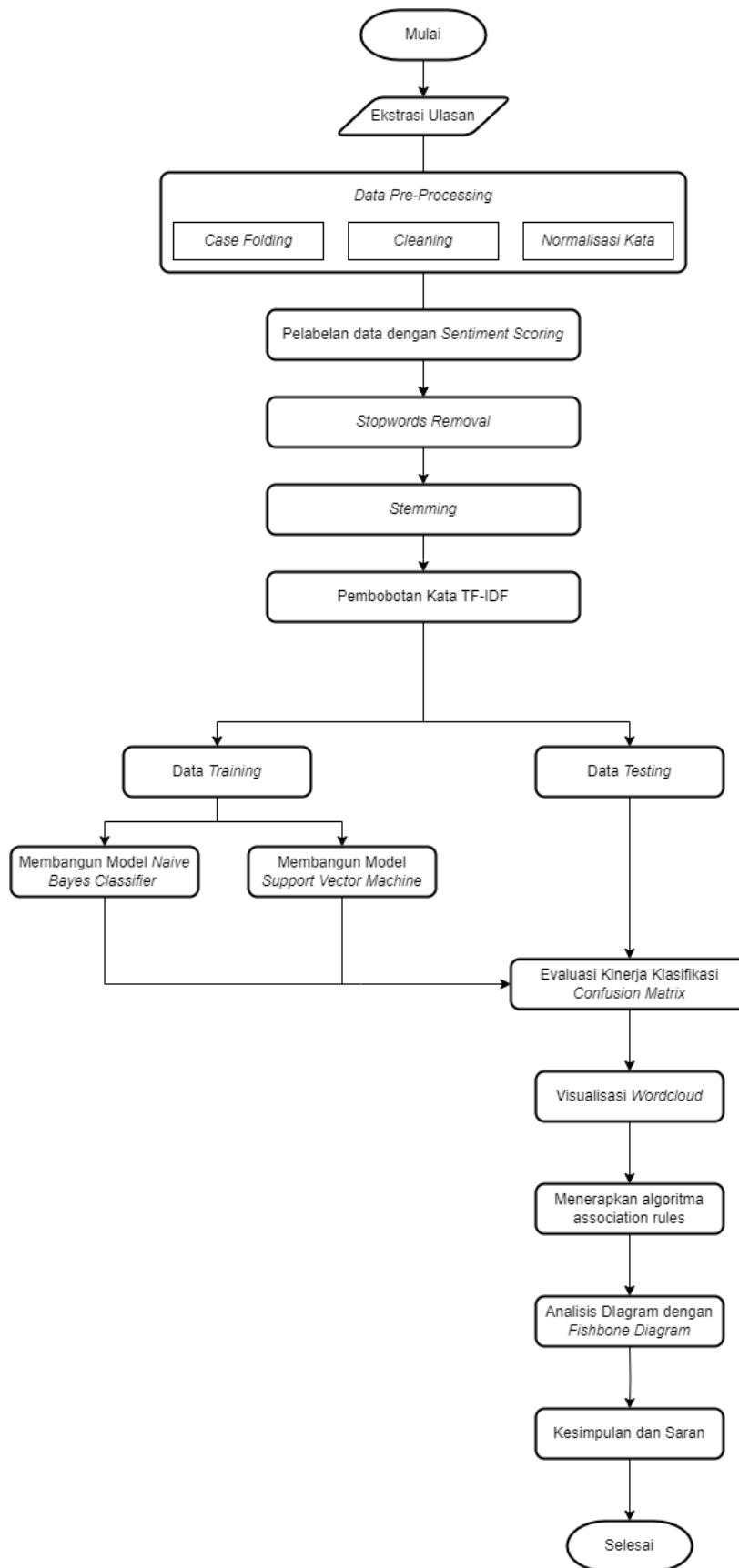
Data yang digunakan dalam penelitian ini adalah data kualitatif yang didapatkan dari proses *crawling* pada *Google Playstore* terkait ulasan aplikasi *Reddoorz*.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terbagi menjadi dua, yaitu variabel bebas dan variabel terikat. Variabel bebas terdiri dari keluhan atau opini pelanggan terhadap aplikasi *Reddoorz*, sedangkan variabel terikat berupa kelas sentimen (positif atau negative) dari opini pelanggan.

3.3 Prosedur Penelitian

Pada penelitian ini terhadap beberapa tahapan yang dilalui sebagai berikut :



Gambar 3.1 Alur Penelitian

Alur di atas merupakan gambaran tahapan yang dilakukan pada penelitian ini, berikut adalah penjelasan dari setiap tahap tersebut :

1. *Extracing Ulasan*

Langkah pertama dalam penelitian ini yaitu mengambil data ulasan berbahasa Indonesia dari *Google Playstore* yang selanjutnya dideteksi duplikatnya.

2. *Pre-processing data*

Setelah data diambil dan akan dianalisis lebih lanjut data akan berada ditahapan *pre-processing* berupa *case folding*, *cleaning*, dan normalisasi kata agar data mudah untuk dianalisis.

3. Sentimen *scoring*

Pada tahap ini data akan diberikan label kelas yaitu positif atau negatif setelah melewati tahapan *pre-processing data*.

4. *Stopwords removal*

Tahapan ini dilakukan untuk menghilangkan kata-kata yang tidak memiliki makna atau memiliki informasi.

5. *Stemming* dan *Tokenizing*

Tahapan ini merupakan pengubahan kata menjadi kata dasar, sehingga kata yang digunakan dapat melalui proses seleksi fitur.

6. Pembobotan Kata TF-IDF

Tahapan ini dilakukan dengan bantuan *scikit-learn modul* menggunakan data hasil.

7. Pembagian Data latih dan Data Uji

Pada tahap ini data akan dibagi menjadi 2 yaitu data latih dan data uji untuk tahap selanjutnya.

8. Klasifikasi menggunakan *Naïve Bayes Classifier*

Pada tahapan ini dibagi atas beberapa langkah, yaitu :

- a. Menghitung probabilitas prior didasari data *training*
- b. Menghitung probabilitas *term* terhadap kelas sentimen
- c. Menentukan hasil klasifikasi berdasarkan probabilitas dengan nilai kelas terbesar

9. Klasifikasi menggunakan *Support Vector Machine*

Pada tahapan ini dibagi atas beberapa langkah, yaitu :

- a. Menghitung estimasi parameter terbaik

- b. Membangun fungsi *hyperplane* dengan parameter terbaik
- 10. Mengevaluasi hasil kinerja Klasifikasi menggunakan *Confusion Matrix*
- 11. Interpretasi dalam bentuk visual *Wordcloud*
- 12. Menerapkan algoritma *Association Rules* berdasarkan hasil klasifikasi terbaik
- 13. Melakukan analisis perbaikan dengan *Fishbone Diagram*
- 14. Menarik kesimpulan dan saran

BAB IV

HASIL DAN PEMBAHASAN

4.1 Ekstrasi Ulasan

Ekstrasi ulasan adalah proses untuk memperoleh data teks berupa ulasan dari *google playstore* dengan metode *scraping* data menggunakan *python* dan menghasilkan 960 ulasan. Data yang diambil berupa tanggal, nama, ulasan pengguna, dan rating aplikasi *Reddoorz*. Data hasil *scraping* didapat dan disimpan dalam bentuk *.xlsx* (excel) . Hasil *scraping* ulasan dapat dilihat pada Gambar 4.1 :

	reviewid	userName	userImage	content	score	thumbsUpCount	CreatedAt	replyContent	repliedAt	appVersion	sortOrder	appid	
20013	2ee6e1ce	Pengguna	https://pl.	Apk nya bagus. Saya pernah co	2	0	2018-09-1	Halo Red	2018-09-1		most_rele	com.reddoorz.app	
119625	865b4de5	Pengguna	https://pl.	Membantu di detik detik terak	4	0	1.5.4	2018-09-1		1.5.4	most_rele	com.reddoorz.app	
40009	f75b5f30	Pengguna	https://pl.	Di aplikasi cuma tertera 1 night	3	2	1.5.5	2018-10-1	Halo Ita,	2018-10-1	1.5.5	most_rele	com.reddoorz.app
20058	880fca3e	Pengguna	https://pl.	Check in nya lama, gak kaya ap	2	2	1.5.5	2018-10-1		1.5.5	most_rele	com.reddoorz.app	
40024	27db87bc	Pengguna	https://pl.	klo mau pesan bukan buat hari	3	0	1.5.5	2018-10-1		1.5.5	most_rele	com.reddoorz.app	
40181	f3a4d7c8	Pengguna	https://pl.	Banyak promo kalo pake aplika	3	0	1.5.5	2018-10-1		1.5.5	most_rele	com.reddoorz.app	
40073	f80c07af	Pengguna	https://pl.	Standar tapi berkualitas, Bersih	3	2	1.5.5	2018-10-1		1.5.5	most_rele	com.reddoorz.app	
40194	d7e6c3c3	Pengguna	https://pl.	Kekuranganya adalah tidak upc	3	0	1.5.5	2018-10-1		1.5.5	most_rele	com.reddoorz.app	
91	ccffe059	Pengguna	https://pl.	Aplikasi Tai lah... kok bisa hotel	1	20	2018-11-1	Dear Hyo,	2018-11-1		most_rele	com.reddoorz.app	
20003								Halo! Teri					
	c503f05f	Pengguna	https://pl.	lumayan hotel budgeting nya. i	2	0	1.5.5	2018-11-1	Terkait pe	2018-07-1	1.5.5	most_rele	com.reddoorz.app
40012	8d4e279c	Pengguna	https://pl.	Harga hotel sih untuk backpac	3	3	1.5.5	2018-11-1		1.5.5	most_rele	com.reddoorz.app	
119618	898207c7	Pengguna	https://pl.	Saya pribadi yang suka travellir	4	1	1.5.5	2018-11-1		1.5.5	most_rele	com.reddoorz.app	
40192	ab362459	Pengguna	https://pl.	Bisa nggak ditambah fitur chat	3	1	1.5.5	2018-11-1		1.5.5	most_rele	com.reddoorz.app	
119655	f5733a64	Pengguna	https://pl.	Saya sudah pembatalan sudah	4	1	1.5.5	2018-11-1		1.5.5	most_rele	com.reddoorz.app	

Gambar 4.1 Scraping Ulasan Aplikasi Reddoorz

Kemudian hasil *scraping* disaring menjadi hanya ulasan yang diberikan pada tahun 2024, data ulasan tersebut berkurang menjadi 362 ulasan yang ditunjukkan pada Gambar 4.2.

	at	content	score
614	2024-01-03 02:03:48	Oke jos	4
615	2024-01-03 10:45:24	Parah aplikasi ini, pihak pemilik hotel udah ga k	1
616	2024-01-03 14:38:51 ::car...cara yang: dan .l..... jugal ...c	4
617	2024-01-03 18:55:39	Saya cek in di ampera 2, sudah pilih kamar yg a	2
618	2024-01-04 23:36:20	Sangat mudah penggunaannya dan sat set2 ... r	5
619	2024-01-05 18:17:18	Keren	4
620	2024-01-08 08:12:33	Sangat membantu saat liburan mencari pengin	5
621	2024-01-08 12:50:40	Pihak Hotel tidak menerima tamu, tapi ada di a	1
622	2024-01-09 11:08:30	Aplikasi Gak Jelas, Uang Saya Hilang 3 kali diten	1
623	2024-01-14 06:10:13	Mudah pemesanannya	4
624	2024-01-14 12:02:53	lumayan membantu klo bisa pihak hotel di trair	5
625	2024-01-15 11:05:01	Awasss.... Aplikasi udah banyak modus nya.. ua	1
626	2024-01-16 10:46:14	Tolong tambahkan fitur ganti foto profil istri say	3
627	2024-01-18 02:03:32	Sangat membantu dalam mencari tempat-temp	5
628	2024-01-19 15:15:36	Ga bisa masuk hotel karena alasan ga bawa kar	1
629	2024-01-19 15:54:14	Tolong iangan curanglah. masa pesan kamar dg	2

Gambar 4.2 Scapping Ulasan Aplikasi Reddoorz Tahun 2024

4.2 Pre-Processing Data

Data ulasan yang didapatkan tidak dapat langsung digunakan dalam penelitian, hal ini disebabkan karena data masih mengandung karakter, simbol, dan struktur yang beragam. Oleh sebab itu, diperlukan tahapan *pre-processing data* untuk mengubah tiap ulasan sehingga memiliki struktur yang sama dan siap diolah. *Pre-proceesing* adalah sebuah tahapan awal dalam mengoalh teks dan menjadi sebuah faktor yang dapat meningkatkan tingkat akurasi dari sebuah model dan efisiensi waktu pemrosesan (Chandrasekar & Qian, 2016). Proses ini membutuhkan sintaks dari python yang sesuai dengan Gambar 4.3. *Package* tersebut merupakan kerangka kerja untuk mengelola data berupa *text mining*.

```
[ ] #dt adalah dataframe yang digunakan membuat model analisis sentimen
datanegatif = negatif[['Comment']]
```

Gambar 4.3 Sintaks Pre-processing data

4.2.1 Case Folding

Case Folding adalah proses penyeragaman bentuk huruf pada dokumen ke dalam bentuk *uppercase* atau *lowercase* (Rianto et al, 2021). Hal ini bertujuan untuk menyamakan bentuk teks yang akan diteliti. Tahap ini dilakukan untuk menyeragamkan arti suatu kata yang ditulis sehingga tidak terdeteksi berbeda saat tahap pembobotan kata. Hasil proses ini ditunjukkan pada Tabel 4.1 :

Tabel 4.1 *Sampel Casefolding*

No	Sebelum <i>Casefolding</i>	Setelah <i>Casefolding</i>
1	Sangat membantu saat liburan mencari penginapan	sangat membantu saat liburan mencari penginapan
2	Pihak Hotel tidak menerima tamu, tapi ada di aplikasi, sudah bayar mahal tidak bisa menginap, tidak bisa di refund juga. Ini masuk ke penipuan	pihak hotel tidak menerima tamu, tapi ada di aplikasi, sudah bayar mahal tidak bisa menginap, tidak bisa di refund juga. ini masuk ke penipuan
3	Aplikasi Gak Jelas, Uang Saya Hilang 3 kali ditempat yg berbeda Akibat Booking di <i>Reddoorz</i> ini. Dan pihak <i>Reddoorz</i> tidak bertanggung jawab. Kapok pesan melalui aplikasi ini, mending hapus aja aplikasinya daripada uang kalian hilang.	aplikasi tidak jelas, uang saya hilang 3 kali ditempat yang berbeda akibat booking di <i>Reddoorz</i> ini. dan pihak <i>Reddoorz</i> tidak bertanggung jawab. kapok pesan melalui aplikasi ini, mending hapus aja aplikasinya daripada uang kalian hilang.

4.2.2 *Remove Number*

Remove number merupakan tahapan untuk menghaapus angka yang terdapat pada ulasa. Hal ini dilakukan karena angka tidak dapat memberikan penilaian maupun menunjukkan suatu persamaan. Hasil dari proses ini ditunjukkan pada Tabel 4.2

Tabel 4.2 *Sample Remove Number*

No	Sebelum <i>Removenumber</i>	Setelah <i>Removenumber</i>
1	sangat membantu saat liburan mencari penginapan	sangat membantu saat liburan mencari penginapan
2	pihak hotel tidak menerima tamu, tapi ada di aplikasi, sudah bayar mahal tidak bisa menginap, tidak bisa di refund juga. ini masuk ke penipuan	pihak hotel tidak menerima tamu, tapi ada di aplikasi, sudah bayar mahal tidak bisa menginap, tidak bisa di refund juga. ini masuk ke penipuan
3	aplikasi tidak jelas, uang saya hilang 3 kali ditempat yang berbeda akibat booking di <i>Reddoorz</i> ini. dan pihak <i>Reddoorz</i> tidak bertanggung jawab. kapok pesan melalui aplikasi ini, mending hapus aja aplikasinya daripada uang kalian hilang.	aplikasi tidak jelas, uang saya hilang kali ditempat yang berbeda akibat booking di <i>Reddoorz</i> ini. dan pihak <i>Reddoorz</i> tidak bertanggung jawab. kapok pesan melalui aplikasi ini, mending hapus aja aplikasinya daripada uang kalian hilang.

4.2.3 *Remove Punctuation*

Proses ini merupakan penghapusan karakter-karakter yang tidak termasuk alphabet, seperti simbol ataupun tanda baca. Proses ini diperlukan karena penelitian hanya memerlukan teks. Hasil dari *remove punctuation* ditunjukkan pada Tabel 4.3 ;

Tabel 4.3 Sample Remove Punctuation

No	Sebelum Remove Punctuation	Setelah Remove Punctuation
1	awasss.... aplikasi sudah banyak modus nya.. uang kamu pada bakal kecuras.. hotel banyak yang tidak bekerja sama lagi dan kita pelanggan terombang ambing... aplikasi an*j*ng	awasss aplikasi sudah banyak modus nya uang kamu pada bakal kecuras hotel banyak yang tidak bekerja sama lagi dan kita pelanggan terombang ambing aplikasi anjng
2	... ::car...cara yang: dan !!..... jugaclick..... yang ::...ca. menyediakan berbagai si.. - : m!.	car cara yang dan juga click yang ca menyediakan berbagai si m aku
3	harus teliti pilih kamar, jangan asal murah,, murah tapi harus tetap kenyamanan nomor satu. terimakasih u.	harus teliti pilih kamar jangan asal murah murah tapi harus tetap kenyamanan nomor satu terimakasih

4.2.4 Normalisasi kata

Normalisasi kata adalah tahapan mengubah kata tidak baku menjadi baku sehingga dapat mempermudah pemahaman terhadap ulasan. Proses ini dilakukan dengan tujuan untuk menghilangkan bahasa tidak baku kepada bahasa baku yang menggunakan aturan Kamus Besar Bahasa Indonesia (Rianto et al., 2021). Kata tidak baku yang dimaksud adalah bahasa daerah/asing, bahasa gaul, ataupun singkatan serta beberapa kata salah ketik. Kamus normalisasi kata dapat dilihat pada Lampiran 2. Hasil dari tahap ini dapat dilihat pada Tabel 4.4 :

Tabel 4.4 Sampel Normalisasi Kata

No	Sebelum Normalisasi	Setelah Normalisasi
1	car cara yang dan juga click yang ca menyediakan berbagai si m aku	cara cara yang dan juga klik yang menyediakan berbagai m aku
2	awasss aplikasi sudah banyak modus nya uang kamu pada bakal kekuras hotel banyak yang tidak bekerja sama lagi dan kita pelanggan terombang ambing aplikasi anjng	awas aplikasi sudah banyak modus nya uang kamu pada bakal kekuras hotel banyak yang tidak bekerja sama lagi dan kita pelanggan terombang ambing aplikasi anjing
3	sangat membantu dalam mencari tempattempat penginapan yang recommended semoga bisa lebih bagus lagi pelayanan nya	sangat membantu dalam mencari tempat penginapan yang recommended semoga bisa lebih bagus lagi pelayanan nya

4.3 Sentimen Scoring

Metode klasifikasi pada analisis sentimen membutuhkan data yang telah diberi label, tahap ini adalah sentimen *scoring*. Metode ini merupakan teknik yang digunakan untuk melabelkan data, prose ini membutuhkan tiga kamus yaitu kamus sentimen, kamus *boosterwords*, dan kamus negasi (Rahman, 2022)

Kamus yang digunakan dalam proses ini memuat kata-kata yang memiliki nilai. Nilai dari masing-masing data akan dihitung berdasar kamus yang digunakan sehingga menghasilkan keputusan yang memuat label positif dan negatif. Ketentuan dalam proses ini berdasarkan program yang digunakan, sebagai berikut :

1. Setiap kata pada dokumen akan dinilai sesuai dengan kata dalam kamus sentimen, apabila kata pada dokumen tidak terdapat dalam kamus sentimen maka akan diberi nilai 0.

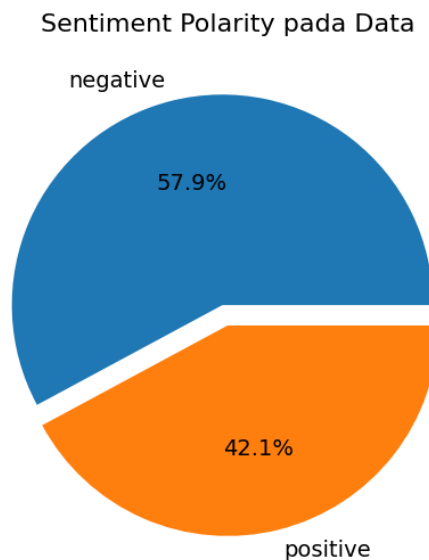
2. Setiap kata yang diikuti kata yang terdapat dalam kamus negasi akan mendapatkan nilai yang berlawanan dari nilai yang terdapat dalam kamus sentimen, contohnya kata berhasil memiliki nilai +5, namun jika diikuti kata negasi “tidak” maka nilai akan berubah menjadi -5.
3. Kata yang memiliki nilai >0 berdasarkan kamus sentimen dan sebelum kata tersebut termasuk dalam kamus *boosterwords*, maka nilai akan ditambahkan dengan nilai yang terdapat dalam kamus *boosterwords*.
4. Kata yang memiliki nilai <0 berdasarkan kamus sentimen dan sebelum kata tersebut termasuk dalam kamus *boosterwords*, maka nilai akan dikurangi dengan nilai yang terdapat dalam kamus *boosterwords*.

Total nilai yang diperoleh akan digunakan untuk memberi label pada masing-masing dokumen, apabila nilai akhir pada suatu dokumen bernilai \geq maka dokumen diberi label positif, sedangkan jika nilai akhir pada suatu dokumen bernilai <0 maka dokumen diberi label negatif. Berikut merupakan contoh perhitungan sentimen *scoring* yang terdapat pada Tabel 4.5:

Tabel 4.5 Contoh Perhitungan Sentimen Scoring

No	Ulasan	Skor	Label Sentimen
1	ok mantap	ok (sentimen) = 4	Positif
		mantap (sentimen) = 5	
		Total Skor = 9	
2	bagus proses cepat	bagus (sentimen) = 2	Positif
		proses (sentimen) = 1	
		cepat (sentimen) = 3	
		Total Skor = 6	
3	mau bayar dipersulit aplikasi rusak	mau (sentimen) = 5	Negatif
		bayar (sentimen) = 1	
		dipersulit = 0	
		aplikasi (sentimen) = -4	
		rusak (sentimen) = -4	
		Total Skor = -2	

Bedasarkan pelabeblan data yang dilakukan, hasil yang didapatkan adalah sentimen negatif memiliki persentase yang lebih tinggi dibandingkan sentimen positif dengan nilai 57.9% untuk sentimen negatif dan 42.1% untuk sentimen positif. Berikut merupakan gambar diagram yang menunjukkan perbandingan antara sentimen positif dan negatif :



Gambar 4.4 Sentimen Polarity

4.4 *Stopwords Removal*

Dokumen ulasan biasanya memuat kata-kata yang frekuensinya sering tetapi tidak memiliki makna. Kata-kata tersebut akan dihilangkan untuk mengurangi dimensi data sehingga tersisa dimensi data yang diperlukan. Oleh sebab itu, dibutuhkan proses *stopwords removal*. Pada proses ini dibutuhkan daftar kata yang akan dihapus, yaitu *stoplist*. Terdapat beberapa cara yang dapat dilakukan untuk menghilangkan *stopword*, salah satunya adalah menggunakan *library* yang telah tersedia yaitu NLTK (NLTK, 2022) dan Sastrawi (Sastrawi, 2018) Hasil dari *stopwords removal* dapat dilihat pada Tabel 4.6 :

Tabel 4.6 Sample Stopwords Removal

No	Sebelum <i>Stopwords</i>	Setelah <i>Stopwords</i>
1	parah ini aplikasi sudah bayar pas datang ke lokasi taunya penginapan tersebut sudah tidak bekerja sama dengan <i>Reddoorz</i>	parah aplikasi bayar lokasi taunya penginapan <i>Reddoorz</i>
2	kita pesan hotel tanggal dan januari kita pesan sebulan sebelumnya dan kita transfer pembayaran tanggal sampai di tempat ternyata hotel yang kita pesan sudah tidak bekerja sama lagi dengan aplikasi ini kita ditipu oleh aplikasi ini mana lagi high season lagi kita luntang lantung tdk dapat hotel di manapun kita ajukan refund malah tidak kembali semua dipotong ribu memang niat nipu kayaknya tidak ada tanggung jawabnya	pesan hotel tanggal januari pesan sebulan transfer pembayaran tanggal hotel pesan aplikasi ditipu aplikasi high season luntang lantung hotel manapun ajukan refund dipotong ribu niat menipu kayaknya tanggung
3	masih baru pesan cepat responnya nanti kalau sudah menginap saya tambah bintang lagi tentang pelayanan dan kondisi hotelnya maaf saya turunkan lagi satu ternyata pemesanan kamar yang twin room tanggal januari sampai januari menjadi single room di invoice nya malah harus	pesan cepat responnya menginap bintang pelayanan kondisi hotelnya maaf turunkan pemesanan kamar twin room tanggal januari januari single room invoice kamar mohon cek ulang terima kasih solusinya bulat uang payah

tambah kamar lagi jadinya
mohon di cek ulang terima
kasih waduh solusinya bulet
malah tambah uang payah

4.5 *Stemming dan Tokenizing*

Stemming adalah proses yang memiliki tujuan untuk mengukur ukuran kosakata dengan memetakan varian berdasarkan akar kata (Singh & Gupta, 2019). Proses ini dibutuhkan agar kata-kata dalam ulasan menjadi lebih sederhana sehingga memudahkan klasifikasi pada penelitian ini. Sedangkan *tokenizing* merupakan proses pemisahan setiap kata dalam suatu kalimat yang bertujuan untuk memotong kalimat berdasarkan tiap kata yang menyusun kalimat tersebut. Proses ini dilakukan dengan cara mengubah kalimat kedalam token (Rianto et al., 2021). Hasil dari proses *stemming* ini dapat dilihat pada Tabel 4.7 :

Tabel 4.7 Sample Stemming

No	Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
1	membantu liburan mencari penginapan	bantu libur cari inap
2	hotel menerima tamu aplikasi bayar mahal menginap refund masuk penipuan	hotel terima tamu aplikasi bayar mahal inap refund masuk tipu

3	aplikasi uang hilang ditempat berbeda akibat booking <i>Reddoorz</i> <i>Reddoorz</i> bertanggung kapok pesan aplikasi mending hapus aplikasinya uang hilang	aplikasi uang hilang tempat beda akibat booking <i>Reddoorz</i> <i>Reddoorz</i> tanggung kapok pesan aplikasi mending hapus aplikasi uang hilang
---	---	--

Berikut merupakan tahap lanjutan dari *stemming* yaitu *tokenizing*. Hasil *tokenizing* ditunjukkan oleh Tabel 4. 8 :

Tabel 4.8 Sample Tokenizing

No	Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
1	membantu liburan mencari penginapan	['bantu', 'libur', 'cari', 'inap']
2	hotel menerima tamu aplikasi bayar mahal menginap refund masuk penipuan	['hotel', 'terima', 'tamu', 'aplikasi', 'bayar', 'mahal', 'inap', 'refund', 'masuk', 'tipu']
3	aplikasi uang hilang ditempat berbeda akibat booking <i>Reddoorz</i> <i>Reddoorz</i> bertanggung kapok pesan aplikasi mending hapus aplikasinya uang hilang	['aplikasi', 'uang', 'hilang', 'tempat', 'beda', 'akibat', 'booking', 'Reddoorz', 'Reddoorz', 'tanggung', 'kapok', 'pesan', 'aplikasi', 'mending', 'hapus', 'aplikasi', 'uang', 'hilang']

4.6 Pembobotan kata TF-IDF

Analisis sentimen adalah penelitian yang memanfaatkan data jenis kualitatif, dimana data tersebut perlu ditransformasikan menjadi data kuantitatif. Pembobotan kata dibutuhkan agar kata-kata memiliki nilai dan dapat dihitung dan diolah. Pembobotan kata dalam

penelitian ini menggunakan perhitungan *Term Frequency-Inverse Document Frequency* (*TF-IDF*). Pada proses ini semua kata diproses sehingga memiliki bobot tersendiri. Hasil pembobotan kata dapat dilihat pada Tabel 4.9 :

Tabel 4.9 Hasil Pembobotan Kata TF-IDF

No	ac	acuh	ada	...	yang
0	0	0	0	...	0
1	0	0	0.18373321	...	0.14865
2	0	0	0	...	0
3	0.18197869	0	0.10289007	...	0.08325
⋮	⋮	⋮	⋮	⋮	⋮
319	0	0	0	...	0

4.7 Klasifikasi Sentimen

4.7.1 Data Latih dan Data Uji

Data yang melalui tahap *pre-processing* hingga pembobotan kata akan dibagi menjadi menjadi 2, yaitu data latih dan data uji. Data latih digunakan untuk melatih model klasifikasi yang sudah dibentuk sedangkan data uji digunakan untuk membentuk model klasifikasi. Hasil dari perbandingan kedua data tersebut didapatkan dengan metode *trial and error* dalam mendapatkan perbandingan terbaik, pada penelitian ini didapatkan perbandingan antara data latih dan data uji sebesar 80%:20% Algoritma *Naïve Bayes Classifier* dengan jumlah data latih sebanyak 254 data dan data uji sebesar 64 data.

4.7.2 Algoritma *Naïve Bayes Classifier*

Algoritma *Naïve Bayes Classifier* yang digunakan pada penelitian ini adalah dengan metode *Multinomial*, yaitu dengan menentukan sentimen melalui perhitungan *prior probability* dari setiap kata kelas terlebih dahulu, kemudian *conditional probability* dari setiap kata dan setiap dokumen dihitung untuk masuk kedalam suatu kelas (Rahman, 2022). Model *Machine Learning* yang telah selesai dilatih menggunakan algoritma *Naïve Bayes Classifier* akan menghasilkan sebuah prediksi yang dianalisis menggunakan *Confusion Matrix*. Pada *Confusion Matrix* memiliki beberapa informasi yang dapat dimanfaatkan untuk mengevaluasi kinerja model, diantaranya adalah

akurasi, error, presisi, *recall*, dan *f-score*. Berikut merupakan hasil *Confusion Matrix* dari algoritma *Naïve Bayes Classifier* yang ditunjukkan pada Tabel 4.11 :

Tabel 4.10 Hasil *Confusion Matrix* Model *Naïve Bayes Classifier*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	1	27
	Negatif	14	7

Berdasarkan Tabel 4.11, berikut merupakan hasil dari perhitungan parameter performansi dalam mengukur kinerja klasifikasi yang ditunjukkan pada Gambar 4.5 :

	precision	recall	f1-score	support
0	0.79	1.00	0.89	27
1	1.00	0.67	0.80	21
accuracy			0.85	48
macro avg	0.90	0.83	0.84	48
weighted avg	0.88	0.85	0.85	48

Gambar 4.5 Parameter Performansi

Hasil evaluasi kinerja *Naïve Bayes Classifier* diketahui bahwa nilai akurasi model ini sebesar 85,41%, sehingga dapat diartikan bahwa 85,41% data terklarifikasi dengan benar oleh sistem. Nilai yang lain berupa *precision* diperoleh sebesar 84%, nilai ini dipengaruhi oleh *false positive* (FP) yang menunjukkan seberapa banyak model dapat memprediksi kelas positif dari data yang masuk kedalam kelas positif. Sedangkan nilai *recall* pada model ini adalah 85%, dimana nilai ini dipengaruhi oleh *false negative* (FN) dan *true positive* (TP) yang artinya keberhasilan sistem untuk dapat mengenali kembali sampel positif pada data uji.

4.7.3 Algoritma *Support Vector Machine*

Algoritma *Support Vector Machine* menentukan sentimen dengan mencari *hyperline* terbaik. Model ini dimodifikasi dengan menggunakan *kernel* karena analisis sentimen tergolong dalam klasifikasi *non-linear* (Rahman, 2022). *Kernel* yang digunakan dalam penelitian ini adalah *Linear* hal ini dilakukan karena *linear* memiliki nilai error yang

tergolong kecil. Sama halnya dengan algoritma *Naïve Bayes Classifier*, model dari *machine learning* yang telah selesai dilatih menggunakan algoritma *Support Vector Machine* akan menghasilkan sebuah prediksi yang dianalisis menggunakan *Confusion Matrix*. Pada *Confusion Matrix* memiliki beberapa informasi yang dapat dimanfaatkan untuk mengevaluasi kinerja model, diantaranya adalah akurasi, error, presisi, *recall*, dan *f-score*. Berikut merupakan hasil *Confusion Matrix* dari algoritma *Support Vector Machine* yang ditunjukkan pada Tabel 4.11 :

Tabel 4.11 Hasil *Confusion Matrix* Model *Support Vector Machine*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	5	32
	Negatif	36	1

Berdasarkan Tabel 4.11, berikut merupakan hasil dari perhitungan parameter performansi dalam mengukur kinerja klasifikasi yang ditunjukkan pada Gambar 4.6 :

	precision	recall	f1-score	support
negative	0.97	0.86	0.91	37
positive	0.88	0.97	0.92	37
accuracy			0.92	74
macro avg	0.92	0.92	0.92	74
weighted avg	0.92	0.92	0.92	74

Gambar 4.6 Parameter Performansi *Support Vector Machine*

Berdasarkan Gambar 4.6 dapat diketahui bahwa parameter performansi pada algoritma *Support Vector Machine* memiliki nilai akurasi sebesar 91.89%, sehingga dapat diartikan bahwa 91.89% data terklarifikasi dengan benar oleh sistem. Sedangkan nilai *precision* pada model ini sebesar 92% yang diartikan bahwa sebanyak 92% model dapat memprediksi kelas positif dari data yang masuk kedalam kelas positif. Kemudian nilai lain yang muncul adalah *recall* adalah 92% yang dapat diartikan bahwa sebanyak 92% keberhasilan sistem untuk dapat mengenali kembali sampel positif pada data uji.

4.7.4 Perbandingan Hasil Kinerja Klasifikasi

Perbandingan hasil kinerja klasifikasi algoritma *Naïve bayes Classifier* dan *Support Vector Machine* dapat dilihat dari nilai akurasi akhir yang diperoleh model kedua algoritma tersebut, algoritma yang memiliki nilai akurasi lebih tinggi merupakan algoritma terbaik dalam mengklasifikasikan sentimen ulasan pengguna terhadap penggunaan aplikasi *Reddoorz*. Berikut merupakan hasil perbandingan nilai akurasi yang ditunjukkan pada Tabel 4.12 :

Tabel 4.12 Hasil Akhir Nilai Akurasi

Model	Nilai Akurasi
<i>Naïve bayes Classifier</i>	85.41%
<i>Support Vector Machine</i>	91.89%

Berdasarkan tabel 4.12, dapat diketahui bahwa algoritma *Naïve bayes Classifier* memiliki nilai sebesar 85.41%, sedangkan algoritma *Support Vector Machine* memiliki nilai akurasi sebesar 91.89% Sehingga dapat disimpulkan bahwa algoritma *Support Vector Machine* memiliki kinerja lebih baik dibandingkan dengan algoritma *Naïve bayes Classifier* dalam mengklasifikasi sentimen ulasan pengguna terhadap penggunaan aplikasi *Reddoorz* karena memiliki nilai akurasi yang lebih tinggi.

Faktor yang dapat mempengaruhi hasil prediksi adalah taraf kesalahan pada pelabelan data yang dilakukan, pada tahap pengklasifikasian algoritma *Naïve bayes Classifier* memiliki kesalahan pelabelan data yang lebih banyak dibandingkan algoritma *Support Vector Machine*, sehingga hal tersebut mempengaruhi keberhasilan *classifier* dalam memprediksi kelas untuk setiap dokumen yang dijadikan data uji. Kesalahan tersebut dapat dilihat pada tabel *confusion matrix* pada Tabel 4.10 dan Tabel 4.10. Kesimpulan berdasarkan kedua tabel tersebut adalah nilai *False Positive* (FP) dan *False Negative* (FN) sebagai nilai yang menunjukkan kesalahan pada pelabelan data dalam data uji pada algoritma *Naïve bayes Classifier* memiliki nilai yang lebih tinggi dibandingkan dengan algoritma *Classifier* memiliki kesalahan pelabelan data yang lebih banyak dibandingkan algoritma *Support Vector Machine*.



Gambar 4.8 Wordcloud Ulasan Sentimen Negatif

Berdasarkan pada Gambar 4.8 dapat diketahui bahwa kata yang sering muncul pada ulasan sentimen negatif adalah “aplikasi”,”hotel”, dan ”refund”. Hal ini menunjukkan bahwa pengguna banyak beropini negatif mengenai aplikasi yang dinilai sulit untuk memberikan uang *refund* jika terjadi kendala dalam pemesanan. Kata-kata lain yang juga sering muncul adalah “respon” dan “customer service” yang menunjukkan bahwa respon dari *customer service* dalam menyelesaikan kendala pengguna dinilai kurang tepat dan tanggap sehingga pengguna menjadi kecewa. Berdasarkan keluhan pengguna diharapkan *Reddoorz* dapat memperbaiki sistem untuk memudahkan pengguna dalam memesan hotel, melakukan *refund*, dan memperbaiki *customer service* agar lebih cepat dan tanggap sehingga pengguna akan merasa puas dan berlangganan dengan nyaman.

4.9 Association Rules

Association rules memiliki tujuan untuk mencari pola atau asosiasi antara kata yang sering muncul secara bersamaan dengan tujuan untuk memperkuat pencarian informasi dan dapat memprediksi suatu penyebab permasalahan dari suatu topik pembicaraan (DAELI, 2022). Pengolahan data ini menggunakan algoritma apriori dan pemrograman *python* dengan *software Google Collaboratory*, dengan data yang diekstraksi adalah data ulasan dengan sentimen negatif yang berisi kritik dan keluhan atas penggunaan aplikasi *reddoorz* dengan jumlah data sebanyak 160 data.

Pengolahan data *association rules* dengan algoritma apriori ini menggunakan *minimum support* sebesar 0,01 dan *minimum confidence* sebesar 0,5. Berikut adalah kata yang sering muncul atau digunakan pada ulasan negatif pengguna *reddoorz* :

Tabel 4.13 Common Words

No	Kata	Jumlah Kata
1	aplikasi	129
2	hotel	120
3	<i>reddoorz</i>	63
4	pesan	59
5	bayar	48
6	uang	38
7	<i>refund</i>	35
8	kamar	34
9	batal	31
10	kecewa	26

Pada gambar 4.13 terlihat bahwa 10 kata yang sering muncul dalam keluhan pelanggan. Kata pertama yang sering muncul adalah kata “aplikasi” sebanyak 129 kata, lalu kata kedua yang sering muncul adalah kata “hotel” dengan jumlah sebanyak 120 kata, kemudian kata “*reddoorz*” dengan jumlah kata sebanyak 63 kata, kata keempat yang sering muncul adalah “pesan” dengan jumlah kata sebanyak 48 kata, kemudian kata selanjutnya adalah “bayar” dengan jumlah kata sebanyak 48, lalu kata berikutnya adalah “uang” sebanyak 38 kata, kata selanjutnya adalah “*refund*” dengan jumlah kata sebanyak 35 kata, lalu kata berikutnya adalah “kamar” sebanyak 34 kata, kemudian kata “batal” dengan jumlah kata 31 kata, dan yang terakhir “kecewa” dengan jumlah kata sebanyak 26 kata. Dari 10 kata yang sering muncul akan dicari asosiasi kata yang berhubungan dengan kata-kata tersebut menggunakan nilai *support* sebagai persentase dari kata yang ada dalam data, sedangkan nilai *confident* sebagai nilai yang menunjukkan kuatnya asosiasi antar kata yang terbentuk. Berikut merupakan *association rules* berdasarkan kata kunci ditemukan dari *frequent word* yang ditunjukkan pada tabel 4.14 hingga table 4.24

Tabel 4.14 *Association Rules* Aplikasi

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Aplikasi	batal	0.1006289	0.19753086
	bayar	0.1761006	0.34567901
	booking	0.0943396	0.18518519
	hotel	0.2893082	0.56790123
	foto	0.0566038	0.11111111
	kamar	0.0943396	0.18518519
	kecewa	0.0691824	0.13580247
	kembali	0.0880503	0.17283951
	masuk	0.0691824	0.13580247
	pakai	0.0754717	0.14814815
	pas	0.0754717	0.14814815
	pesan	0.1949686	0.38271605
	reddoorz	0.1761006	0.34567901
	refund	0.1006289	0.19753086
	tipu	0.1194969	0.2345679

Tabel 4.15 *Association Rules* Hotel

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Hotel	aplikasi	0.289308	0.67647059
	batal	0.062893	0.14705882
	bayar	0.150943	0.35294118
	booking	0.075472	0.17647059
	customer	0.044025	0.10294118
	dana	0.050314	0.11764706
	hubung	0.044025	0.10294118
	kamar	0.100629	0.23529412
	kecewa	0.075472	0.17647059
	kembali	0.09434	0.22058824
	kerjasama	0.069182	0.16176471
	konfirmasi	0.044025	0.10294118

masuk	0.069182	0.16176471
penuh	0.050314	0.11764706
pesan	0.081761	0.19117647
pindah	0.144654	0.33823529
reddoorz	0.050314	0.11764706
refund	0.132075	0.30882353
sesuai	0.075472	0.17647059
tipu	0.050314	0.11764706
uang	0.138365	0.32352941

Tabel 4.16 Association Rules Reddoorz

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
<i>Reddoorz</i>	saja	0.031447	0.11627907
	aplikasi	0.176101	0.65116279
	batal	0.044025	0.1627907
	bayar	0.09434	0.34883721
	booking	0.050314	0.18604651
	buruk	0.031447	0.11627907
	customer	0.031447	0.11627907
	dana	0.044025	0.1627907
	hotel	0.132075	0.48837209
	hubung	0.037736	0.13953488
	inap	0.056604	0.20930233
	kali	0.044025	0.1627907
	kamar	0.050314	0.18604651
	kecewa	0.069182	0.25581395
	kembali	0.069182	0.25581395
	kerja	0.044025	0.1627907
	kerjasama	0.044025	0.1627907
	langsung	0.031447	0.11627907
	masuk	0.044025	0.1627907
	pakai	0.037736	0.13953488

pas	0.031447	0.11627907
pesan	0.119497	0.44186047
refund	0.081761	0.30232558
sesuai	0.037736	0.13953488
tipu	0.050314	0.18604651
tolong	0.031447	0.11627907
transfer	0.031447	0.11627907
tutup	0.031447	0.11627907
uang	0.09434	0.34883721

Tabel 4.17 Association Rules Pesan

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Pesan	aplikasi	0.19497	0.673913
	gagal	0.03145	0.1086957
	batal	0.07547	0.2608696
	bayar	0.1195	0.4130435
	booking	0.03145	0.1086957
	customer	0.04403	0.1521739
	dana	0.03145	0.1086957
	hotel	0.14465	0.5
	inap	0.05031	0.173913
	kali	0.05031	0.173913
	kamar	0.08805	0.3043478
	kecewa	0.06918	0.2391304
	kembali	0.06289	0.2173913
	lokasi	0.03145	0.1086957
	masuk	0.03774	0.1304348
	pakai	0.03774	0.1304348
	reddoorz	0.03145	0.1086957
	respon	0.1195	0.4130435
	service	0.06289	0.2173913
	tipu	0.04403	0.1521739

tolong	0.03145	0.1086957
uang	0.05031	0.173913
via	0.03145	0.1086957

Tabel 4.18 Association Rules Bayar

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Bayar	aplikasi	0.176101	0.6363636
	batal	0.031447	0.1136364
	booking	0.044025	0.1590909
	hotel	0.150943	0.5454545
	gagal	0.044025	0.1590909
	kamar	0.08805	0.3181818
	kecewa	0.031447	0.1136364
	kembali	0.056604	0.2045455
	kerjasama	0.050314	0.1818182
	lokasi	0.031447	0.1136364
	masuk	0.031447	0.1136364
	pakai	0.044025	0.1590909
	pas	0.056604	0.2045455
	pesan	0.119497	0.4318182
	reddoorz	0.09434	0.3409091
	refund	0.056604	0.2045455
	tipu	0.050314	0.1818182
	transfer	0.037736	0.1363636
	uang	0.08805	0.3181818
	via	0.044025	0.1590909

Tabel 4.19 Association Rules Uang

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Uang	tambah	0.025157	0.1176471
	aplikasi	0.18239	0.8529412
	batal	0.056604	0.2647059

booking	0.08805	0.4117647
hangus	0.056604	0.2647059
hilang	0.031447	0.1470588
hotel	0.025157	0.1176471
inap	0.138365	0.6470588
kali	0.031447	0.1470588
kamar	0.031447	0.1470588
kembali	0.050314	0.2352941
kerjasama	0.075472	0.3529412
lokasi	0.044025	0.2058824
masuk	0.031447	0.1470588
pakai	0.044025	0.2058824
parah	0.025157	0.1176471
pas	0.025157	0.1176471
pesan	0.044025	0.2058824
reddoorz	0.100629	0.4705882
refund	0.09434	0.4411765
tipu	0.056604	0.2647059
tolong	0.062893	0.2941176
transfer	0.025157	0.1176471

Tabel 4.20 Association Rules Refund

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
<i>Refund</i>	aplikasi	0.100629	0.57142857
	bantu	0.031447	0.17857143
	batal	0.044025	0.25
	bayar	0.056604	0.32142857
	booking	0.025157	0.14285714
	buruk	0.018868	0.10714286
	customer	0.025157	0.14285714
	dana	0.018868	0.10714286
	hati	0.025157	0.14285714

hotel	0.075472	0.42857143
hubung	0.018868	0.10714286
inap	0.025157	0.14285714
kamar	0.037736	0.21428571
kecewa	0.025157	0.14285714
keluh	0.018868	0.10714286
kembali	0.025157	0.14285714
kerja	0.018868	0.10714286
pas	0.018868	0.10714286
pesan	0.062893	0.35714286
proses	0.018868	0.10714286
reddoorz	0.081761	0.46428571
respon	0.018868	0.10714286
ribu	0.025157	0.14285714
service	0.018868	0.10714286
sesuai	0.018868	0.10714286
susah	0.018868	0.10714286
tanggal	0.018868	0.10714286
tipu	0.044025	0.25
tolong	0.025157	0.14285714
transfer	0.025157	0.14285714
tutup	0.018868	0.10714286
uang	0.056604	0.32142857

Tabel 4.21 Association Rules Kamar

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Kamar	buruk	0.031447	0.18518519
	anak	0.018868	0.11111111
	aplikasi	0.09434	0.55555556
	batal	0.018868	0.11111111
	bayar	0.08805	0.51851852
	booking	0.025157	0.14814815

dana	0.018868	0.11111111
hati	0.018868	0.11111111
hotel	0.100629	0.59259259
inap	0.050314	0.2962963
jam	0.025157	0.14814815
kali	0.031447	0.18518519
kasih	0.025157	0.14814815
kecewa	0.044025	0.25925926
kembali	0.044025	0.25925926
kerjasama	0.018868	0.11111111
konfirmasi	0.018868	0.11111111
lokasi	0.037736	0.22222222
lunas	0.018868	0.11111111
masuk	0.025157	0.14814815
bau	0.018868	0.11111111
pakai	0.037736	0.22222222
penuh	0.031447	0.18518519
pesan	0.025157	0.14814815
pindah	0.08805	0.51851852
reddoorz	0.018868	0.11111111
refund	0.050314	0.2962963
respon	0.037736	0.22222222
sedia	0.018868	0.11111111
sesuai	0.025157	0.14814815
tera	0.031447	0.18518519
tidur	0.025157	0.14814815
tipu	0.018868	0.11111111
tolong	0.031447	0.18518519
transfer	0.025157	0.14814815
uang	0.018868	0.11111111
via	0.050314	0.2962963

Tabel 4.22 Association Rules Batal

<i>Conclusion</i>	<i>Premises</i>	<i>Support</i>	<i>Confidence</i>
Batal	aplikasi	0.10063	0.615384615
	bayar	0.03145	0.192307692
	buruk	0.01887	0.115384615
	customer	0.02516	0.153846154
	dana	0.01887	0.115384615
	hotel	0.06289	0.384615385
	hubung	0.01887	0.115384615
	kamar	0.01887	0.115384615
	kecewa	0.02516	0.153846154
	kembali	0.03145	0.192307692
	masuk	0.02516	0.153846154
	parah	0.01887	0.115384615
	pesan	0.07547	0.461538462
	reddoorz	0.04403	0.269230769
	refund	0.04403	0.269230769
	respon	0.01887	0.115384615
	service	0.01887	0.115384615
	sesuai	0.01887	0.115384615
	tanggal	0.01887	0.115384615
	tipu	0.02516	0.153846154
	transfer	0.01887	0.115384615
	uang	0.0566	0.346153846

4.10 Analisis Diagram *Fishbone*

Berdasarkan hasil dari pengolahan data ulasan pengguna aplikasi *reddoorz* dengan sentimen negatif menggunakan *association rules*, maka diperoleh kata kunci sebagai masalah dari aplikasi *reddoorz* yakni pada kata “aplikasi”, “hotel”, “reddoorz”, “pesan”, “bayar”, “uang”, “*refund*”, “kamar”, dan “batal”. Dari permasalahan yang muncul kemudian dianalisis untuk mengetahui potensi dari akar masalah. Potensi dari akar masalah yang digunakan berdasarkan *premises* dari hasil *association rules*. Kata

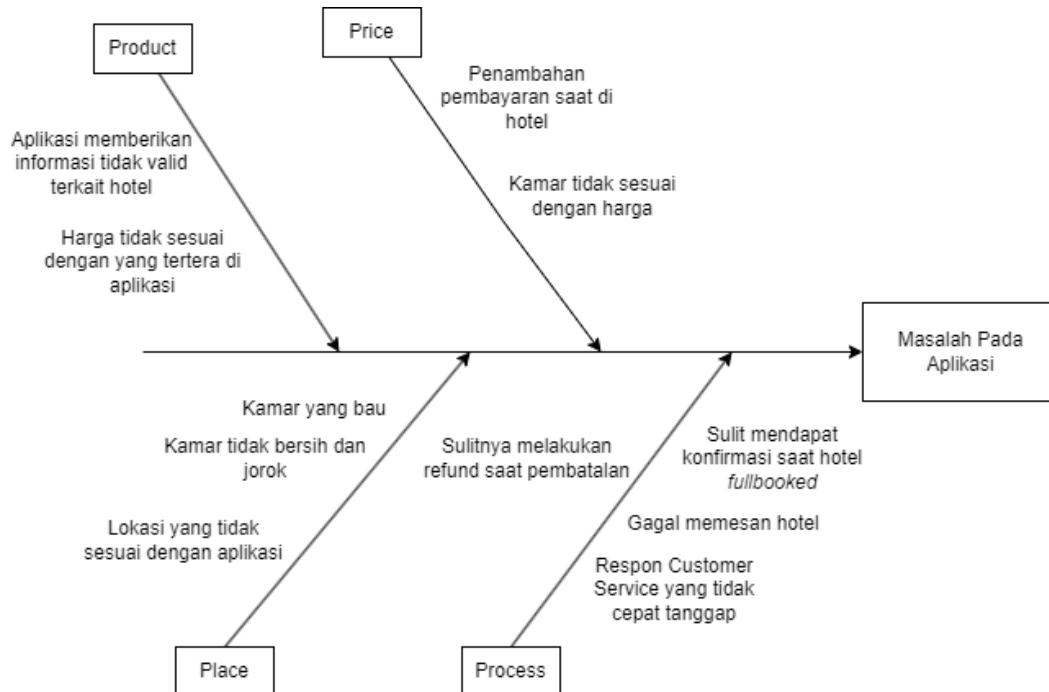
“aplikasi”, “pesan”, dan “reddoorz” dijadikan sebagai satu kesatuan karena memiliki *premises* yang serupa, kemudian kata “kamar” dan “hotel” dijadikan satu kesatuan, dan kata “bayar”, “uang”, “refund”, dan “batal” juga memiliki *premises* yang serupa sehingga dijadikan satu kesatuan. Berikut merupakan premis yang ada pada hasil *association rule* yang ditunjukkan pada tabel

Tabel 4.23 *Premises*

<i>Conclucion</i>	<i>Premises</i>
Aplikasi	Foto
	Lokasi
	Pesan
Hotel	Konfirmasi
	Penuh
Reddoorz	Customer
	respon
	Kerjasama
Pesan	Gagal
	respon
	Batal
Bayar	Batal
	Tambah
	Gagal
Uang	Hangus
	Hilang
Refund	Proses
	Susah
Kamar	Buruk
	Kecewa
	Bau
	Sesuai

Analisis akar masalah menggunakan alat bantu yaitu *fishbone diagram* dengan beberapa empat aspek yang disertakan berdasarkan *markering mix 4P*, aspek ini

berhubungan dengan permasalahan yang ditemukan pada aplikasi *Reddoorz* meliputi *product*, *price*, *place*, dan *process*. Berikut merupakan hasil dari pembuatan *fishbone diagram* untuk setiap masalah yang ditunjukkan pada Gambar 4.9 hingga Gambar 4.11 :



Gambar 4.9 *Fishbone Diagram* Masalah Aplikasi

Metode *marketing mix* yang digunakan pada *fishbone* diagram dapat membantu dalam menguraikan masalah. *Marketing mix* yang terdiri dari *product*, *places*, *prices*, and *promotions* dapat berpengaruh dalam pembelian online melalui aplikasi salah satu contohnya adalah keputusan pembelian pakaian wanita secara online, dimana semakin besar nilai produk maka semakin tinggi pula keputusan untuk membeli (Tjahjono, 2013)

Pemasaran melalui internet menjadi pasar yang menjanjikan karena dapat diakses dimana saja dan kapan saja, banyak faktor yang mempengaruhi dalam pembelian suatu produk salah satunya produk, harga, tempat, promosi, dan rangsangan lingkungan (Mas'ari et al., 2020) Sehingga metode *fishbone* dengan mengkombinasikan *marketing mix* untuk setiap batang masalah dapat membantu mengetahui faktor eksternal yang mempengaruhi dalam keputusan pemesanan kamar hotel melalui aplikasi *reddoorz*, dimana akar masalah telah diketahui dan akan dijabarkan dengan mudah menggunakan *marketing mix*.

4.10.1 Rekomendasi dari Analisis *Fishbone Diagram*

Pembuatan *fishbone* diagram yang ditunjukkan pada Gambar 4.9 hingga Gambar 4.11 menggunakan empat aspek yaitu *product*, *price*, *place*, dan *process*. Aspek produk terkait dengan masalah yang berkaitan pada fitur aplikasi *reddoorz*. Aspek *price* berkaitan dengan masalah harga yang diberikan untuk sebuah kamar bagi pengguna aplikasi *reddoorz*. Aspek *place* berhubungan dengan masalah pada kamar atau hotel pada aplikasi *reddoorz*. Aspek terakhir adalah *process* yaitu berkaitan dengan masalah yang ditemui pengguna ketika menjalankan aplikasi. Berdasarkan dari analisis *cause root* yang telah dilakukan untuk setiap kata kunci menggunakan *fishbone diagram*, maka dapat disimpulkan rekomendasi perbaikan menurut setiap aspek yang tertera ditunjukkan pada Tabel 4.23 :

Tabel 4.24 Rekomendasi Perbaikan

Aspek	Permasalahan	Rekomendasi Perbaikan
<i>Product</i>	Aplikasi memberikan informasi yang tidak valid terkait hotel (Foto)	Memastikan secara berkala bahwa foto yang terpasang pada aplikasi sudah sesuai.
	Harga tidak sesuai dengan yang tertera pada aplikasi	Memastikan secara berkala untuk harga yang tertera pada aplikasi karena sistem dynamic pricing yang kadang terjadi kendala
<i>Price</i>	Penambahan pembayaran saat di hotel	Melakukan <i>quality control</i> untuk setiap hotel agar menjalankan prosedur sesuai SOP
	Kamar tidak sesuai dengan harga	Membuat kesepakatan dengan pemilik hotel agar memberikan harga yang sesuai dengan kamar yang diberikan atau tidak <i>overprice</i> .

<i>Place</i>	Kamar yang bau	Melakukan pengecekan secara berkala untuk memastikan kamar yang dijual layak
	Lokasi yang tidak sesuai dengan aplikasi	Memastikan lokasi yang dicantumkan pada aplikasi sudah sesuai karena tidak sedikit hotel yang memiliki akses yang kurang memadai
	Kamar tidak bersih dan jorok	Memastikan bahwa SOP yang diberikan kepada setiap hotel dijalankan dengan benar
<i>Process</i>	<i>Respon Customer Service</i> yang tidak cepat tanggap	Memberikan pelatihan dan melakukan <i>quality control</i> terhadap kinerja <i>customer service</i> dapat melalui <i>feedback</i> dari pengguna,
	Gagal dalam pemesanan hotel	Melakukan <i>quality control</i> secara berkala terhadap kinerja aplikasi
	Sulitnya melakukan <i>refund</i> saat pembatalan	Memberikan panduan <i>refund</i> yang dapat diakses dengan mudah oleh pengguna.
	Sulit mendapat konfirmasi saat hotel <i>fullbooked</i>	Memastikan pihak hotel telah menjalankan prosedur pengisian kamar dengan sesuai sehingga tidak terjadi gap antara diaplikasi dan hotel

BAB V

KESIMPULAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan analisis sentimen terhadap ulasan pengguna aplikasi *Reddoorz* pada *Google Playstore* dapat disimpulkan bahwa :

1. Model klasifikasi dengan kinerja klasifikasi terbaik berdasarkan nilai akurasi pada penelitian ini adalah algoritma *Support Vector Machine* dengan nilai akurasi sebesar 91.89%, sehingga dapat disimpulkan bahwa algoritma *Support Vector Machine* dinyatakan lebih baik daripada algoritma *Naïve bayes Classifier* dengan nilai akurasi 85.41% dalam mengklasifikasikan sentimen ulasan pengguna *Reddoorz* pada *Google Playstore*.
2. Berdasarkan algoritma *association rules* dengan metode apriori dapat diketahui bahwa kata yang sering muncul pada ulasan aplikasi *Reddoorz* di *Google Playstore* adalah “aplikasi” sebanyak 129 kata, “hotel” sebanyak 120 kata, “reddoorz” sebanyak 63 kata, “pesan” sebanyak 59 kata, “bayar” sebanyak 48 kata, “uang” sebanyak 38 kata, “refund” sebanyak 35 kata, “kamar” sebanyak 34 kata, “batal” sebanyak 31 kata, dan yang terakhir adalah “kecewa” sebanyak 26 kata.
3. Berdasarkan masalah yang telah ditemukan menggunakan analisis *fishbone diagram* maka rekomendasi perbaikan yang tepat meliputi pengecekan secara berkala terhadap informasi yang diberikan pada aplikasi seperti foto, lokasi, maupun harga, melakukan *quality control* secara berkala terhadap kamar yang dijual agar memenuhi SOP, melakukan pengecekan berkala terhadap kinerja hotel agar menjalankan sesuai SOP terkait pembayaran hotel, melakukan kesepakatan kepada pemilik hotel terkait harga yang sesuai dengan kamar, memberikan pelatihan dan melakukan *quality control* terhadap kinerja *customer service*, melakukan pengecekan secara berkala terhadap kinerja aplikasi, dan memastikan pihak hotel telah menjalankan prosedur pengisian kamar dengan sesuai sehingga tidak terjadi gap antara diaplikasikan dan hotel

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan yang didapat, peneliti memberikan saran sebagai berikut :

1. Bagi penelitian selanjutnya untuk algoritma *Naïve Bayes Classifier* ditambahkan algoritma lain untuk meningkatkan akurasi. Selain itu, analisis sentimen dapat dilakukan dengan ragam metode yang lain untuk membandingkan tingkat akurasi tiap metode.
2. Bagi aplikasi *Reddoorz* perlu meningkatkan pelayanan dan dapat memberikan *feedback* terhadap keluhan pengguna dengan lebih cepat tanggap berdasarkan ulasan yang diberikan pengguna agar tetap dapat bersaing dengan *online* hotel lain di Indonesia.

DAFTAR PUSTAKA

- Addam, M. K., Ng, H. T., & Chieu, L. (2002). Bayesian Online Classifiers for Text Classification and Filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in International ACM SIGIR Conference on Research and Development in*. Singapore.
- Agustina, C. N., Novita, R., Mustakim, & Rozanda, N. E. (2024). The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm. *Procedia Computer Science Volume 234*, 156-163.
- Alhanin, Y., & Juzaidinn, M. (2011). The Enhancement of Arabic Stemming by Using Light Stemming and Dictionary-Based Stemming . *Journal of Software Engineering and Application*.
- B, Y, A., Shaker, & Kumar, B. (2023). *Proceedings of the 1st International Conference on Innovation in Information Technology and Business (ICIITB 2022)*, (pp. 96-113).
- Bagnasco, S., Berzano, D., Guarise, A., Lusso, S., Masera, M., & Vallero, S. (2015). Monitoring of IaaS and scientific applications on the Cloud using the Elasticsearch ecosystem. *Journal of Physics: Conference Series vol. 608, no. 1*.
- Buololo, E. (2020). *Data Mining untuk Perguruan Tinggi*. Yogyakarta: Deepublish Publisher.
- Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. *Heliyon*.
- Chandrasekar, P., & Qian, K. (2016). The Impact of Data Preprocessing on the Performance of a Naïve Bayes Classifier. *Proceedings - International Computer Software and Applications Conference*, pp. 618-619.
- DAELI, N. C. (2022). *INTEGRASI SIX SIGMA DAN ASSOCIATION RULE UNTUK KELUHAN PELANGGAN PROVIDER BY.U* . Semarang: UNIVERSITAS DIPONEGORO.

- Developers, S.-L. (2022). *Naive Bayes*.
- Fanisaa, S., Fauzi, M., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2, No. 8,*, 2766-2770.
- fatmawati, F., & Affandes, M. (2017). Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) Pada Akun Facebook Group iRaise Helpdesk. *Jurnal CoreIT Vol. 3, No. 1,*, 24-30.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm. *Procedia Computer Science Vol 161,* 765-772.
- Ginting, J. N., Sitompul, P. N., Siagian, M. v., Nduru, S. W., & Tanjung, F. S. (2023). Pelatihan Digital Marketing Hotel Menggunakan Teknologi Reddoorz Sebagai Potensi Bisnis. *ULINA : Jurnal Pengabdian kepada Masyarakat Vol 1, No.2* , 6-10.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concept and Techniques*. San Fransisco: Morgan Kaufmann Publishers.
- Heryana, A. (2020, May). *UJI CHI SQUARE*. Retrieved from ResearchGate : Uji Chi Square:
https://www.researchgate.net/publication/341539841_UJI_CHI_SQUARE
- Hidayah, N., & Sahibu, S. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 5, No.4,*, 820-826.

- Hubert, Phoenix, P., Sudaryono, R., & Suhartono, D. (2021). Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier. *Procedia Computer Science vol 179*, 498-506.
- Irvantoro, D., Saifudin, I., & Umilasari, R. (2019). Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naive Bayes Classifier Untuk. *repository.unmuhjember.ac.id*.
- Khosmah, S., & Ariwibowo, A. (2020). Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol 4, No. 4*, 648-654.
- Kim, H.-C., Park, J.-H., Kim, D.-W., & Lee, J. (2020). Multilabel naïve Bayes classification considering label dependence. *Pattern Recognition Letters Volume 136*, 279-285.
- Kresmanto, M. A., Hanggara, B. T., & Prakoso, B. S. (2020). Analisis Pengalaman Pengguna pada Aplikasi Mobile Booking Hotel dengan menggunakan Metode User Experience Questionnaire (UEQ) . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3637-3646.
- Latif, R. M., Shah, S. U., Ijaz, F., Abdullah, M. T., Farhan, M., & Karim, A. (2019). Data Scraping from Google Play Store and Visualization of its Content for Analytics. *International Conference on Computing, Mathematics and Engineering Technologies*.
- Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., . . . Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*.
- Ling, J., Kencana, I., & Oka, T. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi-Square. *E-Jurnal Matematika Vol 3, No. 3*, 92-99.
- Mahesh, B. (2019). Machine Learning Algorithms - A Review | Enhanced Reader. *International Journal of Science and Research (IJSR)*, 18(8), 381–386. .

- Maitra, S., Madan, S., kandwal, r., & Mahajan, P. (2018). Mining Authentic Student Feedback for Faculty Using Naive Bayes Classifier. *Prodecia Computer Science Vol 132*, 1171-1183.
- Mas'ari, A., IHamdy, M., & Safira, M. (2020). Analisa Strategi Marketing Mix Menggunakan Konsep 4p (Price, Product, Place, Promotion) pada PT. Haluan Riau. . *Jurnal Teknik Industri: Jurnal Hasil Penelitian dan Karya Ilmiah dalam Bidang Teknik Industri*, 79-86.
- McCallun, A., & Nigam, K. (1998). A Comparison of Event Models for Naïve Bayes Text Classification. *Workshop on Learning for Text Categorization AAAI*.
- Mujilahwati, S. (2016). Pre-Processing Text Mining pada Data Twitter. *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*.
- Narayan, Y. (2021). Comparative analysis of SVM and Naive Bayes classifier for the SEMG signal classification. (pp. 3241-3245). *Materials Today: Proceedings*.
- Niazi, K. A., Akhtar, W., AKhan, H. A., Yang, Y., & Athar, S. (2019). Hostpot Diagnosis For Solar Photovoltaic Modules Using a Naive Bayes Classifier. *Solar Energy Vol 190*, 34-43.
- NLTK. (2022). *NLTK Documentation*. Retrieved from <https://www.nltk.org/>
- Nurjanah, M., & Astusti, I. F. (2013). *Jurnal Informatika Vol 8, No. 3*.
- Pisner, D. A., & Schnyer, D. (2019). Support vector machine. *In Machine Learning: Methods and Applications to Brain Disorders*.
- Play, G. (2023, 10 2). Retrieved from Persyaratan Layanan Google Play.: https://play.google.com/intl/id_id/about/play-terms/
- PURWALESTARI, E. C. (2024). *ANALISIS SENTIMEN ULASAN APLIKASI DISNEY+ HOTSTAR PADA SITUS GOOGLE PLAYSTORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER DENGAN ALGORITMA MODIFIED PARTICLE SWARM OPTIMIZATION (MPSO)*. Semarang: Universitas Diponegoro.
- Rahman, H. A. (2022). *Analisis Sentimen Pada Perusahaan Penyedia Jasa Logistik J&T Menggunakan Algoritma Multinomial Naive Bayes dan Support Vector Machine*. Semarang: Universitas Diponegoro.

- Ramasubramanian, K., & Singh, A. (2017). *Machine Learning Using R*. New York: Apress Media.
- Ratiasasadara, P. W. (2022). *Analisis Sentimen Penerapan PPKM Pada Twitter Menggunakan Naive Bayes Classifier dengan Selkese Fitur Chi-Square*. Semarang: Universitas Diponegoro.
- Rianto, Wibowo, & Santosa. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*.
- Rozy, I. F., Ardiansyah, R., & Rebeka, N. (2019). Penerapan Normalisasi Kata Tidak Baku Menggunakan Levenshtein Distance pada Analisa Menggunakan Levenshtein Distance pada Analisa. *Seminar Informatika Aplikatif Polinema (SIAP)*.
- Rozy, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS Vol. 6, No. 1,* 37-43.
- Ruswati, Guffroni, A., & Rianto, R. (2018). Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm. *Scientific Journal of Informatics, 5(2)*, 91-104.
- Salloum, S., Al-Emran, M., Monem, A., & Shaalan, K. (2019). Using text mining techniques for extracting information from research articles. *Studies in Computational Intelligence, 740*, 373-379.
- Sanchez-Franco, M. J., Navarro-Garcia, A., & Rondan-Cataluna, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research, 499-506*.
- Sastrawi. (2018). *har07/PySastrawi: Indonesian stemmer*. Retrieved from Python port of PHP Sastrawi project: <https://github.com/har07/PySastrawi>
- Singh, J., & Gupta, V. (2019). A novel unsupervised corpus-based stemming technique using lexicon and corpus statistics. *Knowledge-Based Systems, 147-162*.

- SINIWI, L. M. (2021). *QUERY EXPANSION RANKING PADA ANALISIS SENTIMEN MENGGUNAKAN KLASIFIKASI MULTINOMIAL NAÏVE BAYES*. Semarang: Universitas Diponegoro.
- Taheri, S., & Mammadov, M. (2013). Learning the Naive Bayes Classifier with Optimization Models. *International Journal of Applied Mathematics and Computer Science*, 787-795.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining (1st ed)*. Boston: Pearson Education, Inc.
- Tjahjono, A. (2013). Analisis Marketing Mix, Lingkungan Sosial, Psikologi Terhadap Keputusan Pembelian Online Pakaian Wanita . *Jurnal Strategi Pemasaran*, 1-9.
- Wahid, D. H., & Azhari, S. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS Vol 10*.
- Wardani, N. S. (2020). *Analisis Sentimen Pemandangan Ibu Kota Negara Dengan Klasifikasi Naive Bayes Untuk Model Bernoulli dan Multinomial*. Semarang: Universitas Diponegoro.
- Wardani, N., Prahutama, A., & Kartikasari, P. (2020). Analisis Sentimen Pemandangan Ibu Kota Negara dengan Klasifikasi Naïve Bayes untuk Model Bernoulli dan Multinomial. *Jurnal Gaussian Vol. 9, No. 3,*, 237-246.

LAMPIRAN

Lampiran 1 Data Scrapping

No	Username	at	Content
1	Khairi Bayu S	2024-01-03 02:03:48	Oke jos
2	Hikmat	1/3/2024 10:45	Parah aplikasi ini, pihak pemilik hotel udah ga kerjasama lagi sama reddoors tp kamarnya terap dijual. Mana uang udah masuk dan ga bisa balik. Saya sangat tidak merekomendasikan kepada yg lain. Hati2 jgn sampai ketipu ya bro. Saya harap pihak google segera menutup aplikasi ini agar tdk makan korban yang lainnya
3	Saifudin m	1/3/2024 14:38 ::car...cara yang: dan !..... jugalclick..... yang ::ca. menyediakan berbagai si.. - : M!.aku
4	Aida Fitria	1/3/2024 18:55	Saya cek in di ampera 2, sudah pilih kamar yg ada balkonnnya dengan harga lebih mahal, tapi tetap tidak di kasih yg ada balkonnnya, tau gt saya pilih yg standar aja lebih murah, saya tidak puas
5	Rahmad Karue	1/4/2024 23:36	Sangat mudah penggunaannya dan sat set2 ... mantap
6	Fakhri Fuad Afif	1/8/2024 12:50	Pihak Hotel tidak menerima tamu, tapi ada di aplikasi, sudah bayar mahal tidak bisa menginap, tidak bisa di refund juga. Ini masuk ke penipuan
7	Hendi Sismaja	1/9/2024 11:08	Aplikasi Gak Jelas, Uang Saya Hilang 3 kali ditempat yg berbeda Akibat Booking di reddoorz ini. Dan pihak reddoorz tidak bertanggung jawab. Kapok pesan melalui aplikasi ini, mending hapus aja aplikasinya daripada uang kalian hilang.
8	M DG. ERANG	1/14/2024 6:10	Mudah pemesanannya
9	ochie creatives	1/14/2024 12:02	lumayan membantu klo bisa pihak hotel di training sesuai sop terbaik
∴	∴	∴	∴
320	Mochamad Alie	6/4/2024 9:36	Jelas lebih bagus dari pada oyo
321	Ade Rukmana	6/5/2024 5:50	Aderuk.mana Yube

Lampiran 2 Kamus Normalisasi

No	Slang	Formal	No	Slang	Formal
1	blm	belum	34	lu	kamu
2	okeee	ok	35	udah	sudah
3	gak	tidak	36	dgn	dengan
4	app	aplikasi	37	entar	nanti
5	tk	terimakasih	38	sdah	sudah
6	boking	pesan	39	redoorz	reddoorz
7	aplikasi	aplikasi	40	resepsinonis	resepsionis
8	luv	love	41	ko	kok
9	bru	baru	42	jg	juga
10	nginep	menginap	43	mlah	malah
11	poll	pol	44	smpe	sampai
12	n	dan	45	d	di
13	mantapp	mantap	46	sma	sama
14	kerem	keren	47	thx u	terimakasih
15	dr	dari	48	dsb	dan sebagainya
16	rd	reddoorz	49	yg	yang
17	mntap	mantap	50	nyampe	sampai
18	terjangko	terjangkau	51	cance	cancel
19	uda	sudah	52	trus	terus
20	thks	terimakasih	53	jugal	juga
21	jd	jadi	54	pemesana	pemesanan
22	gua	aku	55	rb	ribu
23	jmlh	jumlah	56	family	keluarga
24	gw	saya	57	tgl	tanggal
25	byar	bayar	58	cek in	masuk
26	okee	ok	59	reddoors	reddoorz
27	yg	yang	60	car	cara
28	oke	ok	61	anjng	anjing
29	tp	tapi	62	click	klik
30	jgn	jangan	63	sat	cepat
31	gt	gitu	:	:	:
32	klo	kalau	556	mana	bagaimana
33	awasss	awas	557	ok	setuju

Lampiran 3 Kamus Indonesian Sentiment Lexicon Positif

No	Kata	Skor	No	Kata	Skor
1	hai	3	34	thanks	3
2	merekam	2	35	pengembangan	3
3	ekstensif	3	36	diva	2
4	paripurna	1	37	punya	3
5	detail	2	38	tidak segan	2
6	pernik	3	39	detailnya	1
7	belas	2	40	tak segan	2
8	welas	4	41	aktivasi	2
9	kabung	1	42	asih	3
10	rahayu	4	43	kasih sayang	5
11	maaf	2	44	kekaguman	4
12	hello	2	45	kehangatan	4
13	promo	3	46	afeksi	2
14	terimakasih	5	47	renjana	2
15	cover	3	48	amor	2
16	mohon	2	49	cinta kasih	5
17	mengawal	2	50	tresna	2
18	statistik	1	51	filantropi	2
19	keuangan	3	52	cintrong	2
20	jalan terbuka	3	53	dinamika	3
21	banyaknya	3	54	tuhan	3
22	lebar	3	55	merespon	3
23	bentang	1	56	makmur	4
24	hendaknya	1	57	suka cita	4
25	silahkan	3	58	pengguna	1
26	semboyan	2	59	tunggu	1
27	ditunggu	2	60	lotre	2
28	akses	2	61	nggak	1
29	penerangan	2	62	kupon	3
30	hi	1	63	terpelihara	4
31	dibantu	2	∴	∴	∴
32	makasih	4	3593	asrama	3
33	halo	1	3594	orisinal	3

Lampiran 4 Kamus Indonesian Sentiment Lexicon Negatif

No	Kata	Skor	No	Kata	Skor
1	putus tali		34	coba	-1
	gantung	-2			
2	gelebah	-2	35	kangen	-3
3	gobar hati	-2	36	kalau	-1
4	isak	-5	37	maunya	-1
5	larat hati	-3	38	seandainya	-1
6	nelangsa	-3	39	marilah	-1
7	remuk redam	-5	40	bener	-1
8	tidak segan	-2	41	yaudah	-4
9	gemar	-1	42	nggak	-3
10	tak segan	-1	43	gatau	-1
11	sesal	-4	44	apaan	-4
12	pengen	-2	45	ngakak	-2
13	penghayatan	-2	46	atuh	-1
14	absorpsi	-1	47	sekali	-1
15	linu	-4	48	menarik hati	-1
16	salah benang	-1	49	cedayam	-2
17	sakit	-5	50	kece	-3
18	lara	-5	51	termakan	-1
19	zuhud	-1	52	belum	-1
20	mencederai	-4	53	malem	-1
21	mengingkari	-4	54	mencekau	-2
22	maaf	-3	55	menduga	-1
23	mengkhianat	-4	56	menyuarakan	-1
24	mencelakai	-5	57	memprediksi	-1
25	mulu	-1	58	membunyikan	-1
26	ngga	-2	59	menerka	-1
27	borong	-1	60	menaksir	-1
28	lever	-2	61	mengantisipasi	-1
29	kasian	-3	62	nangis	-5
30	gamau	-4	63	rompok	-2
31	doang	-1	:	:	:
32	pulas	-1	6655	tolol	-5
33	abis	-2	6656	tai	-5

Lampiran 5 Kamus Stopwords Indonesia

No	Kata	No	Kata	No	Kata
1	ada	34	ataupun	67	belum
2	adalah	35	awal	68	belumlah
3	adanya	36	awalnya	69	benar
4	adapun	37	bagai	70	benarkah
5	agak	38	bagaikan	71	benarlah
6	agaknya	39	bagaimana	72	berada
7	agar	40	bagaimanakah	73	berakhir
8	akan	41	bagaimanapun	74	berakhirilah
9	akankah	42		75	berakhirnya
10	akhir	43		76	berapa
11	akhiri	44	bahkan	77	berapakah
12	akhirnya	45	bahwa	78	berapalah
13	aku	46	bahwasanya	79	berapapun
14	akulah	47	baik	80	berarti
15	amat	48	bakal	81	berawal
16	amatlah	49	bakalan	82	berbagai
17	anda	50	balik	83	berdatangan
18	andalah	51	banyak	84	beri
19	antar	52	bapak	85	berikan
20	antara	53	baru	86	berikut
21	antaranya	54	bawah	87	berikutnya
22	apa	55	beberapa	88	berjumlah
23	apaan	56	begini	89	berkali-kali
24	apabila	57	beginian	90	berkata
25	apakah	58	beginikah	91	berkehendak
26	apalagi	59	beginilah	92	berkeinginan
27	apatah	60	begitu	93	berkenaan
28	artinya	61	begitukah	94	berlainan
29	asal	62	begitulah	95	berlalu
30	asalkan	63	begitupun	96	berlangsung
31	atas	64	bekerja	:	:
32	atau	65	belakang	758	yakni
33	ataukah	66	belakangan	759	yang

1. *Import Library*

```
import json
import pandas as pd

import numpy as np
import matplotlib.pyplot as plt

from tqdm import tqdm

from pygments import highlight
from pygments.lexers import JsonLexer
from pygments.formatters import TerminalFormatter

from google_play_scraper import Sort, reviews, app
```

2. *Scrapping Data Ulasan Aplikasi Reddoorz*

```
app_reviews = []

for ap in tqdm(app_packages):
    for score in list(range(1,6)):
        for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]:
            rvs, _ = reviews(
                ap,
                lang='id',
                country='id',
                sort = sort_order,
                count=200 if score==3 else 100,
                filter_score_with = score
            )
            for r in rvs:
                r['sortOrder'] = 'most_relevant' if sort_order ==
Sort.MOST_RELEVANT else 'newest'
                r['appId'] = ap
                app_reviews.extend(rvs)

app_reviews_df = pd.DataFrame(app_reviews)
app_reviews_df = app_reviews_df.sort_values(["appId", "at"])

app_reviews_df
```

Lampiran 7 Sintaks Python Pre-Processing Data

```
#Casefolding
lowerlist = []
for index, row in datanegatif.iterrows():
    lowerlist.append(row['content'].lower())
datanegatif['Lower']=lowerlist
```

```
#Remove Number
def removenumber(data):
    data = re.sub(r'\d+', '', data)
    return data
```

```
#Remove Punctuation
def punctuation(data):
    data = re.sub(r"[,.;@#?!&$]+\ *", " ", data)
    data = data.translate(str.maketrans('', '', string.punctuation))
    return data
```

```
#Remove non ASCII chars
def asciunicode(data):
    data = data.encode('ascii', 'ignore').decode('utf-8')
    data = re.sub(r'^\x00-\x7f', r'', data)
    data = re.sub(r'(\u[0-9A-Fa-f]+)', r'', data)
    data = re.sub(r"^[A-Za-z0-9^,!.\/'+-=]", " ", data)
    data = re.sub(r'\\u\\w\\w\\w\\w', ' ', data)
    return data
```

```
def normalisasi(dataframe, namakolom, namakolombaru):
    def scanner(sentence):
        words = sentence.lower().split()
        kalimat = []
        for word in words:
            try:
                kalimat.append(str(kamus[word]))
            except:
                kalimat.append(str(word))
        return ' '.join(kalimat)
    dataframe[namakolombaru]=dataframe.apply(lambda x:
scanner(str(x[namakolom])), axis=1)
    return dataframe)
```

```

#Apply Tokenizing
tokenasi = []
for index, row in datanegatif.iterrows():
    tokenasi.append(tokenize(row['Normalisasi']))
datanegatif['Tokenizing'] = tokenasi

```

```

# Pembuatan Kamus Lexicon
lexicon_positive = {}
import csv
with open('/content/drive/MyDrive/Thesis-Wulan/positive.csv', 'r')
as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        print (row)
        lexicon_positive[row[0]] = int(row[1])
    lexicon_negative = dict()
import csv
with open('/content/drive/MyDrive/Thesis-Wulan/negative.csv', 'r')
as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        lexicon_negative[row[0]] = int(row[1])
def sentiment_analysis_lexicon_indonesia(text):
    score = 0
    words = text.split()
    for word in words:
        if word in lexicon_positive:
            score += lexicon_positive[word]
        if word in lexicon_negative:
            score -= abs(lexicon_negative[word]) # Mengurangkan
    bobot negatif
    return score

```

```

def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]
datanegatif['Stopremover'] = df['Tokenizing'].apply(lambda x:
stopwords_removal(ast.literal_eval(x)))
factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()

```

```

def stemmed_wrapper(term):
    return stemmer.stem(term)

```

```

term_dict = {}

for document in datanegatif['Stopremover']:
    for term in ast.literal_eval(document) :
        if term not in term_dict:
            term_dict[term] = ' '

print(len(term_dict))
print("-----")

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    if term == term_dict[term] :
        print(term,":", term_dict[term])

# print(term_dict)
print("-----")

#Terapkan stemmer
stem = []
for index, row in datanegatif.iterrows():
    stem.append(stemmer.stem(row['Stopremover']))
datanegatif['Stemmer'] = stem

```

Lampiran 8 TF-IDF

```

import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import TfidfVectorizer

#menggabungkan daftar token menjadi dokumen string tunggal
review_reddoorz = pd.read_excel('/content/drive/My Drive/Thesis-
Wulan/stemmer.xlsx')
def join_text_list(texts):
    texts = ast.literal_eval(texts)
    return ' '.join([text for text in texts])
review_reddoorz['content_text'] =
review_reddoorz['Tokenizing'].apply(join_text_list)

cv = CountVectorizer()
word_count_vector=cv.fit_transform(review_reddoorz['content_text'])
tf = pd.DataFrame(word_count_vector.toarray(),
columns = cv.get_feature_names_out())
tf.to_excel('reddoorz_tf.xlsx') # Export to excel file

```

```

tfidf_transformer = TfidfTransformer()
xtfidf = tfidf_transformer.fit_transform(word_count_vector)
idf =
pd.DataFrame({'feature_name':cv.get_feature_names_out(),'idf_weights':tfidf_transformer.idf_})
idf.to_excel('reddoorz_idf.xlsx') # Export to excel file

```

```

tf_idf =
pd.DataFrame(xtfidf.toarray(),columns=cv.get_feature_names_out())
tf_idf.to_excel('reddoorz_tfidf.xlsx') # Export to excel file

```

Lampiran 9 Naive Bayes Classifier

```

from sklearn import model_selection
from sklearn import metrics
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score

```

```

tfidf = pd.read_excel('reddoorz_tfidf.xlsx')
tfidf_rd = tfidf.loc[:, ~tfidf.columns.str.contains('^Unnamed')]
review_reddoorz = pd.read_excel('/content/drive/My Drive/Thesis-
Wulan/data_label.xlsx')

```

```

x = tfidf_rd
polarity_encode = {'negative' : 0, 'positive' : 1}
y = review_reddoorz['polarity'].map(polarity_encode).values
xtrain, xtest, ytrain, ytest=train_test_split(x,y,test_size=0.2,
random_state=7)
print(ytrain.shape)
print(ytest.shape)

```

```

mnb = MultinomialNB().fit(xtrain, ytrain)
ypred = mnb.predict(xtest)
print(ypred)

```

```

conf_mat = confusion_matrix(ytest,ypred)
print(pd.crosstab(ytest, ypred, rownames=['True'],
colnames=['Predicted'], margins=True))
print('=====\n')
print(metrics.classification_report(ytest,ypred))
print("Accuracy:", accuracy_score(ytest,ypred))
print("Precision:", precision_score(ytest,ypred))
print("Recall:", recall_score(ytest,ypred))
print("F-measure:", f1_score(ytest,ypred))

```

Lampiran 10 Support Vector Machine

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.metrics import classification_report, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
from imblearn.over_sampling import SMOTE
polarity_counts = data['polarity'].value_counts()

# Plot the distribution of unique values in the 'polarity' column
plt.figure(figsize=(10, 6))
sns.barplot(x=polarity_counts.index, y=polarity_counts.values,
palette='viridis')
plt.title('Distribution of Unique Values in Polarity Column')
plt.xlabel('Polarity')
plt.ylabel('Count')
plt.show()
vectorizer = TfidfVectorizer(max_features=5000)
X_tfidf = vectorizer.fit_transform(X)
smote = SMOTE(random_state=42)
X, y = smote.fit_resample(X_tfidf, y)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
svm = SVC(kernel="rbf", gamma=0.5, C=1.0)
svm.fit(X_train, y_train)

```

Lampiran 11 Apriori

```

pip install apyori
from apyori import apriori
from collections import Counter
import re

```

```
#Cari 10 kata terbanyak
from collections import Counter
Counter(" ".join(datanegatif["Stemmer"]).split()).most_common(10)
```

```
#Association Rule dengan Algoritma Apriori
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
te = TransactionEncoder()
Dataneg = datanegatif['Stemmer'].apply(lambda x: x.split(" "))
Dataneg
te_ary = te.fit(Dataneg).transform(Dataneg)
frequent_itemsets = apriori(Dataneg2, min_support = 0.004,
max_len=2, use_colnames=True)
res = association_rules(frequent_itemsets, metric='confidence',
min_threshold=0.1)
ARaplikasi =
res1[res1['antecedents'].isin(["frozenset({'aplikasi'})"]) == True]
ARhotel = res1[res1['antecedents'].isin(["frozenset({'hotel'})"])
== True]
```

