

**ANALISIS SENTIMEN PERSEPSI PENGGUNA PEDULILINDUNGI
MENGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER* (NBC) DAN
SUPPORT VECTOR MACHINE (SVM)**

TUGAS AKHIR

**Diajukan Sebagai salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Pada Jurusan Teknik Industri Fakultas Teknologi Industri**



Disusun Oleh:

Nama : Rizki Rahmattullah

No. Mahasiswa : 18522291

**PROGRAM STUDI TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2022

LEMBAR PERNYATAAN KEASLIAN

Demi Allah saya mengakui bahwa karya ini merupakan karya saya sendiri kecuali kutipan dan ringkasan yang salah satunya telah dicantumkan sumbernya. Jika ditemukan dikemudian hari ternyata terbukti pengakuan saya yang tidak benar dan melanggar peraturan yang sah dalam karya tulis ini dan hak kekayaan intelektual maka saya bersedia ijazah yang saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 17 Agustus 2022



Rizki Rahmattullah
18522291



SURAT KETERANGAN PENELITIAN
Nomor : 238/A/Ka.Lab DATMIN/FTI-UIII/VIII/2022

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Rizki Rahmattullah
No. Mhs : 18522291
Dosen Pembimbing : Annisa Uswatun Khasanah, ST., M.B.A., M.Sc

Telah selesai melaksanakan penelitian yang berjudul "ANALISIS SENTIMEN PERSEPSI PENGGUNA PEDULI LINDUNGI MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (NBC) DAN SUPPPORT VECTOR MACHINE (SVM)" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 30 Maret 2022 sampai dengan tanggal 05 Mei 2022

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 26 Agustus 2022

Kepala Laboratorium
Data Mining

Annisa Uswatun Khasanah, ST., M.B.A., M.Sc.

LEMBAR PENGESAHAN DOSEN PEMBIMBING

**ANALISIS SENTIMEN PERSEPSI PENGGUNA PEDULILINDUNGI
MENGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER* (NBC) DAN
SUPPPORT VECTOR MACHINE (SVM)**



Dosen Pembimbing,

Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

LEMBAR PENGESAHAN DOSEN PENGUJI**ANALISIS SENTIMEN PERSEPSI PENGGUNA PEDULILINDUNGI
MENGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER* (NBC) DAN
SUPPPORT VECTOR MACHINE (SVM)****TUGAS AKHIR**

Disusun Oleh:

Nama : Rizki Rahmattullah

No. Mahasiswa : 18522291

Telah dipertahankan didepan sidang penguji sebagai salah satu syarat untuk
memperoleh gelar Sarjana Strata S-1 Teknik Industri

Yogyakarta, 22 Agustus 2022

Tim Penguji

Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

Ketua

Sri Indrawati, S.T., M.Eng

Anggota I

Abdullah 'Azzam, S.T., M.T.

Anggota II

Mengetahui,

Ketua Program Studi Teknik Industri

Fakultas Teknologi Industri

Universitas Islam Indonesia

Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.

LEMBAR PERSEMBAHAN

*Dengan Menyebut Nama Allah yang Maha Pengasih lagi Maha Penyayang.
Puji Syukur atas Kehadiran Allah SWT atas nikmat yang telah beliau berikan, Skripsi
ini saya persembahkan kepada:*

Bapak Januardi dan Ibu Elyza selaku kedua orang tua saya. Terima kasih telah selalu mendoakan, mendidik, dan memberi semangat kepada saya. Terima kasih yang selalu menjadi sosok utama dan segalanya didalam hidup saya yang tidak pernah tergantikan. Terima kasih atas pengorbanan dan segala hal yang tidak bisa diungkapkan kata dan tak akan pernah bisa dibalas dengan apapun, This litte las anent for you.

Kakak saya Reza Mayasari dan Rahmi Afrizani, terima kasih telah memberikan dukungan, semangat, dan motivasi sampai pada tahap ini untuk menjadi las an dalam menyelsaikan studi hingga akhir.

Serta kerabat, sahabat, dan teman-teman saya yang selalu mendukung, menemani dan membantu saya setiap hari dari awal hingga akhir masa perkulihan ini.



MOTTO

يُحِبُّ اللَّهُ الْعَامِلَ إِذَا عَمِلَ أَنْ تَحْسِنَ. رواه الطز ان نى

“Allah mencintai pekerjaan yang apabila bekerja ia menyelesaikannya dengan baik”.

(HR. Thabrani)

مَنْ خَرَجَ فِي طَلَبِ الْعِلْمِ فَهُوَ فِي سَبِيلِ اللَّهِ

“Barang siapa keluar untuk mencari Ilmu maka dia berada di jalan Allah “.

(HR. Turmidzi)

إِنَّا أَعْطَيْنَاكَ الْكَوْثَرَ

"Sesungguhnya Kami telah memberimu (Muhammad) nikmat yang banyak."

(Al-Kautsar: 1)

KATA PENGANTAR

Assalamualaikum Wr. Wb.

Alhamdulillahirobbil'alamin, segala puji syukur penulis panjatkan atas kehadiran Allah SWT atas limpahan rahmat dan hidayah yang diberikan-Nya. Shalawat serta salam senantiasa tercurah kepada Baginda Nabi Muhammad SAW atas petunjuk untuk selalu berada di jalan-Nya sehingga penulis menyusun dan menyelesaikan laporan ini. Laporan Tugas Akhir/Skripsi yang berjudul “Analisis Sentimen Persepsi Pengguna PeduliLindungi Menggunakan Algoritma *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM)” dapat terlaksana dan terselesaikan dengan baik. Adapun Tugas Akhir ini disusun sebagai salah satu syarat utama untuk menyelesaikan studi Strata-1 pada Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Laporan Tugas Akhir ini kiranya tidak akan selesai tanpa bantuan dari beberapa pihak yang terus membantu dan mendorong penulis untuk menyelesaikannya. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Ir. Hari Purnomo, M.T. selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
2. Dr. Drs. Imam Djati Widodo, M.Eng.Sc. selaku Ketua Jurusan Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T. M.Sc., Ph.D., IPM. selaku Ketua Program Studi Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.
4. Ibu Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc. selaku dosen pembimbing Tugas Akhir yang telah membimbing dalam pelaksanaan dan penyusunan laporan Tugas Akhir
5. Kedua orang tua penulis, Bapak Januardi dan Ibu Elyza yang telah memberikan doa, semangat, motivasi dan kasih sayang yang diberikan kepada saya sehingga penyelesaian Tugas Akhir ini dapat terlaksana dengan lancar.
6. Keluarga Laboratorium Data Mining angkatan 2017, 2018, dan 2019, kepala laboratorium Ibu Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc., laboran mas Bayu Hertanta, dan teman-teman yang telah memberikan doa dan dukungan terutama sahabat seperjuangan Tyas, Acad, Lala, Jihan, dan Rahma.

7. Sahabat penulis Cindy Munawaroh P.P, Salsarillah Syaraini E, dan Fikri Septrian Anggara, yang telah senantiasa berbagi ilmu dan dukungan, serta mendengarkan keluh kesah penulis.
8. Sahabat seperjuangan Rizky Restiana, Fahrul Triyulianto R, yang telah senantiasa memberikan doa, berbagi ilmu serta mendengarkan keluh kesah penulis.
9. Teman-teman Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia khususnya Angkatan 2018.

Demikian pnulisan laporan Tugas Akhir ini disusun, penulis menyadari dalam menyusun dan menulis laporan Tugas Akhir ini masih memiliki kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun untuk perbaikan di masa yang akan datang dari seluruh pembaca. Penulis juga berharap semoga laporan ini bisa bermanfaat bagi pembacanya ataupun penelitian selanjutnya.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 17 Agustus 2022



Rizki Rahmatullah

NIM 18522291

ABSTRAK

Perkembangan virus corona di Indonesia merupakan salah satu negara dengan kasus tertinggi. Hal ini membuat pemerintah berinisiatif melakukan pemantauan untuk meminimalisir penyebarannya dengan membuat pelayanan yang bernama PeduliLindungi. Aplikasi PeduliLindungi mengandalkan masyarakat untuk membagikan data lokasinya. Aplikasi ini sudah di-*download* sebanyak 50 juta kali dan ulasan yang ada cukup memenuhi harapan pengguna. Namun, perlu adanya evaluasi untuk meningkatkan pelayanan tersebut. Untuk mendapatkan informasi terkait perbaikan layanan diperlukan metode untuk membantu mengekstrak dan memilih ulasan secara efisien. Salah satu metode yang digunakan yaitu *text mining* dengan menggunakan algoritma *Naive Bayes Classifier* (NBC) Dan *Support Vector Machine* (SVM) serta dianalisis dengan menggunakan asosiasi kata. Berdasarkan tiga perbandingan *data training* dan *data testing* dengan percobaan sebanyak tiga kali disetiap perbandingan diketahui bahwa akurasi tertinggi metode SVM memiliki 95.01% dan pada metode NBC akurasi tertinggi memiliki 87.68%. Selanjutnya dilakukan asosiasi kata pada sentimen positif yaitu aplikasi, bantu, vaksin, sertifikat, kasih, terima, baru, bagus, masuk dan manfaat. Sedangkan pada kelas sentimen negatif yaitu aplikasi, vaksin, sertifikat, sulit, masuk, tolong, buka, data, *update*, tanggal, lahir, muncul. Hasil ulasan yang didapatkan selanjutnya akan dilakukan analisis perbaikan dengan diagram fishbone sebagai pemecah permasalahan.

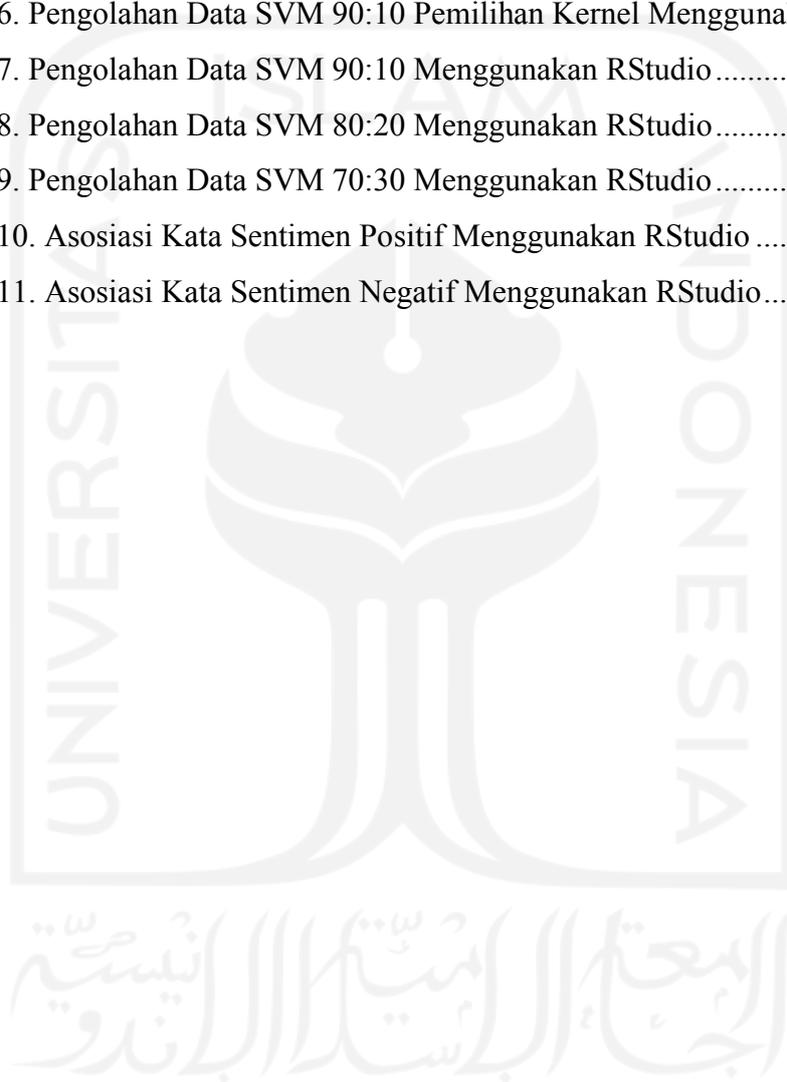
Keyword: Analisis Sentimen, *Machine Learning*, *Naive Bayes Classifier* (NBC), PeduliLindungi, *Support Vector Machine* (SVM)

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN.....	ii
KETERANGAN SELESAI PENELITIAN.....	iii
LEMBAR PENGESAHAN DOSEN PEMBIMBING	iv
LEMBAR PENGESAHAN DOSEN PENGUJI.....	v
LEMBAR PERSEMBAHAN	vi
MOTTO	vii
KATA PENGATAR	viii
ABSTRAK.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan Penelitian.....	5
BAB II KAJIAN LITERATUR.....	8
2.1 Kajian Induktif	8
2.2 Kajian Deduktif.....	17
2.2.1 Machine Learning.....	17
2.2.2 Text Mining	18
2.2.3 Analisis Sentimen	20
2.2.4 Pembobotan (<i>Term Weighting</i>).....	20
2.2.5 <i>Web Scraping</i>	21
2.2.6 Klasifikasi	22
2.2.7 Naïve Bayes Classifier (NBC).....	24
2.2.8 Support Vector Machine (SVM).....	25
2.2.9 Asosiasi Kata	25
BAB III METODOLOGI PENELITIAN	27

3.1	Objek Penelitian	27
3.2	Metode Pengumpulan Data	27
3.3	Variabel Penelitian	27
3.4	Jenis dan Sumber Data	27
3.3.1.	Data Primer	27
3.3.2.	Data Sekunder	28
3.5	Instrumen Penelitian	28
3.6	Metode Analisis Data	28
3.7	Prosedur Eksperimen	29
4.1.	Pengumpulan Data	34
4.2.	Pengolahan Data	36
4.2.1	Analisis Deskriptif	36
4.2.2	Pre-processing Data	37
4.2.2.1.	Data Cleaning	38
4.2.2.2.	Case Folding	38
4.2.2.3.	Tokenizing	39
4.2.2.4.	Stemming	40
4.2.3	Representasi Model	40
4.2.4	Analisis Klasifikasi	44
4.2.4.1.	Pembagian Data Training dan Data Testing	44
4.2.4.2.	Analisis Klasifikasi Pada Algoritma <i>Naïve Bayes Classifier</i>	46
4.2.4.3.	Analisis Klasifikasi Pada Algoritma <i>Support Vector Machine</i>	49
4.2.5	Visualisasi dan Asosiasi	52
4.2.5.1.	Ulasan Positif	52
4.2.5.2.	Ulasan Negatif	55
BAB V	PEMBAHASAN	59
5.1	Gambaran Umum Terkait Persepsi Pengguna Aplikasi PeduliLinungi	59
5.2	Hasil Klasifikasi Pada <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i> ..	60
5.3	Hasil Klasifikasi dan Asosiasi Kata	62
5.4	Rekomendasi Berdasarkan Usulan Negatif	64
BAB VI	PENUTUP	69
6.1	Kesimpulan	69
6.2	Saran	70

DAFTAR PUSTAKA	71
Lampiran 1. Tahap <i>Pre-Processing</i> Menggunakan Python.....	77
Lampiran 2. Analisis Sentimen Menggunakan RStudio.....	79
Lampiran 3. Pengolahan Data NBC 90:10 Menggunakan RStudio	80
Lampiran 4. Pengolahan Data NBC 80:20 Menggunakan RStudio	82
Lampiran 5. Pengolahan Data NBC 70:30 Menggunakan RStudio	84
Lampiran 6. Pengolahan Data SVM 90:10 Pemilihan Kernel Menggunakan RStudio..	86
Lampiran 7. Pengolahan Data SVM 90:10 Menggunakan RStudio.....	87
Lampiran 8. Pengolahan Data SVM 80:20 Menggunakan RStudio.....	88
Lampiran 9. Pengolahan Data SVM 70:30 Menggunakan RStudio.....	89
Lampiran 10. Asosiasi Kata Sentimen Positif Menggunakan RStudio	90
Lampiran 11. Asosiasi Kata Sentimen Negatif Menggunakan RStudio.....	92



DAFTAR TABEL

Tabel 2. 1 Kelebihan Algoritma <i>Support Vector Machine</i> (SVM) dan <i>Naïve Bayes Classifier</i> (NBC)	11
Tabel 4. 1 Pengubahan Kata menjadi huruf kecil	38
Tabel 4. 2 Menghapus angka pada ulasan	39
Tabel 4. 3 Menghapus tanda baca pada ulasan	39
Tabel 4. 4 Pemisahan kata pada ulasan menjadi token	40
Tabel 4. 5 Mengubah kata pada ulasan menjadi kata dasar	40
Tabel 4. 6 Pembobotan Menggunakan Metode Kamus Lexicon.....	41
Tabel 4. 7 Perhitungan Pembobotan	41
Tabel 4. 8 Pelabelan Data Ulasan	42
Tabel 4. 9 Perbandingan Data Training dan Data Testing 1	45
Tabel 4. 10 Perbandingan <i>Data Training</i> dan <i>Data Testing</i> 2.....	45
Tabel 4. 11 Perbandingan <i>Data Training</i> dan <i>Data Testing</i> 3.....	45
Tabel 4. 12 Hasil Klasifikasi NBC Dengan Perbandingan 90:10.....	46
Tabel 4. 13 Evaluasi Dengan <i>Confusion Matrix</i> Pada Percobaan 2	47
Tabel 4. 14 Hasil Klasifikasi NBC Dengan Perbandingan 80:20.....	47
Tabel 4. 15 Evaluasi Dengan <i>Confusion Matrix</i> Pada Percobaan 1	48
Tabel 4. 16 Hasil Klasifikasi NBC Dengan Perbandingan 70:30.....	48
Tabel 4. 17 Evaluasi Dengan <i>Confusion Matrix</i> Pada Percobaan 1	49
Tabel 4. 18 Perbandingan Akurasi Pada SVM	49
Tabel 4. 19 Hasil Klasifikasi SVM Dengan Perbandingan 90:10	50
Tabel 4. 20 Evaluasi Dengan <i>Confusion Matrix</i>	50
Tabel 4. 21 Hasil Klasifikasi SVM Dengan Perbandingan 80:20	50
Tabel 4. 22 Evaluasi Dengan <i>Confusion Matrix</i> Pada Percobaan 3	51
Tabel 4. 23 Hasil Klasifikasi SVM Dengan Perbandingan 70:30	51
Tabel 4. 24 Evaluasi Dengan <i>Confusion Matrix</i>	52
Tabel 5. 1 Perbandingan Akurasi Model Klasifikasi Algoritma NBC dan SVM.....	60
Tabel 5. 2 Usulan Pada Permasalahan Aplikasi PeduliLindungi.....	66

DAFTAR GAMBAR

Gambar 2. 1 Tahapan Text Mining.....	18
Gambar 2. 2 Tahapan Analisis Klasifikasi	23
Gambar 2. 3 Hyperplane pada Support Vector Machine.....	25
Gambar 3. 1 Diagram Alur Penelitian	30
Gambar 4. 1 Interface awal pada situs Appfollow.io.....	34
Gambar 4. 2 Website ulasan PeduliLindungi	35
Gambar 4. 3 Pengambilan Ulasan.....	35
Gambar 4. 4 Export Data	36
Gambar 4. 5 Hasil Scraping Data	36
Gambar 4. 6 Jumlah Ulasan Pengguna PeduliLindungi	37
Gambar 4. 7 Visualisasi Hasil Pelabelan Sentimen sebelum reduksi.....	43
Gambar 4. 8 Visualisasi Hasil Pelabelan Sentimen sebelum reduksi.....	44
Gambar 4. 9 Visualisasi Data Sentimen Positif Menggunakan Barchart	53
Gambar 4. 10 Visualisasi Ulasan Positif Dengan Menggunakan <i>Wordcloud</i>	53
Gambar 4. 11 Asosiasi Kata Pada Ulasan Kelas Sentimen Positif.....	54
Gambar 4. 12 Visualisasi Data Sentimen sentimen Menggunakan <i>Barchart</i>	56
Gambar 4. 13 Visualisasi Ulasan Negatif Dengan Menggunakan <i>Wordcloud</i>	56
Gambar 4. 14 Asosiasi Kata Pada Ulasan Kelas Sentimen Negatif	57
Gambar 5. 1 Diagram Fishbone Terkait Permasalahan Login.....	64
Gambar 5. 2 Diagram Fishbone Terkait Permasalahan Tidak Terdapat Notifikasi.....	65
Gambar 5. 3 Diagram Fishbone Terkait Permasalahan Sulit Melakukan Scan.....	65
Gambar 5. 4 Diagram Fishbone Terkait Permasalahan Sulit Menyimpan Bukti Vaksin.....	66

BAB I PENDAHULUAN

1.1 Latar Belakang

Coronavirus Disease 2019 (COVID-19) adalah penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Virus penyebab COVID-19 ini dinamakan Sars-CoV-2. Virus corona adalah zoonosis (ditularkan antara hewan dan manusia). Adapun hewan yang menjadi sumber penularan COVID-19 masih belum diketahui. COVID-19 dapat menular dari manusia ke manusia melalui percikan bersin/batuk. Orang yang paling berisiko tertular penyakit ini adalah orang yang berkontak erat dengan pasien COVID-19 termasuk perawat yang merawat (Kementerian Kesehatan RI, 2020).

Seiring pesatnya perkembangan virus corona, Indonesia merupakan salah satu negara yang terdampak kasus virus corona tertinggi. Virus ini masuk ke Indonesia pada awal tahun 2020. Dimana, Presiden Joko Widodo melaporkan pertama kali menemukan dua kasus infeksi COVID-19 di Indonesia pada 2 maret 2020 (Djalante et al., 2020). Penambahan jumlah kasus COVID-19 berlangsung cukup cepat dan sudah terjadi penyebaran antar negara. Sampai tanggal 15 Maret 2022, dilaporkan total kasus terkonfirmasi 4.844.279 dengan kasus kematian sebesar 145.321 (Kementerian Kesehatan Republik Indonesia, 2022).

Semakin maraknya kasus COVID-19 di Indonesia membuat pemerintah melakukan Gerakan pencegahan, dengan cara: Menjaga Kesehatan dan kebugaran, mencuci tangan, menjaga jarak (minimal 1 meter), menggunakan masker, meminimalisir berpergian keluar, memantau perkembangan COVID-19, menunda perjalanan ke luar kota/negeri (Kementerian Kesehatan Republik Indonesia, 2022). Selain melakukan hal-hal yang dianjurkan pemerintah, untuk lebih meminimalisir penyebaran virus ini, pemerintah melakukan pemantauan dengan menggunakan bantuan aplikasi yang bernama PeduliLindungi.

Aplikasi PeduliLindungi adalah aplikasi pelayanan yang dikembangkan untuk membantu suatu instansi pemerintahan terkait pelacakan untuk menghentikan penyebaran *Coronavirus Disease (COVID-19)*. Selain itu, aplikasi mengandalkan masyarakat Indonesia untuk saling membagikan data lokasinya saat bepergian agar dapat mengetahui riwayat kontak dengan penderita COVID-19 (Windy, 2021). Aplikasi ini

merupakan salah satu alternatif yang sangat efektif dan efisien dalam memantau masyarakat yang sedang melakukan aktivitas diluar rumah. Selain itu digunakan untuk aktivitas luar ruangan, aplikasi ini kerap digunakan saat melakukan perjalanan jauh, seperti: melakukan perjalanan ke luar kota maupun ke luar negeri dengan memakan waktu yang lama.

Saat ini aplikasi PeduliLindungi sudah didownload sebanyak 50 juta kali dengan *rating* 4,4 dengan 803 ribu ulasan pada *Google Play Store* sedangkan pada *AppStore* *rating* dalam penggunaan aplikasi ini sebesar 2.6 dari 5. Terjadinya perbedaan pemberian *rating* dan ulasan pada aplikasi PeduliLindungi ditoko aplikasi pada dua sistem operasi yang berbeda dikarenakan masyarakat Indonesia sendiri dalam menggunakan *smartphone* dominan menggunakan dengan sistem operasi berbasis *android* (Darmadi, et al., 2021). *Rating* yang diberikan oleh pengguna pada aplikasi PeduliLindungi tidak memiliki hubungan dengan ulasan yang diberikan pengguna. Hal ini terjadi dikarenakan cara pengguna dalam memberikan ulasan dan *rating* yang dilakukan secara *online*. Mengulas dengan memberikan *rating* dan *review* suatu aplikasi secara *online* dapat membuat perbedaan dalam pemberian saran perbaikan kepada pemerintah.

Ulasan yang lumayan banyak diberikan oleh pengguna pada *Google Play Store* tidak menjamin bahwa aplikasi yang digunakan tidak memiliki kekurangan. Berbagai ulasan positif dan negatif menunjukkan bahwa pelayanan yang diberikan oleh PeduliLindungi cukup memenuhi harapan pengguna aplikasi PeduliLindungi. Keluhan-keluhan yang diberikan oleh pengguna aplikasi PeduliLindungi terkait pelayanan yang diberikan karena fitur yang diberikan sulit pahami dan selalu terjadi gagal login, selain itu juga dari pihak layanan kurang memberikan cepat memberikan respon terkait keluhan pengguna (Rachmatunnisa, 2021). Oleh sebab itu, perlu adanya evaluasi untuk meningkatkan pelayanan dan kinerja kepada pengguna/masyarakat. Dimana, evaluasi yang akan dilakukan dilihat berdasarkan dari ulasan yang diberikan.

Untuk mendapatkan suatu informasi yang dibutuhkan terkait pelayanan saat menggunakan suatu produk atau aplikasi, dapat digunakan salah satu upayanya yaitu dengan melihat dari ulasan pengguna atau *user*. Didalam suatu komentar atau ulasan mengenai suatu produk, tidak semuanya memberikan reaksi yang positif melainkan terdapat juga reaksi negatif. Dalam meningkatkan kualitas pelayanan, suatu perusahaan akan melihat dari ulasan yang telah diberikan oleh pengguna ketika menggunakan aplikasi tersebut. Semakin banyak ulasan yang diberikan, maka akan semakin sulit pula

perusahaan akan mendapatkan informasi yang dibutuhkan. Untuk mendapatkan informasi yang sesuai dengan kebutuhan, diperlukan suatu metode yang dapat membantu dalam dalam mengekstrak dan memilih secara efektif dan efisien. Metode yang dapat digunakan dalam membantu permasalahan tersebut salah satunya dengan menggunakan metode *Text Mining*. Salah satu teknik yang terdapat pada metode *text mining* sendiri yaitu analisis sentimen. Analisis Sentimen digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini (Sari & Wibowo, 2019). Informasi yang didapatkan dengan dilakukannya analisis sentimen tersebut dapat dijadikan perusahaan dalam melakukan landasan dalam menginovasi maupun meningkatkan kinerja aplikasi. Pada penelitian ini, dalam melakukan analisis sentimen dibantu dengan pendekatan klasifikasi. Metode ini tergolong kedalam *supervised learning*. dalam sepuluh tahun terakhir dapat dilihat kemajuan yang belum terjadi sebelumnya, salah satu kemajuan yaitu pada *machine learning*. algoritma pada *machine learning* ini terdiri dari *Bayesian Classifier*, *Decision Tree*, *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Neural Networks* (Ranjan, et al., 2015).

Dalam mendapatkan informasi dari pengguna untuk dijadikan landasan, dilakukanlah analisis sentimen dengan pendekatan klasifikasi. Dalam penelitian ini terdapat dua algoritma yang digunakan, yaitu *Naïve Bayes Classifier* (NBC) dan *Support Machine Vector* (SVM). Digunakannya algoritma *Naïve Bayes Classifier* (NBC) ini merupakan algoritma klasifikasi statistik yang dapat digunakan untuk mengestimasi probabilitas kedalam *sub group* dan memiliki performansi akurasi tinggi dan dapat digunakan pada *database* yang besar (Saleh, 2015). Selain itu juga, algoritma ini pada umumnya digunakan dalam melakukan pengkategorisasian teks (Samuel, et al., 2014) Menggunakan algoritma *Support Vector Machine* (SVM) karena dapat menemukan *hyperplane* yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). Selain itu, saat menentukan jarak menggunakan *Support Vector Process* komputasi menjadi cepat (Vapnik & Cortes, 1995).

Pemilihan algoritma yang digunakan pada penelitian ini sendiri didasarkan karena pada algoritma *Random Forest* jika dalam proses pembuatan pohon keputusannya banyak maka hal ini akan mengakibatkan algoritma ini menjadi lambat dan sulit untuk diinterpretasikan (Chan, et al., 2022), pada algoritma *K-Nearest Neighbor* (KNN) karena rentan terhadap variabel yang non-informatif dan dimensionalitas yang tinggi (Yustanti,

2012), pada algoritma *Decision Tree* karena untuk melatih model membutuhkan waktu dan memori yang cukup banyak, selain itu juga ketika terdapat perubahan kecil yang terdapat pada data maka akan menyebabkan perubahan besar pada struktur *decision tree* (Gie, 2021), sedangkan pada algoritma *Deep Learning* karena jika menggunakan data yang sedikit saat melakukan pengujian maka akan sulit untuk melakukan pengintegrasian dalam menyalurkan hasilnya (APPKEY, 2019).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah didapatkan, maka dapat diketahui beberapa rumusan masalah yang akan dikaji pada penelitian ini:

1. Bagaimana analisis sentimen ulasan yang didapatkan pada aplikasi PeduliLindungi pada situs *Google Play* dan pembentukan kata yang didapatkan berdasarkan klasifikasi kata dan asosiasi kata menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM)?
2. Bagaimana performansi yang dihasilkan dari penggunaan algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM)?
3. Bagaimana identifikasi faktor permasalahan yang dihadapi pengguna aplikasi PeduliLindungi menggunakan diagram *fishbone*?

1.3 Tujuan Penelitian

Berdasarkan masalah diatas, penelitian ini dibuat untuk menjawab beberapa tujuan dibawah ini

1. Mampu mengetahui analisis sentimen ulasan yang didapatkan pada aplikasi PeduliLindungi pada situs *Google Play* dan membentuk dan menganalisa kata yang didapatkan berdasarkan klasifikasi kata dan asosiasi kata kedalam kelas positif maupun negatif dengan menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).
2. Mampu menganalisa performansi yang dihasilkan dari penggunaan algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).
3. Mampu mengidentifikasi faktor permasalahan yang dihadapi pengguna aplikasi PeduliLindungi dengan menggunakan diagram *fishbone*.

1.4 Batasan Masalah

Pada penelitian kali ini, terdapat beberapa batasan masalah, diantaranya adalah:

1. Data yang digunakan merupakan data ulasan mengenai aplikasi PeduliLindungi pada situs *Google Play*.
2. Ulasan aplikasi PeduliLindungi diambil dari tanggal 1 November 2021 sampai 30 April 2022 dengan menggunakan versi 4.3 hingga 4.4.3.1.
3. Data ulasan yang digunakan merupakan data ulasan yang diambil menggunakan Bahasa Indonesia.
4. Penelitian ini menggunakan software Python dan Rstudio.

1.5 Manfaat Penelitian

Manfaat pada penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum mengenai persepsi pengguna aplikasi PeduliLindungi berdasarkan data ulasan pada situs *Google Play*.
2. Hasil analisis sentimen yang dihasilkan dapat dijadikan sebagai bahan evaluasi pihak perusahaan terhadap kepuasan pengguna aplikasi PeduliLindungi yang berguna untuk melakukan pengembangan pelayanan dalam memantau masyarakat yang beraktivitas diluar ruangan.
3. Memberikan *insight* baru mengenai analisis sentimen dengan menggunakan dua algoritma yang berbeda dengan kapasitas penggunaan data yang berbeda.
4. Memudahkan pihak yang memiliki kepentingan untuk mencari atau melihat informasi yang terdapat pada ulasan yang sangat banyak.

1.6 Sistematika Penulisan Penelitian

Dibuatnya sistematika penulisan dalam melakukan penelitian bertujuan agar mampu ditulis secara rapi dan sistematis. Sistematika penulisan ini sendiri dibuat berdasarkan akar permasalahan hingga didapatkan solusi penyelesaian permasalahan yang dibagi kedalam beberapa bab. Berikut merupakan sistematika penulisan pada penelitian kali ini:

BAB I PENDAHULUAN

Bab ini berisikan Latar Belakang, rumusan masalah, batasan masalah, tujuan masalah, manfaat penelitian, dan sistematika penulisan. Secara umum, pada bab ini berisi tentang permasalahan yang akan diangkat

oleh peneliti berdasarkan sumber data yang didapat dengan melihat objek penelitian.

BAB II KAJIAN LITERATUR

Bab ini memiliki 2 sub bab utama yaitu kajian deduktif dan kajian induktif. Kajian deduktif sendiri merupakan uraian teori yang pernah dijelaskan oleh peneliti terdahulu berdasarkan penelitian terdahulu, sedangkan kajian induktif berisi landasan teori yang dijadikan bahan pendukung dalam bentuk fakta dan kesimpulan yang memiliki hubungan dengan penelitian yang dilakukan.

BAB III METODE PENELITIAN

Bab ini berisi mengenai uraian alur dalam melakukan penelitian seperti alur penelitian, populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data, dan metode analisis data. Secara garis besar bab ini menjelaskan bagaimana keseluruhan penelitian dilakukan yang akan dijelaskan dalam diagram alur dan bagaimana data yang didapatkan diproses.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini berisi mengenai data yang didapatkan selama melakukan penelitian dan proses pengolahan data. Proses pengumpulan data merupakan bagaimana data yang telah didapatkan lalu diakumulasi, dan kemudian akan dilanjutkan ke proses pengolahan data dimana maksudnya yaitu proses mempersiapkan data, mengolah, hingga memvisualisasikan data yang didapat.

BAB V ANALISA DAN PEMBAHASAN

Bab ini akan dilakukan analisa dan pembahasan dari hasil yang diperoleh dalam penelitian ini. Analisis yang dilakukan disesuaikan dengan teknik yang digunakan dengan didapatkannya berupa rekomendasi.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisikan ringkasan dari analisis yang telah dilakukan berdasarkan penelitian yang dilakukan. Kesimpulan yang didapatkan merupakan jawaban dari rumusan masalah yang dipertanyakan. Saran yang akan diberikan sendiri merupakan solusi yang dapat diberikan kepada peneliti selanjutnya maupun perusahaan.

DAFTAR PUSTAKA**LAMPIRAN**

BAB II KAJIAN LITERATUR

2.1 Kajian Induktif

Berdasarkan penelitian yang dilakukan oleh (Hidayat & Minati, 2019) mengenai analisis komparasi dalam algoritma klasifikasi *text mining* untuk *translate* Qur'an Inggris dan Indonesia. Dijelaskan bahwa algoritma yang digunakan dalam membantu untuk melakukan analisis komparasi tersebut yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, *K-Nearest Neighbor* (KNN), dan *J48 Decision Tree Classifier*. Hasil penelitian yang didapatkan berdasarkan permasalahan tersebut untuk komparasi pada *translate* Indonesia didapatkan bahwa akurasi SVM tertinggi dengan nilai sebesar 81.4% sedangkan komparasi Inggris didapatkan bahwa *Naïve Bayes* tertinggi dengan nilai sebesar 78.3%.

Penelitian yang dilakukan oleh (Winanto & Budihartini, 2022) mengenai Perbandingan Akurasi Analisis Sentimen Pada *Twitter* Pemerintahan Provinsi DKI Jakarta Selama waktu vaksinasi COVID-19. Menjelaskan bahwa algoritma yang digunakan dalam membantu analisis yaitu dengan *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Hasil penelitian yang didapatkan berdasarkan permasalahan tersebut didapatkan bahwa nilai akurasi yang dihasilkan pada metode KNN sebesar 81.50%, sedangkan akurasi yang dihasilkan pada algoritma *naïve bayes* sebesar 82.50%.

Penelitian yang dilakukan oleh (Utami & Masripah, 2021) mengenai komparasi algoritma klasifikasi sentimen analisis ulasan pembelajaran *online* dan pembelajaran jarak. Menjelaskan bahwa algoritma yang digunakan dalam membantu menyelesaikan penelitian ini yaitu *Naïve Bayes*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM). Hasil penelitian yang dihasilkan dengan menggunakan ketiga algoritma tersebut yaitu, akurasi yang dihasilkan pada algoritma *Naïve Bayes* sebesar 83,33%, menggunakan algoritma *k-Nearest Neighbor* (KNN) sebesar 86,33%, dan sedangkan menggunakan algoritma *Support Vector Machine* (SVM) sebesar 87,67%.

Penelitian yang dilakukan oleh (Kaswidjanti, et al., 2020) mengenai perbandingan akurasi analisis sentimen sosial media menggunakan *Lexicon Based* dan *Support Vector Machine* pada rekomendasi *souvenir*. Pada penelitian ini dijelaskan bahwa algoritma yang digunakan dalam membantu untuk menyelesaikan penelitian ini yaitu *Lexicon*

Based dan *Support Vector Machine* (SVM). Berdasarkan hasil penelitian yang didapatkan, dapat diketahui bahwa akurasi yang dihasilkan menggunakan algoritma *Lexicon Based* sebesar 88% dan sedangkan ketika menggunakan algoritma *Support Vector Machine* (SVM) sebesar 86%.

Berdasarkan penelitian (Qaiser, et al., 2021) mengenai perbandingan teknik mesin pembelajaran untuk analisis sentimen. Pada penelitian ini algoritma yang membantu dalam menyelesaikan penelitian ini yaitu, *Naïve Bayes*, *Decision Tree*, *Support Vector Machine*, dan metode modern seperti *Deep Learning*. Berdasarkan hasil penelitian yang didapatkan, akurasi yang dihasilkan menggunakan metode *Naïve Bayes* sebesar 87.18%, untuk algoritma *Decision Tree* sebesar 68.21%, untuk algoritma *Support Vector Machine* Sebesar 82.05, dan untuk algoritma *Deep Learning* sebesar 93.33%.

Pada penelitian (Sidiq, et al., 2020) mengenai sentimen analisis komentar *toxic* pada *group* facebook *game online* menggunakan klasifikasi *Naïve Bayes*. Berdasarkan hasil penelitian yang didapatkan, nilai akurasi dengan menggunakan metode ini sebesar 75%.

Pada penelitian (Giovani, et al., 2020) mengenai analisis sentimen aplikasi ruang guru di twitter menggunakan algoritma klasifikasi. Berdasarkan hasil yang didapatkan, algoritma yang digunakan pada penelitian ini yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbor* (KNN). Dapat diketahui, bahwa nilai akurasi yang dihasilkan saat menggunakan algoritma SVM yaitu sebesar 76.93%, pada algoritma NB nilai akurasi yang didapatkan sebesar 65.40%, dan saat menggunakan algoritma KNN nilai akurasinya sebesar 70.42%.

Penelitian (Murnawan & Sinaga, 2017) mengenai pemanfaatan analisis sentimen untuk pemeringkatan popularitas tujuan wisata. Berdasarkan hasil yang didapatkan, dapat diketahui bahwa pada penelitian hanya menggunakan algoritma dari *Naïve Bayes* karena metode yang sederhana dan intuitif. Oleh karena itu, dapat diketahui bahwa pada penelitian ini menggunakan 2 tingkat phrase, yaitu pada 100 phrase dengan nilai akurasi yang didapatkan yaitu sebesar 65.65% dan 5000 phrase dengan nilai akurasi yang didapatkan sebesar 82.67%.

Penelitian yang dilakukan oleh (Sari, 2020) yang berjudul analisis sentimen pada review objek wisata dunia fantasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Berdasarkan hasil yang didapatkan dapat diketahui nilai akurasi yang dihasilkan sebesar 77.01%.

Pada penelitian (Muttaqin & Kharisnudin, 2021) yang berjudul analisis sentimen aplikasi Gojek menggunakan *Support Vector Machine* dan *K-Nearest Neighbor*. Berdasarkan hasil yang didapatkan, dapat diketahui bahwa nilai akurasi yang dihasilkan dengan menggunakan metode SVM yaitu sebesar 87.98%, sedangkan dengan menggunakan algoritma KNN dihasilkan nilai akurasi sebesar 82.14%.

Pada penelitian (Afdhal, et al., 2022) Penerapan Algoritma *Random Forest* Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia dengan menggunakan algoritma *Random Forest*. Berdasarkan hasil yang didapatkan, diketahui akurasi yang dihasilkan sebesar 79%.

Pada penelitian yang dilakukan oleh (Adrian, et al., 2021) dengan judul Perbandingan Metode Klasifikasi *Random Forest* dan SVM Pada Analisis Sentimen PSBB didapatkan hasil pengujian yang dilakukan dengan menggunakan algoritma *Random Forest* didapatkan hasil akurasi yang didapatkan sebesar 58%. Sedangkan dilakukan dengan pengujian algoritma *Support Vector Machine* (SVM) didapatkan hasil akurasi sebesar 56%.

Penelitian yang ditulis (Naufal & Kusuma, 2022) dengan judul Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis *Deep Learning* didapatkan hasil akurasi yang dihasilkan 87%.

Pada penelitian yang dilakukan oleh (Pangaribuan, et al., 2021) yang berjudul Analisis Sentimen Aplikasi *E-Learning* Selama Pandemi COVID-19 Dengan Menggunakan Metode *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN). Dari hasil pengolahan data yang dilakukan dengan proporsi perbandingan 80:20 diketahui akurasi yang dihasilkan dari algoritma SVM didapatkan sebesar 83%, sedangkan pengolahan dengan proporsi perbandingan 80:20 menggunakan algoritma CNN sebesar 86%.

Penelitian yang berjudul Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi oleh (Himawan & Eliyanti, 2021) dengan menggunakan metode *Random Forest*, *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC), didapatkan hasil dengan menggunakan algoritma *Random Forest* akurasi yang dihasilkan sebesar 75.81%, dengan menggunakan algoritma NBC akurasi yang dihasilkan sebesar 75.22%, dan untuk algoritma SVM akurasi yang dihasilkan sebesar 77.58%.

Tabel 2. 1 Kelebihan Algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC)

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
Hidayat & Minati, 2019	Analisis Komparasi Dalam Klasifikasi <i>Text Mining</i> Untuk <i>Translate</i> Qur'an Inggris Dan Indonesia	Komparasi Algoritma <i>Support Vector Machine</i> (SVM), <i>Naïve Bayes</i> , <i>K-Nearest Neighbor</i> (KNN), dan <i>J48 Decision Tree Classifier</i>	Penelitian ini mengkomparasi algoritma klasifikasi untuk melakukan <i>translate</i> Qur'an Inggris dan Indonesia. Berdasarkan hasil pembahasan yang telah dilakukan tersebut, dapat dilihat bahwa hasil akurasi tertinggi dihasilkan oleh algoritma SVM.	Pendidikan
Winanto & Budihartini, 2022	Perbandingan Akurasi Analisis Sentimen Pada <i>Twitter</i> Pemerintahan Provinsi DKI Jakarta Selama waktu vaksinasi COVID-19	Akurasi <i>Naïve Bayes</i> dan <i>k-Nearest Neighbor</i> (KNN)	Berdasarkan penelitian yang telah dilakukan, dapat dilihat bahwa nilai akurasi tertinggi atau performansi terbaik terdapat pada algoritma NB.	Kesehatan
Utami & Masripah, 2021	Komparasi Klasifikasi Analisis Sentimen Ulasan	Algoritma <i>Naïve Bayes</i> , <i>K-Nearest Neighbor</i> (k-NN), dan <i>Support Vector Machine</i> (SVM)	Mengkomparasi algoritma pada ulasan pembelajaran online dan pembelajaran jarak dapat dilihat	Pendidikan

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
	Pembelajaran <i>Online</i> Dan Pembelajaran Jarak		bahwa performansi terbaik didapatkan pada algoritma SVM.	
Kaswidjanti, et al., 2020	Perbandingan Akurasi Analisis Sentimen Sosial Media Menggunakan <i>Lexicon Based</i> Dan <i>Support Vector Machine</i> Pada Rekomendasi <i>Souvenir</i>	<i>Lexicon Based</i> dan <i>Support Vector Machine</i> (SVM)	Perbandingan analisis sentimen pada sosial media untuk merekomendasikan <i>souvenir</i> , dapat diketahui bahwa performansi terbaik terdapat pada algoritma <i>Lexicon based</i> dengan nilai akurasi sebesar 88%.	Bisnis
Qaiser, et al., 2021	Perbandingan Teknik Mesin Pembelajaran Untuk Analisis Sentimen	<i>Naïve Bayes</i> , <i>Decision Tree</i> , <i>Support Vector Machine</i> , dan metode modern dari <i>Deep Learning</i>	Penelitian yang bertujuan untuk membandingkan teknik pembelajaran untuk analisis sentimen, dapat diketahui bahwa algoritma dengan performansi terbaik yaitu metode <i>modern</i> dari <i>deep learning</i> dengan nilai akurasi sebesar 93.33%.	Pendidikan

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
Sidiq, et al., 2020	Sentimen Analisis Komentar <i>Toxic</i> Pada <i>Group Facebook Game</i> <i>Online</i> Menggunakan Klasifikasi <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Analisis sentimen pada komentar <i>toxic group facebook game online</i> dapat diketahui bahwa nilai akurasi yang didapatkan sebesar 75%.	Pendidikan
Giovani, et al., 2020	Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi	<i>Naïve Bayes</i> , <i>K-Nearest Neighbor</i> (KNN), dan <i>Support Vector Machine</i> (SVM)	Penelitian yang telah dilakukan untuk mengklasifikasi aplikasi ruang guru di twitter menghasilkan performansi terbaik pada algoritma SVM dengan nilai akurasi yang didapatkan sebesar 76.93%.	Pendidikan
Murnawan & Sinaga, 2017	Pemanfaatan Sentimen Untuk Peningkatan Popularitas Tujuan Wisata	Analisis <i>Naïve Bayes</i>	Pemanfaatan sentimen yang dilakukan untuk pemeringkatan popularitas tujuan wisata dapat diketahui bahwa penelitian yang dilakukan semakin banyak phrase yang digunakan, maka semakin baik pula performansi yang didapatkan.	Pariwisata

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
Sari, 2020	Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma <i>K-Nearest Neighbor</i> (KNN)	<i>K-Nearest Neighbor</i> (KNN)	Penelitian yang dilakukan pada <i>review</i> objek wisata dunia fantasi didapatkan performansi dengan nilai akurasi sebesar 77.01%.	Pariwisata
Muttaqin & Kharisnudin, 2021	Analisis Sentimen Aplikasi Gojek Menggunakan <i>Support Vector Machine</i> Dan <i>K-Nearest Neighbor</i>	<i>Support Vector Machine</i> (SVM) dan <i>K-Nearest Neighbor</i> (KNN)	Berdasarkan perbandingan yang dilakukan pada metode SVM dan KNN, dapat diketahui bahwa algoritma SVM memiliki performa terbaik dengan nilai akurasi sebesar 87.98%.	Bisnis
Afdhal, et al., 2022	Penerapan Algoritma <i>Random Forest</i> Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia dengan menggunakan	<i>Random Forest</i>	Berdasarkan penelitian yang dilakukan pada analisis sentiment di You Tube terkait Islamfobia didapatkan performansi dengan nilai akurasi sebesar 79%.	Agama

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
	algoritma <i>Forest</i>	<i>Random</i>		
Adrian, et al., 2021	Perbandingan Metode Klasifikasi <i>Random Forest</i> dan SVM Pada Analisis Sentimen PSBB	<i>Random Forest</i> dan <i>Random Support Vector Machine (SVM)</i>	Berdasarkan dari hasil perbandingan pada penelitian analisis sentiment PSBB didapatkan hasil pengujian pada algoritma <i>Random Forest</i> 58% dan pada algoritma SVM sebesar 56%	Kesehatan
Naufal & Kusuma, 2022	Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis <i>Deep Learning</i>	<i>Deep Learning</i>	Penelitian sentimen di Twitter yang dilakukan terhadap kebijakan pemberlakuan pembatasan kegiatan masyarakat didapatkan nilai akurasinya sebesar 87%.	Sosial

Penulis	Judul	Metode	Hasil Penelitian	Mapping Topik
Pangaribuan , et al., 2021	Analisis Sentimen Aplikasi <i>E-Learning</i> Selama Pandemi COVID-19 Menggunakan Metode <i>Support Vector Machine</i> (SVM) dan <i>Convolutional Neural</i> <i>Network</i> (CNN)	Sentimen <i>Support Vector Machine</i> (SVM) dan <i>Convolutional Neural</i> <i>Network</i> (CNN)	Dari penelitian analisis sentimen yang dilakukan pada aplikasi <i>E-Learning</i> selama pandemi. Didapatkan nilai akurasi pada algoritma SVM sebesar 83% dan akurasi yang dihasilkan pada algoritma CNN sebesar 86%.	Pendidikan
Himawan & Eliyanti, 2021	Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi	Akurasi <i>Random Forest</i> , <i>Support</i> <i>Vector Machine</i> (SVM) dan <i>Naïve Bayes</i> <i>Classifier</i> (NBC)	Penelitian yang dilakukan untuk mengetahui sentimen pengguna twitter terhadap pemerintah provinsi DKI Jakarta di masa pandemi. Didapatkan nilai akurasi pada algoritma <i>Random Forest</i> sebesar 75.81%, pada algoritma NBC sebesar 75.22%, dan untuk algoritma SVM akurasi yang dihasilkan sebesar 77.58%	Politik

2.2 Kajian Deduktif

2.2.1 Machine Learning

Machine learning merupakan salah satu alat yang sangat menjanjikan untuk melakukan klasifikasi (Hossain, 2019). *Machine learning* sendiri memiliki tujuan untuk mencari fungsi yang belum diketahui atau terstruktur antara variabel *input* dan *output* (Kwekha-Rashid, Abduljabbar and Alhayani, 2021). Dalam membangun model matematika *machine learning* atau pembelajaran mesin menggunakan sampel yang sering disebut sebagai *data training* untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk menjalankan tugasnya (Bishop, 2006).

Machine learning berkaitan dengan pertanyaan terkait bagaimana membangun suatu program komputer agar dapat meningkat secara otomatis yang didasarkan dari pengalaman (Mitchell, 1997). Pembelajaran mesin dibagi menjadi beberapa kategori, yaitu sebagai berikut:

a. *Supervised Learning*

Teknik *supervised learning* merupakan teknik yang sepenuhnya diberikan label. Hal ini didasarkan dalam mengumpulkan sampel data memiliki suatu label. *Supervised learning* dikelompokkan lebih lanjut dalam masalah klasifikasi dan regresi. Dimana, masalah klasifikasi adalah ketika variabel *output* berbentuk kategori. Sedangkan pada regresi ketika variabel output berupa nilai riil (Brownlee, 2016).

b. *Unsupervised Learning*

Teknik *unsupervised learning* merupakan suatu teknik yang sepenuhnya tidak memiliki label. Hal ini dikarenakan tidak ada kebutuhan untuk pemberian label dalam kumpulan data dan hasilnya tidak mengidentifikasi contoh di kelas yang telah ditentukan (Thupae, et al., 2018). *Unsupervised learning* dikelompokkan lebih lanjut dalam permasalahan *clustering* dan asosiasi. Dimana, masalah pengelompokkan (*clustering*) adalah tempat untuk mengelompokkan data berdasarkan kemiripannya. Sedangkan, permasalahan asosiasi adalah aturan yang menggambarkan sebagian besar data yang ada seperti pelanggan yang membeli produk A juga berkemungkinan membeli produk B (Brownlee, 2016).

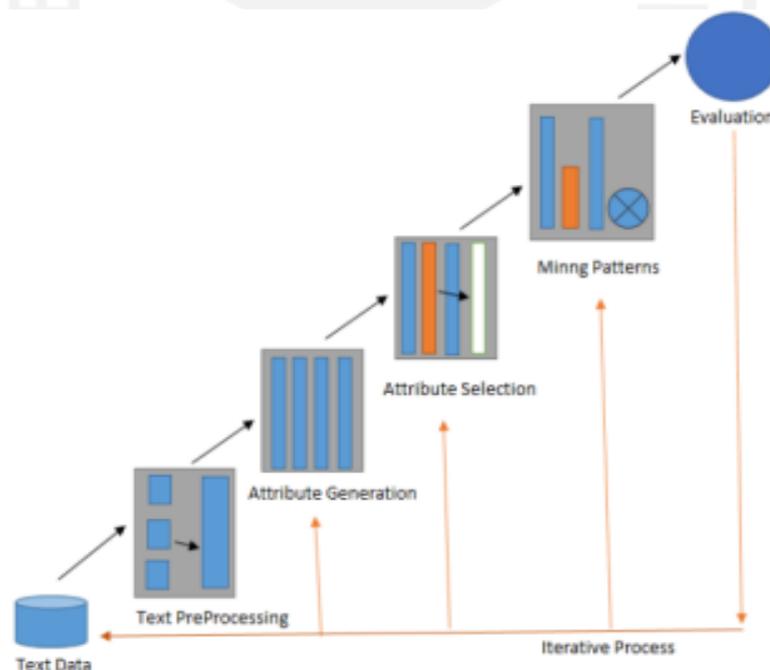
c. *Reinforcement Learning*

Teknik ini terletak pada *supervised learning* dan *unsupervised learning*. Teknik pembelajaran ini dilakukan dengan cara coba-coba dan akan terus berinteraksi dengan

lingkungan yang dinamis (Qiang & Zhongli, 2011). Metode ini berdasarkan pada model proses pengambilan keputusan Markov (*Markov Decision Process*) yaitu mencakup 2 jenis metode. Seperti algoritma SARSA, dimana akan mempelajari pengalaman model, kemudian mendapatkan strategi yang optimal. Dan algoritma *Temporal Difference* dan algoritma *Q-Learning* yang secara langsung menghitung strategi yang optimal tanpa pengetahuan model (Qiang & Zhongli, 2011).

2.2.2 Text Mining

Text mining merupakan suatu proses menambang data yang berupa teks dengan sumber data biasanya didapatkan dari dokumen untuk bertujuan dalam mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan untuk mendapatkan suatu informasi yang bermanfaat. Tujuan dari *text mining* sendiri adalah mengekstrak informasi yang berguna dari sumber data. Sumber data yang digunakan pada *text mining* adalah sekumpulan dokumen yang memiliki format yang tidak terstruktur melalui identifikasi dan eksplorasi pola yang menarik. Tugas khusus *text mining* antara lain, pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*) (Soepomo, 2014).



Gambar 2. 1 Tahapan *Text Mining*

Sumber: (Kumar and Bhatia, 2013)

Pada *Text Mining* untuk mendapatkan suatu informasi secara efisiensi, terdapat beberapa tahapan didalamnya (Kumar and Bhatia, 2013), yaitu:

a. Pre-Processing

Pada tahapan pre-processing ini, terdapat beberapa tahapan yang perlu dilakukan:

1. Pembersihan Teks

Tahapan ini merupakan tahapan penghapusan teks yang tidak diperlukan atau yang tidak diinginkan seperti penghapusan iklan yang terdapat pada tampilan awal web, melakukan normalisasi dengan mengkonversi teks dari format *binary* menjadi tabel, gambar, dan formula.

2. Tokenisasi

Tahap ini merupakan tahap pemecahan teks dari dokumen menjadi pecahan-pecahan. Dimana pada tokenisasi sama dengan kata, karakter, ataupun sub kata. Tokenisasi tidak hanya dapat digunakan pada Bahasa Indonesia, namun dapat menggunakan Bahasa dari negara lain. Tokenisasi sendiri memiliki banyak arti.

3. *Part of Speech Tagging* (POS)

Tahapan ini melakukan penandaan pada setiap token dengan input yang diberikan berupa teks. Pemberian tanda memiliki fungsi pada kata-kata yang ambigu atau yang tidak memiliki arti.

b. Transformasi Teks (*Text Transformation*)

Dokumen data yang akan di diwakilkan berupa kata (fitur) dengan berisi kejadian. Terdapat 2 pendekatan utama dalam representasi dokumen, yaitu kantong kata dan *vector space*.

c. Pemilihan Fitur (*Feature Selection*)

Tahap ini juga mengetahui penyaringan variabel, dimana proses penyeleksian *subset* fitur yang penting untuk digunakan dalam kreasi model. Asumsi yang terpenting ketika menggunakan fitur seleksi adalah banyaknya pengulangan dan fitur yang tidak relevan.

d. Data Mining

Pada poin ini proses *text mining* digabung dengan proses *data mining*. Teknik klasik pada data mining menggunakan struktur database yang hasilnya dari tahapan sebelumnya.

e. Evaluasi

Tahap ini akan dilakukan evaluasi, setelah dilakukan evaluasi hasil yang didapatkan bisa dibuang atau hasil yang didapatkan bisa digunakan pada *input* rangkaian berikutnya.

2.2.3 Analisis Sentimen

Analisis sentimen atau penggalian opini merupakan salah satu tugas utama dalam *Natural Language Processing* (NLP) dimana mempelajari pendapat seseorang terhadap entitas tertentu. Analisis sentimen sendiri berasal dari perspektif pengguna produk (Fang and Zhan, 2015). Selain berdasarkan pendapat seseorang, analisis sentimen dapat digunakan pada perilaku atau emosi seseorang dapat dijadikan sebagai bahan evaluasi (Isnain, Marga and Alita, 2021). Tugas dasar saat melakukan analisis sentimen adalah mengklasifikasikan polaritas teks yang didapatkan kedalam suatu dokumen, kalimat, atau pendapat dengan memiliki aspek positif dan negatif (Wahyudi and Kusumawardana, 2021).

Analisis Sentimen juga berguna pada saat melakukan klasifikasi produk, jasa, target konsumen di pasar, hingga memonitoring sosial media untuk meningkatkan branding dari suatu *merk* produk (Arviana, 2021). Metode ini dapat bertujuan untuk mengetahui keunggulan dan kelemahan dari produk yang kita punya berdasarkan data yang diperoleh melalui internet. Hal tersebut bertujuan sebagai umpan balik dari pengguna untuk dijadikan bahan evaluasi untuk mengembangkan produk.

2.2.4 Pembobotan (*Term Weighting*)

Pembobotan kata merupakan metode yang standar dalam melakukan ekstraksi. Metode ini mencoba untuk menemukan aspek-aspek eksplisit dari ulasan dalam domain tertentu. Pendekatan pada metode ini sangat relevan karena ketika pengguna memberikan ulasannya, mereka akan menggunakan kata-kata yang memiliki kemiripan arti (Hu & Liu, 2004). Pembobotan merupakan metode untuk mengubah input dari suatu data menjadi suatu fitur vektor. Metode pembobotan yang umum digunakan adalah *bag-of-feature*. Sebagai contoh terdapat sederetan fitur seperti pada vektor $\{f_1, f_2, \dots, f_n\}$ dimana yaitu sekumpulan fitur-fitur sebanyak n yang sudah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vektor dari data tersebut adalah vektor. Terdapat beberapa metode yang dapat digunakan pada saat melakukan pembobotan, yaitu:

a. *Term Presence* (TP)

Term presence (TP) merupakan suatu metode pembobotan yang ditandai dengan angka 1 atau 0. Dimana metode ini melihat keberadaannya dari daftar kata-kata (term) atau fitur yang terdapat pada corpus pada suatu dokumen. Nilai 1 diberikan pada *feature vector* jika fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang dilakukan pembobotan

dan hal tersebut tidak menghiraukan jumlah kemuculan. Nilai 0 diberikan pada *feature space* jika fitur tersebut tidak terdapat pada dokumen (O'Keefe, 2009).

b. *Term Frequency* (TF)

Term frequency (TF) merupakan suatu metode yang digunakan hanya untuk menggambarkan hitungan kemunculan dari suatu kata atau fitur dalam suatu dokumen. Metode ini juga memiliki kesamaan dengan metode TP yang telah dijelaskan sebelumnya, dimana yang membedakan metode ini dengan metode sebelumnya yaitu jika metode ini menghitung kemunculan sarti suatu kata atau fitur acuan pada suatu dokumen dan tidak hanya keberadaan dari fitur tersebut (O'Keefe, 2009).

c. *Term Frequency - Inverse Document Frequency* (TF-IDF)

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan suatu algoritma pembobotan yang tersusun dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Output* dengan fitur yang sering muncul pada suatu dokumen akan menghasilkan TF-IDF yang tinggi, sedangkan fitur yang sering muncul pada suatu dokumen akan menghasilkan nilai yang rendah. Dengan metode ini, peneliti dapat mengetahui fitur-fitur yang penting akan memiliki nilai tinggi dan fitur yang tidak penting akan memiliki nilai rendah (O'Keefe, 2009).

2.2.5 *Web Scraping*

Web scraping merupakan suatu teknik yang digunakan untuk mengumpulkan informasi dari website secara otomatis tanpa harus menyalin secara manual. Tujuan dari teknik yaitu untuk mencari informasi yang diperlukan lalu dikumpulkan kedalam web yang baru dengan cara pengambilan dan ekstraksi (A. Yani, Pratiwi and Muhandi, 2019). *Web scraping* juga bisa disebut suatu alat bantu dalam *drop shipper* untuk mendapatkan data produk dalam skala besar (Maulana, Susanto and Kusumaningrum, 2019). Pengindeksan yang dilakukan teknik ini yaitu dengan cara menelusuri dokumen-dokumen HTML dari suatu *website* yang akan diambil informasinya untuk di-tag ke HTML agar bisa mendapatkan informasi yang diambil untuk ditirukan pada aplikasi *web scraping* yang akan dibuat (F. Maria Rosario B, 2017). *Web scraping* sendiri memiliki beberapa Langkah (A. Yani, Pratiwi and Muhandi, 2019), yaitu:

a. *Create Scraping Template*

Pembuat program mempelajari dokumen HTML dari website yang akan diambil informasinya untuk tag HTML yang mengapit informasi yang akan diambil.

b. Explore Site Navigation

Pembuat program mempelajari teknik navigasi pada website yang akan diambil informasinya untuk ditirukan pada aplikasi web scraper yang akan dibuat.

c. Automate Navigation and Extraction

Berdasarkan informasi yang didapat pada langkah 1 dan 2 diatas, aplikasi web scraper dibuat untuk megotomatisasi pengambilan informasi dari website yang ditentukan.

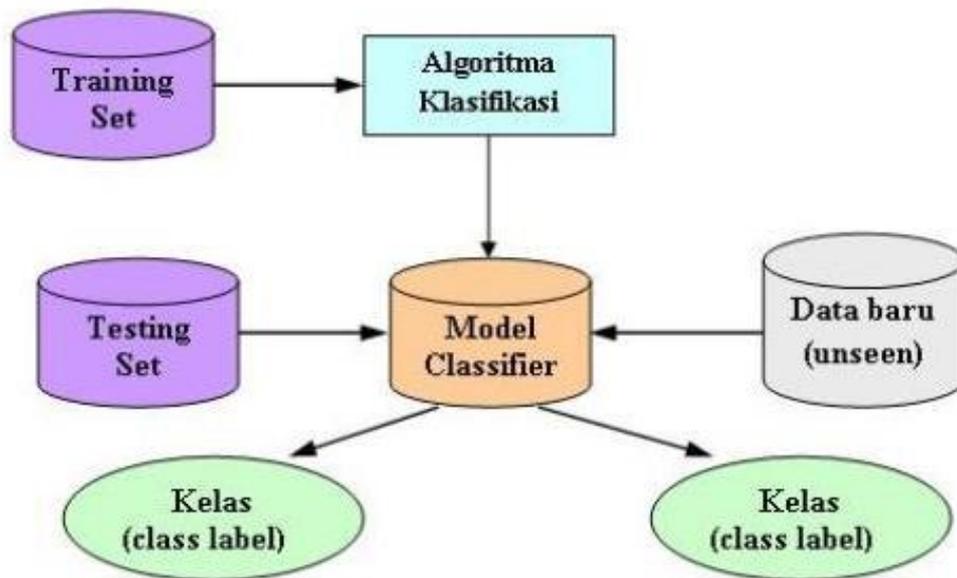
d. Extracted Data and Package History

Informasi yang didapat dari langkah 3 disimpan dalam tabel database.

2.2.6 Klasifikasi

Klasifikasi merupakan teknik multivariat untuk memisahkan data yang berbeda dari suatu objek (observasi) dan mengalokasikan objek (observasi) baru tersebut ke dalam kelompok yang telah ditetapkan sebelumnya. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan yang kecil (Johnson. R, 2007). Analisis klasifikasi merupakan suatu teknik yang digunakan dalam menghasilkan suatu model prediksi yang mendeskripsikan suatu label kelas.

Menurut (Han & Kamber, 2006) *classification* memiliki dua proses, yaitu tahap pertama dengan membangun suatu model berdasarkan dari serangkaian *data class* yang sering disebut *learned model*. Model yang dibangun dilakukan dengan menganalisa *record database*, dimana setiap *record* akan diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang sering disebut atribut *class label*. Oleh karena itu dengan adanya atribut label kelas ini, maka tahapan klasifikasi ini juga disebut sebagai *supervised learning*. Tahap ini juga disebut sebagai tahap pembelajaran, dimana algoritma klasifikasi ini membangun sebuah model klasifikasi dengan cara menganalisis *data training*. Tahap pembelajaran dapat dipandang sebagai tahap pembentukan fungsi atau tahap pemetaan $y = f(x)$ dengan y merupakan kelas hasil prediksi dan x adalah *record class* yang ingin diprediksi.



Gambar 2. 2 Tahapan Analisis Klasifikasi

Sumber: <https://philips.wordpress.com/2006/08/10/training-set-testing-set-dan-validation-set/>

Kualitas dari hasil analisis klasifikasi yang didapatkan dapat dinilai dan dievaluasi berdasarkan:

1. *Accuracy*

Akurasi adalah jumlah proporsi prediksi yang benar. Dimana, akurasi sendiri biasa digunakan sebagai tingkat ketepatan antara nilai aktual dengan nilai prediksi (Manning, et al., 2009).

2. *Precision*

Presisi adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks yang terpilih oleh system. Presisi digunakan untuk melihat tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem (Manning, et al., 2009).

3. *Recall*

Recall adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks relevan yang ada pada koleksi. *Recall* biasanya digunakan untuk melihat ukuran keberhasilan sistem dalam menemukan Kembali informasi (Manning, et al., 2009).

4. *F-Measure*

F-Measure merupakan kata-kata harmonis yang berasal dari nilai *recall* dan nilai presisi, sehingga dapat memberikan penilaian kinerja yang lebih seimbang. F-measure biasa digunakan untuk mengukur kinerja suatu sistem secara menyeluruh dalam pengklasifikasian (Manning, et al., 2009).

2.2.7 Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier (NBC) merupakan salah satu metode pengklasifikasian statistik. Dimana, pengklasifikasian ini dapat memprediksi probabilitas keanggotaan kelas suatu data tuple yang akan masuk ke dalam kelas-kelas tertentu sesuai dengan perhitungan probabilitasnya (Handayani & Pribadi, 2015). Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang mesin pembelajaran karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan melakukan perhitungan yang sederhana (Aggarwal & C, 2012). Metode NBC ini didasarkan pada teorema *Bayes*, dimana teorema ini dikombinasikan dengan “*Naïve*” yang dapat diartikan setiap atribut atau variabel bersifat *independen*. Metode NBC ini dilatih dengan efisien dalam pembelajaran terawasi, dimana proses NBC ini sendiri mengasumsikan ada tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada tidaknya fitur lain di kelas yang sama (Setiawan, 2014).

Pada dasarnya teorema *Bayes* merupakan aturan dari *Naïve Bayes Classifier*. Dimana, NBC merupakan suatu perhitungan yang sederhana karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas (Arifin & Sasongo, 2018). Berikut merupakan persamaan dari teorema *Bayes* (Feldman & J, 2007).

$$P(X) = \frac{P(X|H)P(H)}{P(X)} \dots\dots\dots \text{pers (1)}$$

Dimana:

X = Data tuple hasil pengujian dari suatu set data yang telah ditentukan yang masuk ke kelas tertentu

H = Hipotesis data X yang akan masuk ke dalam kelas yang lebih spesifik

P(H|X) = Probabilitas X merupakan data tuple yang akan masuk kedalam kelas spesifik

P(H) = Probabilitas prior atau sebelumnya

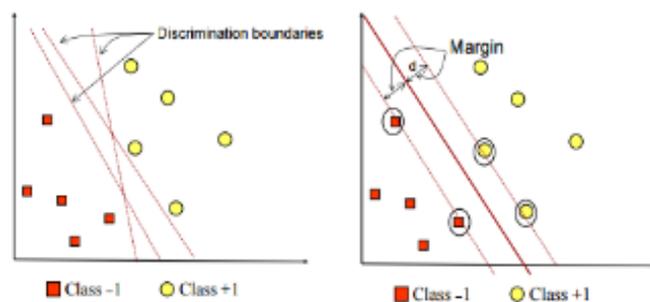
P(X|H) = Probabilitas posterior dengan X dikondisikan pada H

P(X) = Probabilitas sebelum dari X

2.2.8 Support Vector Machine (SVM)

Support Vector Machine (SVM) dikenalkan pertama kali oleh Vapnik tahun 1992 sebagai salah satu metode *learning machine* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*. Metode ini menggunakan hipotesis berupa fungsi – fungsi linier dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan (Siagian, 2011). Hal ini berfungsi untuk memisahkan dua kelas data yang berbeda, yaitu positif (+1) dan negatif (-1). Pada gambar 2.4 untuk data positif (+1) disimbolkan dengan warna kuning dan data negatif (-1) disimbolkan dengan warna merah.

Problem yang terjadi pada klasifikasi dapat dijelaskan dengan usaha menemukan garis *hyperplane* yang memisahkan antara kedua kelompok tersebut. Garis solid pada Gambar dibawah menunjukkan *hyperplane* yang terbaik, yaitu yang terdapat pada tengah – tengah kedua *class*, sedangkan titik merah dan kuning yang berada didalam lingkaran hitam adalah *support vector*.



Gambar 2. 3 *Hyperplane* pada *Support Vector Machine*

Sumber: (Parapat, et al., 2018)

2.2.9 Asosiasi Kata

Asosiasi kata biasanya digunakan untuk mengetahui kata apa saja yang muncul pada suatu dokumen. Asosiasi kata dapat digunakan untuk mengetahui keterkaitan dan hubungan antar kata, seperti antar dua kata atau lebih digunakan secara bersamaan dalam suatu dokumen. Asosiasi kata dapat juga dilihat dari nilai korelasi antar kata, dimana nilai korelasi berkisar antara -1 sampai 1. Jika nilai mendekati 1 atau -1 maka hubungan antar kata tersebut semakin kuat, sedangkan jika nilai mendekati 0 maka hubungan antar kata

semakin lemah. Ada beberapa kategori nilai korelasi yang digunakan sebagai berikut (Farach & Nugraha, 2019).

0: Tidak ada korelasi antar dua variabel

$>0-0,25$: Korelasi lemah

$>0,25 - 0,5$: Korelasi cukup

$>0,5 - 0,75$: Korelasi kuat

1: Korelasi sangat kuat



BAB III METODOLOGI PENELITIAN

3.1 Objek Penelitian

Pada penelitian yang akan dilakukan, objek penelitian yang akan ditunjukkan berasal dari persepsi pengguna pelayanan aplikasi PeduliLindungi melalui *review* pada *database website Google Play*.

3.2 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan pada tugas akhir ini yaitu dengan cara metode *Web Scraping*. Alat yang digunakan pada dalam metode ini yaitu AppFollow yang merupakan website untuk menggali data dari *Google Chrome*.

3.3 Variabel Penelitian

Pada penelitian ini terdapat dua variabel yang digunakan dalam membantu menganalisis sentiment yang didapatkan, yaitu:

a. Tanggal (*Date*)

merupakan tanggal dibuatnya suatu ulasan. Variabel ini akan digunakan dalam penelitian ini untuk membantu dalam melakukan analisis.

b. Ulasan (*Review*)

merupakan ulasan pengguna terkait penggunaan aplikasi. Variabel ini akan digunakan dalam penelitian untuk membantu melakukan analisis dan memberikan perbaikan terkait hasil analisis sentimen yang didapatkan.

3.4 Jenis dan Sumber Data

Data-data yang akan peneliti gunakan untuk penelitian ini merupakan gabungan dari data primer dan data sekunder. Data primer merupakan hasil penggalian data yang diambil secara langsung melalui *website*. Sedangkan data sekunder adalah data yang relevan yang diambil melalui berbagai sumber, seperti buku referensi, internet, dan buku atau informasi dari instansi terkait.

3.3.1. Data Primer

Data primer disini diperoleh berdasarkan dengan penggalian data dengan menggunakan metode *scrapping* dari situs *website* PeduliLindungi menggunakan

bantuan AppFollow pada *Google Chrome* pada alamat website “[PeduliLindungi - Aplikasi di Google Play](#)” dengan ulasan yang akan diambil sebanyak 125.523 data ulasan. Pada penelitian ini, digunakan sampel dari ulasan pengguna aplikasi PeduliLindungi sejak adanya update dengan versi 4.3 hingga 4.4.3.1 pada tanggal 30 Maret 2022 sampai 14 Mei 2022 berdasarkan rating karena memiliki pelayanan dari aplikasi kurang memenuhi standar. Data sampel yang digunakan pada penelitian ini sebanyak 5.077 ulasan.

3.3.2. Data Sekunder

Data sekunder disini diperoleh melalui dari jurnal-jurnal, buku-buku, serta informasi-informasi yang terkait dengan permasalahan yang sedang diangkat dan yang sedang berhubungan dengan masalah penelitian.

3.5 Instrumen Penelitian

Instrument merupakan alat untuk mengumpulkan, mengolah, menganalisis, ataupun menyajikan data-data secara sistematis dan objektif. Pada penelitian kali ini alat yang digunakan dalam membantu dalam melancarkan ekstraksi ulasan untuk diolah yang didapatkan menggunakan python dan Rstudio dan untuk memvisualisasikan menggunakan *wordcloud*. Algoritma yang akan digunakan dalam membantu pengolahan data sendiri menggunakan *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).

3.6 Metode Analisis Data

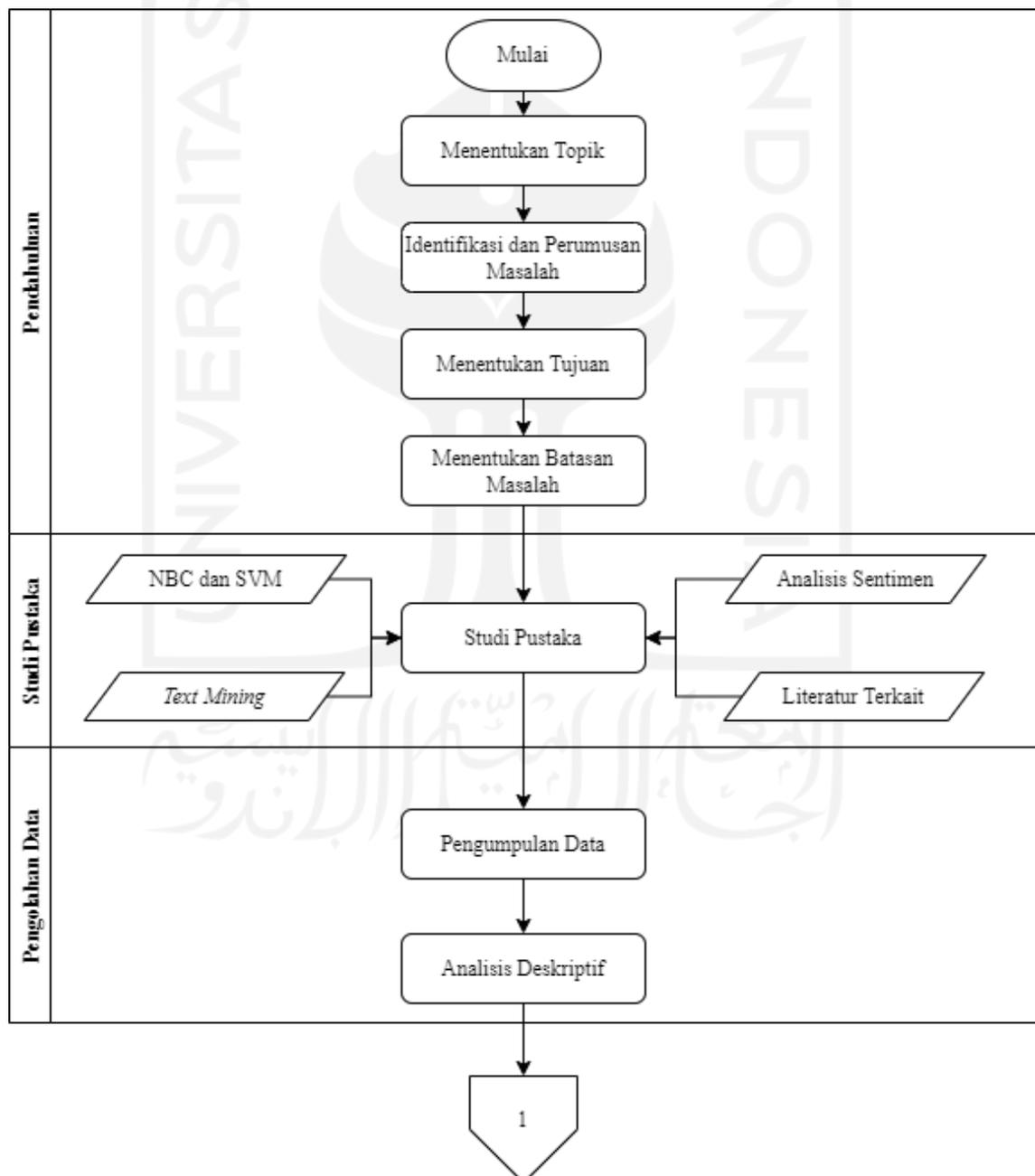
Pada penelitian yang akan dilakukan, digunakan bantuan software Python, Rstudio dan Microsoft Excel. Metode analisis yang akan digunakan pada penelitian ini, yaitu:

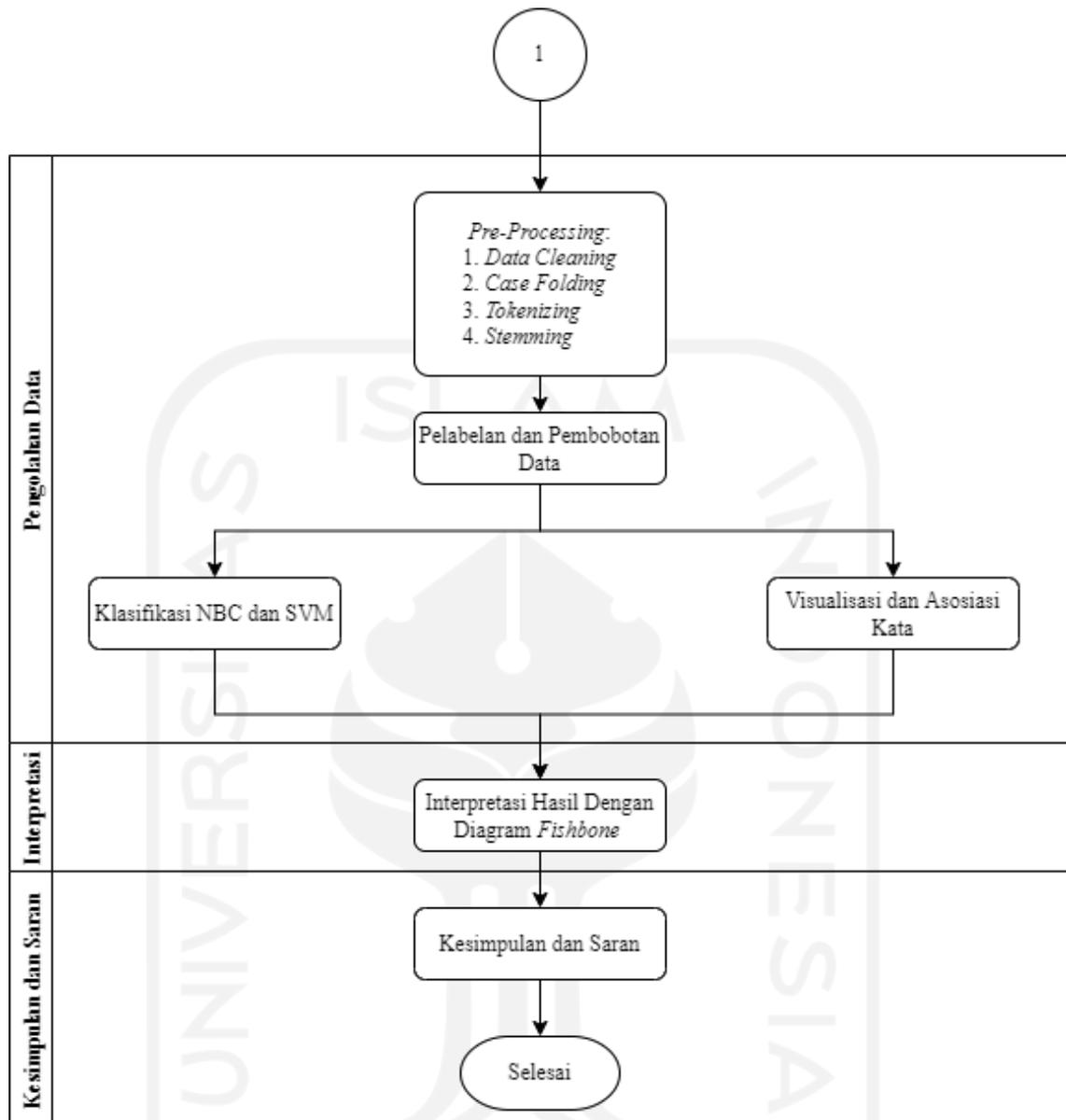
- a. Analisis Deskriptif, metode ini digunakan saat memberikan gambaran dari ulasan yang ada pada aplikasi PeduliLindungi pada situs *Google Play*.
- b. Analisis Sentimen, metode ini dilakukan untuk menganalisis apakah ulasan yang didapatkan termasuk kedalam label positif atau negatif.
- c. Metode *Machine Learning*, pada metode menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) untuk dilakukan pengklasifikasian yang dimasukkan kedalam label positif atau negatif.
- d. *Wordcloud*, metode analisis ini digunakan dalam memvisualisasikan ulasan yang sering muncul di aplikasi PeduliLindungi.

- e. *Diagram Fishbone*, metode analisis data ini digunakan untuk mengetahui faktor-faktor yang menjadi permasalahan dari penggunaan aplikasi PeduliLindungi yang didasarkan pada ulasan negatif.

3.7 Prosedur Eksperimen

Prosedur eksperimen merupakan tahapan penelitian yang menunjukkan tahapan-tahapan yang dilakukan dalam melakukan penelitian dari awal hingga akhir. Berikut merupakan diagram alir atau *flowchart* adalah sebagai berikut.





Gambar 3. 1 Diagram Alur Penelitian

Alur penelitian diatas tahapan yang dilakukan dalam proses penelitian, untuk mendetailkan setiap tahapannya berikut adalah keterangan pada setiap tahapan pada diagram alir diatas:

1. Mulai
2. Menentukan Topik

Tahapan penelitian yang dilakukan pertama kali yaitu dengan mencari topik, dimana pada penelitian kali ini topik yang diangkat yaitu perbandingan analisis sentimen pada layanan jasa yang dikembangkan oleh pemerintahan yaitu KOMINFO yaitu layanan PeduliLindungi.

3. Identifikasi dan Perumusan Masalah

Pada tahapan penelitian ini melakukan identifikasi masalah dengan melihat kondisi dan situasi pada saat ini dengan pengamatan yang dilakukan tidak secara langsung. Hasil identifikasi yang telah dilakukan, didapatkan bahwa mengingat aplikasi PeduliLindungi merupakan suatu alat yang menjadi ketergantungan bagi masyarakat selama masa COVID-19 ini yang digunakan untuk membantu mengetahui keberadaan pengguna, maka perlu untuk mengetahui analisis sentimen pengguna dari aplikasi tersebut. Setelah itu, dilakukan perumusan masalah yang akan menjadi acuan pada penelitian kali ini. Dimana rumusan masalah yang terbentuk yaitu bagaimana gambaran umum mengenai penggunaan layanan dari aplikasi PeduliLindungi pada website *Google Play*, bagaimana performa algoritma NBC dan SVM dalam mengklasifikasikan data ulasan pengguna layanan dari aplikasi PeduliLindungi menjadi kelas positif dan negatif, apa informasi yang didapatkan dalam setiap klasifikasi yang telah dilakukan, dan faktor-faktor apa saja yang perlu diperbaiki berdasarkan hasil ulasan negatif yang didapatkan?

4. Menentukan Tujuan

Pada tahap ini dilakukannya penentuan tujuan penelitian. Dimana, disini akan dijelaskan hal-hal yang ingin dicapai dalam penelitian ini. Tujuan dari penelitian ini sendiri yaitu untuk mengetahui gambaran umum mengenai penggunaan layanan dari aplikasi PeduliLindungi pada website *Google Play*, mengetahui performa dari algoritma NBC dan SVM dalam mengklasifikasikan data ulasan pengguna layanan dari aplikasi PeduliLindungi menjadi kelas positif dan negatif, mendapatkan informasi yang didapatkan dalam setiap klasifikasi yang telah dilakukan, serta mengetahui faktor-faktor yang perlu diperbaiki berdasarkan hasil ulasan negatif yang didapatkan.

5. Menentukan Batasan Masalah

Dalam penelitian ini pun terdapat batasan masalah yang harus dipenuhi agar penelitian tetap terarah dan pembahasan tidak melebar kepada hal-hal yang tidak diperlukan. Batasan-batasan masalah dalam penelitian ini yaitu data yang digunakan merupakan data ulasan mengenai aplikasi dari PeduliLindungi pada situs *Google Play* pada tanggal 30 Maret 2022 sampai 14 Mei 2022 dan ulasan yang diambil merupakan ulasan yang menggunakan Bahasa Indonesia.

6. Studi Pustaka

Pada tahap ini dilakukan studi pustaka yang berisi dan berkaitan dengan topik penelitian ini agar dapat mengetahui analisis sentimen dengan menggunakan metode yang sesuai.

Topik-topik yang digunakan dalam studi pustaka yaitu, *Text mining*, *Wordcloud*, Analisis Sentimen, *Machine Learning*, dan lain-lain.

7. Pengumpulan Data

Tahap pengumpulan data ini sendiri berbeda dengan penelitian lainnya karena bukan datang secara langsung ke lapangan atau ke perusahaan. Data yang akan digunakan dan diolah pada penelitian ini yaitu berupa ulasan pada aplikasi PeduliLindungi pada *website Google Play* yang dilakukan dengan mengekstraksi data ulasan secara online menggunakan teknik *scrapping data* yang menggunakan *tools* dari *google chrome*, yaitu AppFollow.

8. Analisis Deskriptif

Tahap ini pada penelitian ini digunakan untuk menggambarkan ulasan secara umum yang didasarkan pada ulasan pengguna PeduliLindungi seperti waktu *review* pengguna.

9. Preprocessing Data

Preprocessing merupakan suatu proses yang memiliki tujuan untuk menyeleksi data-data yang tidak diperlukan agar dapat digunakan dengan baik dan data yang didapatkan akan menjadi lebih terstruktur, selain itu informasi yang didapatkan akan menjadi jelas. Pada preprocessing ini terdapat beberapa proses, yaitu:

- a. *Data Cleaning*: tahap ini merupakan tahap penghapusan teks yang tidak diperlukan seperti simbol, emotikon, tanda baca, dan lain hal sebagainya.
- b. *Case Folding*: tahap ini memiliki tujuan mengubah teks menjadi lowercase, menghapus tanda baca pada teks, dan menghapus angka pada teks ulasan.
- c. *Tokenizing*: tahap ini merupakan tahap pembersihan data dan ekstraksi daftar kata.
- d. *Stemming*: tahap ini merupakan tahap *filtering* atau penyaringan kata yang terdapat pada suatu dokumen.

10. Pelabelan dan Pembobotan Data

a. Klasifikasi NBC dan SVM

Data-data yang telah didapatkan dan diberikan pelabelan, selanjutnya akan dilakukan proses klasifikasi. Pada proses klasifikasi ini, metode *machine learning* yang digunakan dalam membantu proses klasifikasi saat melakukan pengolahan data ulasan ini yaitu *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) yang akan berbentuk kelas positif dan negatif.

b. Visualisasi dan Asosiasi Kata

Visualisasi dan asosiasi kata merupakan proses ekstraksi seluruh informasi yang telah dilakukan pengolahan data, dimana informasi yang didapatkan berupa kata-kata yang sering muncul pada ulasan dari penggunaan aplikasi. Bentuk dari visualisasi dan asosiasi kata akan berbentuk *wordcloud* dan barplot.

11. Interpretasi Hasil Menggunakan Diagram *Fishbone*

Setelah melakukan visualisasi kata dan dapatkan informasi mengenai penggunaan dari aplikasi PeduliLindungi yang menjadikan pengguna memberikan ulasan negatif. Informasi yang didapatkan selanjutnya akan dilakukan analisis faktor-faktor yang menyebabkan permasalahan berdasarkan dari pendapat pengguna. *Tool* yang digunakan dalam membantu menganalisis faktor-faktor tersebut dengan menggunakan diagram *fishbone*.

12. Kesimpulan dan Saran

Tahap terakhir dari penelitian ini yaitu dengan menarik kesimpulan dan saran berdasarkan pembahasan dan analisis yang telah dilakukan.

13. Selesai

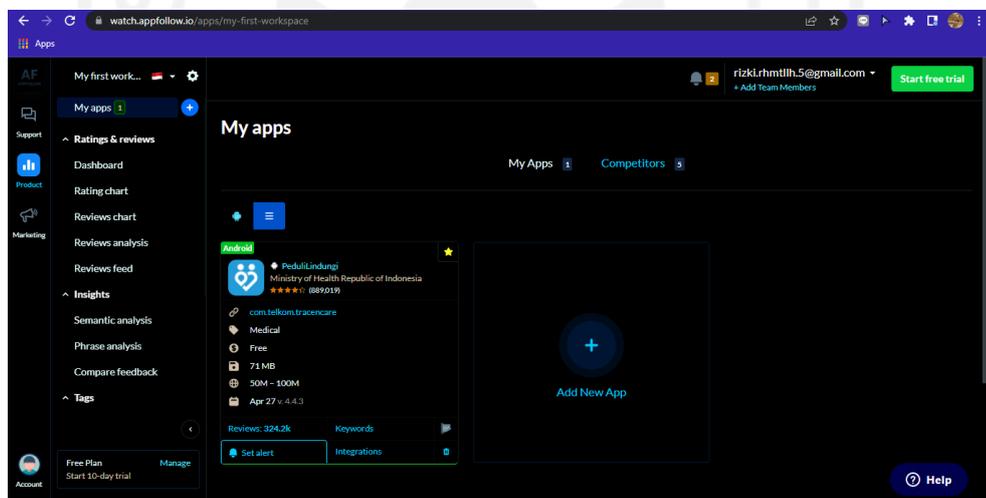
3.1 Framework of the Research

Penelitian yang akan dilakukan didasarkan pada hasil ekstraksi atau penggalan data dari ulasan yang didapatkan pada aplikasi PeduliLindungi menggunakan metode *scrapping*. Dimana, pada penelitian ini dalam mengolah data yang didapatkan dari ulasan-ulasan yang didapatkan melalui *website Google Play* yaitu *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Dalam penelitian ini juga, untuk memvisualisasi data yang didapatkan setelah diolah menggunakan *wordcloud* berdasarkan kata ulasan yang sering muncul.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

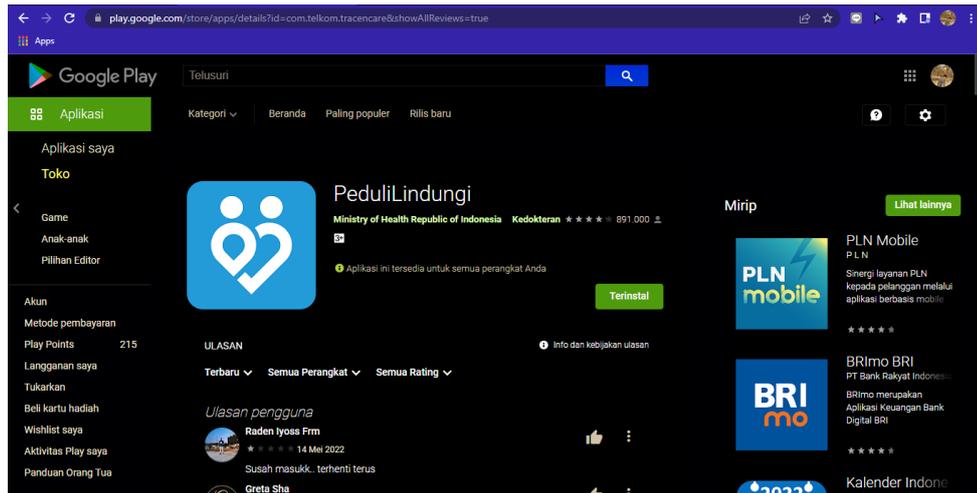
4.1. Pengumpulan Data

Pengambilan data yang dilakukan pada penelitian ini berupa ulasan dari pengguna aplikasi PeduliLindungi pada situs *Google Play* dengan alamat domain: <https://play.google.com/store/apps/details?id=com.telkom.tracencare&showAllReviews=true>. Pengambilan data yang dilakukan dengan menggunakan teknik *scraping*. Banyak tools yang membantu dalam melakukan teknik *scraping*, salah satunya yaitu appfollow. Appfollow merupakan suatu tools yang membantu dalam mengekstraksi data pada situs secara gratis. Pada penelitian ini digunakan tools pada appfollow.io yang dengan alamat domain yang dapat diakses pada: <https://appfollow.io/> yang digunakan dalam mengambil data ulasan yang selanjutnya akan diekspor menjadi file *xlsx* atau *CSV*.



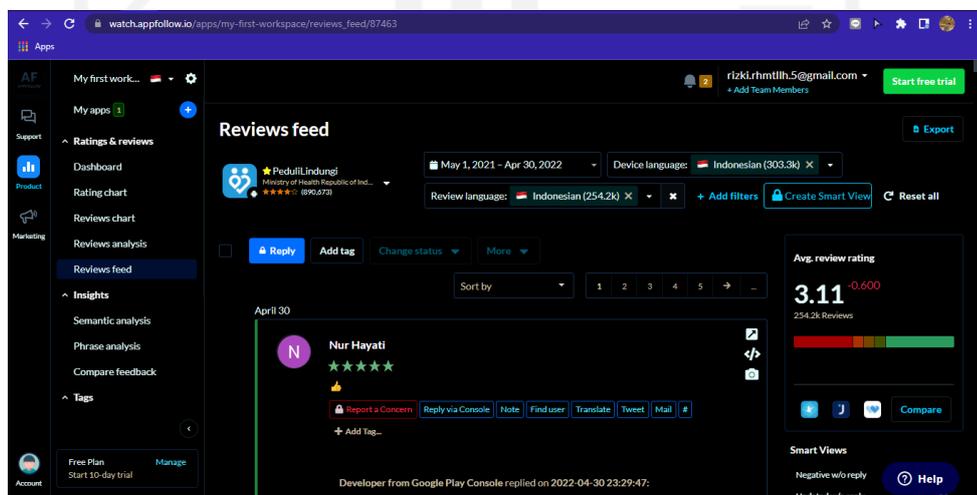
Gambar 4. 1 *Interface* awal pada situs Appfollow.io

Pertama-tama tahapan yang perlu dilakukan dalam melakukan *scraping* data adalah dengan memasuki situs appfollow. Kemudian setelah itu melakukan pencarian aplikasi yang akan dijadikan objek untuk melakukan *scraping* data, dimana untuk memudahkan hal tersebut dapat langsung memasukkan alamat domain <https://play.google.com/store/apps/details?id=com.telkom.tracencare&showAllReviews=true>.



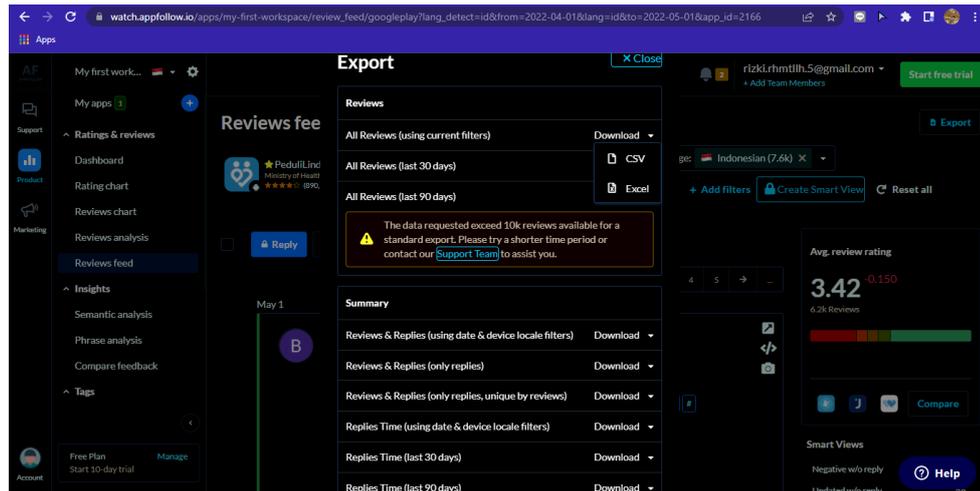
Gambar 4. 2 Website ulasan PeduliLindungi

Setelah menambahkan aplikasi yang akan dijadikan objek penelitian kedalam appfollow, setelah melakukan tahapan tersebut memilih data sesuai dengan batasan yang telah ditentukan. Pertama pada appfollow klik aplikasi PeduliLindungi yang telah ditambahkan, selanjutnya pilih *review feed* pada bagian kiri tampilan *website*. Setelah itu melakukan filterisasi pada tanggal pengambilan data dan Bahasa ulasan yang digunakan. Dalam penelitian ini data yang akan diambil dan digunakan yaitu waktu (tanggal) dan ulasan yang diberikan oleh *user*. Dapat dilihat pada gambar dibawah ini untuk pengambilan ulasan dengan tampilan seperti Gambar 4.3.



Gambar 4. 3 Pengambilan Ulasan

Setelah melakukan filterisasi pada data yang akan diambil, selanjutnya melakukan *export* data atau menyimpan data dalam bentuk *xlsx* atau *csv*.



Gambar 4. 4 Export Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Tanggal	Review															
2	30 March 2022	Fitur klaim sertifikat vaksin masih belum bisa digunakan, saya sudah vaksin ke tiga(boosters) menggunakan vaksin sinopharm, untuk sertifikat diberikan hardcopy dari lembaran															
3	30 March 2022	Masa sya sdah dvaksinasi booster ke tiga pda tanggal 16-02-2022,sertifikat vaksinasi ke 3 belum muncul jga smpai saat ini.															
4	30 March 2022	Update aplikasi untuk memperbaiki bug saat Scan QR yang gagal.															
5	30 March 2022	Aplikasi di update tuh malah makin susah klaim sertifikat vaksin															
6	30 March 2022	Aplikasi yang bagus untuk digunakan ðŸŒŸðŸŒŸðŸŒŸ															
7	30 March 2022	Terimakasih atas aplikasinya															
8	30 March 2022	Gimana nih booster sertifikat kenapa gak bisa diklaim.															
9	30 March 2022	Saya sudah vaksin ke3., tapi sertifikat vaksin tdk ada., sya coba claim masukin smua data tpi gagal padahal data sudah benar...gimana tuh?															
10	30 March 2022	Halo admin, Ini aplikasi nya sudah saya update tapi kenapa malah tidak bisa dibuka sama sekali ya? Semua perizinan aplikasi aman, sinyal aman, ruang penyimpanan aman															
11	30 March 2022	Saya sudah booster tapi tidak ada sertifikat vaksin k 3 nya kenapa ya? Padahal no yang di pakai sama nik sama															
12	30 March 2022	Udah gak ada lagi covid ini, hapus aja dari PlayStore ini aplikasi															
13	30 March 2022	Saya sudah vaksin yang hanya satu kali Johnson & Johnson tgl 29 desember 2021 sertifikat sudah keluar. Tapi baru baru ini Saya cek di akun peduli lindungi malah kembali															
14	30 March 2022	bagussss															
15	30 March 2022	Oke															
16	30 March 2022	Sangat membantu															
17	30 March 2022	Aplikasi karya anak bangsa yang sangat bermanfaat. Maju terus, kami dukung dengan support dan do'a															
18	30 March 2022	Bagus															
19	30 March 2022	Apl Ini gak hisa dibuat 2 akun.gmn..															
20	30 March 2022	Alhamdulillah sehat selalu															

Gambar 4. 5 Hasil Scraping Data

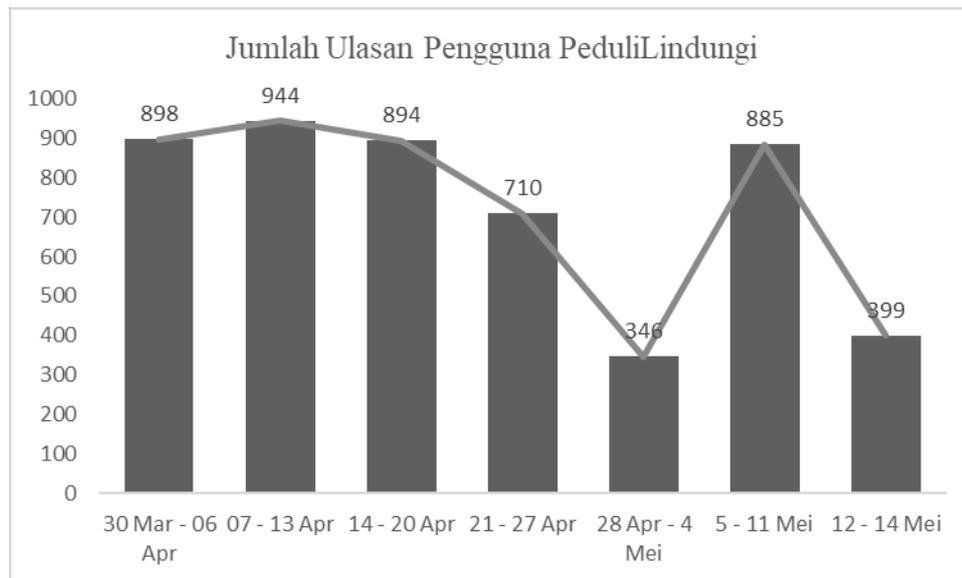
4.2. Pengolahan Data

Pada penelitian ini, data yang diekstraksi dari *website Google Play*, selanjutnya akan diolah untuk mendapatkan suatu informasi yang dapat digunakan dalam memperbaiki aplikasi yang digunakan. Data sampel yang digunakan untuk pengolahan data yang akan dilakukan pada ulasan pengguna aplikasi PeduliLindungi yaitu sebanyak 5.077 ulasan. dimana, dalam pengolahan data penelitian ini terdapat dua pengolahan data yang dilakukan yang terdiri dari analisis deskriptif dan *pre-processing data*, yaitu sebagai berikut:

4.2.1 Analisis Deskriptif

Tahapan pengolahan analisis deskriptif dilakukan dengan tujuan untuk mengetahui gambaran secara umum dari data yang telah didapatkan terkait pelayanan dari aplikasi PeduliLindungi melalui teknik *scraping* yang dilakukan melalui situs *Google Play*. Pada

penelitian ini, aspek yang dilihat yaitu berdasarkan waktu dan jumlah ulasan yang diberikan pengguna.



Gambar 4. 6 Jumlah Ulasan Pengguna PeduliLindungi

Berdasarkan pada gambar diatas, dapat diketahui bahwa pengambilan data yang diambil dari tanggal 30 Maret 2022 hingga dengan 14 Mei 2022 mengalami fluktuasi. Pengguna yang memberikan ulasan terbanyak terdapat pada awal April yaitu tepatnya tanggal 07 – 13 April 2022 dengan jumlah sebesar 944 ulasan. Namun, terjadi penurunan yang drastis dalam memberikan ulasan yang terjadi pada akhir april menuju awal mei dengan ulasan data yang didapatkan sebanyak 346 data ulasan. Hal ini terjadi diperkirakan karena tidak terjadinya perubahan pada tampilan dan kinerja serta aplikasi karena belum dilakukannya update aplikasi ke versi terbaru yang memungkinkan terjadinya penurunan kinerja aplikasi sehingga mengakibatkan pada penurunannya pelayanan. Oleh karena itu sedikit pengguna yang menuliskan keluhannya pada kolom komentar aplikasi PeduliLindungi.

4.2.2 Pre-processing Data

Pre-processing merupakan salah satu tahapan yang sangat penting dalam melakukan penambangan data. *Pre-processing* bertujuan untuk melakukan pembersihan data yang tidak diperlukan agar dapat meningkatkan kinerja. Pada proses ini dilakukan dengan melewati beberapa tahapan dan dibantu dengan menggunakan *tools* Python. Tahapan *pre-processing* yang dilakukan pada penelitian ini antara lain sebagai berikut:

4.2.2.1. Data Cleaning

Tahapan *data cleaning* merupakan proses penghapusan data yang berduplikasi atau kata yang memiliki salinan kalimat yang sama dalam satu data yang telah didapatkan. Data mentah yang didapatkan setelah dilakukan *scraping* data sebanyak 5.077 ulasan, namun setelah dilakukan *data cleaning* didapatkan data bersih sebanyak 3.683 ulasan yang akan digunakan.

4.2.2.2. Case Folding

Pada tahapan ini, terdapat beberapa proses yang dilakukan, seperti mengubah kata menjadi *lowercase*, menghapus angka, dan menghapus tanda baca. Dimana pada tahapan ini sendiri merupakan tahapan normalisasi yang dilakukan untuk mengurangi kosakata dan memungkinkan untuk megeneralisasi lebih baik.

a. Mengubah text menjadi *lowercase* (huruf kecil)

Proses ini merupakan proses perubahan pada kata yang memiliki huruf kapital menjadi huruf kecil. Hal ini dikarenakan huruf kapital dan huruf kecil memiliki arti yang berbeda ketika dideteksi. Adapun contoh dari kalimat ulasan yang dapat dilihat pada tabel 4.1.

Tabel 4. 1 Pengubahan Kata menjadi huruf kecil

<i>Input</i>	<i>Output</i>
Udah tidak ada lagi covid ini, hapus aja dari PlayStore ini aplikasi	udah tidak ada lagi covid ini hapus aja dari playstore ini aplikasi

b. Menghapus angka

Proses ini merupakan proses penghapusan angka yang terdapat pada suatu kalimat. Angka yang dimaksud dapat berupa tanggal, perulangan, dan lain sebagainya. Karena hal ini dilakukan agar dalam melakukan analisis sentiment selanjutnya memudahkan kamus lexicon dalam memproses ulasan yang didapatkan. Adapun contoh dari kalimat ulasan terdapat pada tabel 4.2.

Tabel 4. 2 Menghapus angka pada ulasan

<i>Input</i>	<i>Output</i>
mengubah tahun dan bulan aja masih harus geser jauh ditambah ngeblank lagi. Bayangin kalau lahir tahun 60-70 an berapa kali harus tekan ² ?	Mengubah tahun dan bulan aja masih harus geser jauh ditambah ngeblank lagi. Bayangin kalau lahir tahun an berapa kali harus tekan?

c. Menghapus tanda baca

Tanda baca yang terdapat pada suatu kalimat akan dianggap sebagai pemisah kata (delimiter) dan tidak memiliki pengaruh apapun pada proses *text mining*. Sehingga komponen-komponen tanda baca yang terdapat pada suatu kalimat tersebut perlu dihapus. Berikut merupakan contoh kalimat ulasan pada tabel 4.3.

Tabel 4. 3 Menghapus tanda baca pada ulasan

<i>Input</i>	<i>Output</i>
Tidak bisa berkomentar. SERBA LEMOT, mempersulit, memperhambat. R I B E T - L A M B A T	Tidak bisa berkomentar SERBA LEMOT mempersulit memperhambat R I B E T L A M B A T

4.2.2.3. Tokenizing

Tokenizing merupakan tahapan pemisahan kata dari suatu kalimat yang tidak memiliki hubungan kata. Dimana, kata-kata yang telah dipisah disebut sebagai token. Token-token yang telah didapatkan akan membantu dalam memahami konteks atau mengembangkan model NLP. Proses tokenisasi sendiri membantu dalam menggambarkan arti kata dengan menganalisis urutan kata. *Tokenizing* juga bertujuan untuk mempermudah dalam menghitung frekuensi kemunculan dalam dokumen. Berikut merupakan contoh kalimat ulasan pada proses *tokenizing*.

Tabel 4. 4 Pemisahan kata pada ulasan menjadi token

<i>Input</i>	<i>Output</i>
update aplikasi untuk memperbaiki bug saat scan qr yang gagal	['update', 'aplikasi', 'baik', 'bug', 'scan', 'qr', 'gagal']

4.2.2.4. Stemming

Stemming merupakan suatu proses *filtering* yang dilakukan pada suatu data. Dimana pada tahapan ini akan dilakukan perubahan kata yang mengandung imbuhan induktif dan deduktif menjadi kata dasar. Berikut merupakan contoh kalimat pada proses *stemming*.

Tabel 4. 5 Mengubah kata pada ulasan menjadi kata dasar

<i>Input</i>	<i>Output</i>
"Update"	"Update"
"Aplikasi"	"Aplikasi"
"Memperbaiki"	"Baik"
"Dibuka"	"Buka"
"Penyimpanan"	"Simpan"

4.2.3 Representasi Model

Representasi model merupakan suatu situasi yang digunakan untuk menemukan suatu solusi masalah. Pada penelitian ini dilakukan dengan membuat suatu model yang berbentuk token kata agar dapat dihitung dan diolah dengan melakukan analisis sentimen agar mendapatkan skor pembobotan untuk melakukan sentimen.

4.2.3.1. Perhitungan Skor Sentimen dan Pelabelan Kelas Sentimen

Tahap ini merupakan tahap penting dalam melakukan analisis sentimen. Proses pada analisis sentimen yaitu dengan mengekstraksi dan mengolah data menggunakan algoritma yang dilakukan secara otomatis untuk mendapatkan informasi yang dibutuhkan terkait ulasan pengguna.

1. Perhitungan Skor Sentimen

Perhitungan skor sentimen dilakukan untuk melihat ulasan yang memiliki nilai negatif dan positif yang diwakilkan dengan data yang memiliki nilai positif dan negatif. Berikut merupakan contoh dari perhitungan skor sentimen yang dilakukan dengan menggunakan bantuan *software Rstudio* yang ditunjukkan pada tabel dibawah ini.

Tabel 4. 6 Pembobotan Menggunakan Metode Kamus Lexicon

No	Ulasan	Score
1	"['terima', 'kasih', 'aplikasi']"	1
2	"['vaksinasi', 'boster', 'tanggal', 'sertifikat', 'vaksinasi', 'muncul', 'smpai']"	0
3	"['update', 'aplikasi', 'baik', 'bug', 'scan', 'qr', 'gagal']"	-2
4	"['aplikasi', 'update', 'tuh', 'susah', 'klaim', 'sertifikat', 'vaksin']"	-2
5	"['aplikasi', 'bagus']"	0

Berdasarkan pada tabel 4.6 diatas, dapat dilihat bahwa dalam hasil skor sentimen yang didapatkan bervariasi. Pada ulasan yang pertama skor sentimen yang didapatkan bernilai positif, maka berdasarkan hal ini dapat dilihat bahwa kata yang dominan pada kalimat tersebut berisi kata yang mengandung makna positif atau membangun. Pada kalimat ketiga dapat dilihat bahwa skor sentimen yang didapatkan negatif, maka hal ini dapat dilihat bahwa kalimat tersebut berisikan kata-kata kritik negatif. Dalam menentukan pembobotan pada ulasan yang didapatkan juga dapat dilakukan secara manual seperti tabel 4.7 dibawah.

Tabel 4. 7 Perhitungan Pembobotan

Ulasan	Kata Negatif	Kata Positif
['aplikasi', 'karya', 'anak', 'bangsa', 'manfaat', 'maju', 'dukung', 'support', 'doa']	-	Manfaat, Maju
Jumlah	0	2

Berdasarkan dari tabel diatas diketahui bahwa terdapat 2 kata positif dan 0 kata negatif berdasarkan kamus lexicon. Kata positif yang terdeteksi yaitu “manfaat” dan “maju”. Oleh karena itu perhitungan skor sentimennya sebagai berikut:

$$\text{Skor} = 2 - 0 = 2$$

2. Pelabelan Skor Sentimen

Setelah dilakukan perhitungan skor sentimen pada data ulasan yang telah didapatkan, maka hal selanjutnya yang dilakukan yaitu dengan melakukan pelabelan kelas sentimen. Pelabelan sendiri dilakukan berdasarkan tiga kelas kategori yang berbeda yaitu kelas kategori sentimen positif, kelas kategori sentimen netral, dan kelas kategori sentimen negatif. Kriteria pelabelan dapat dilihat berdasarkan ketentuan (AnkThon, 2021):

Sentimen positif: skor > 0

Sentimen netral: skor = 0

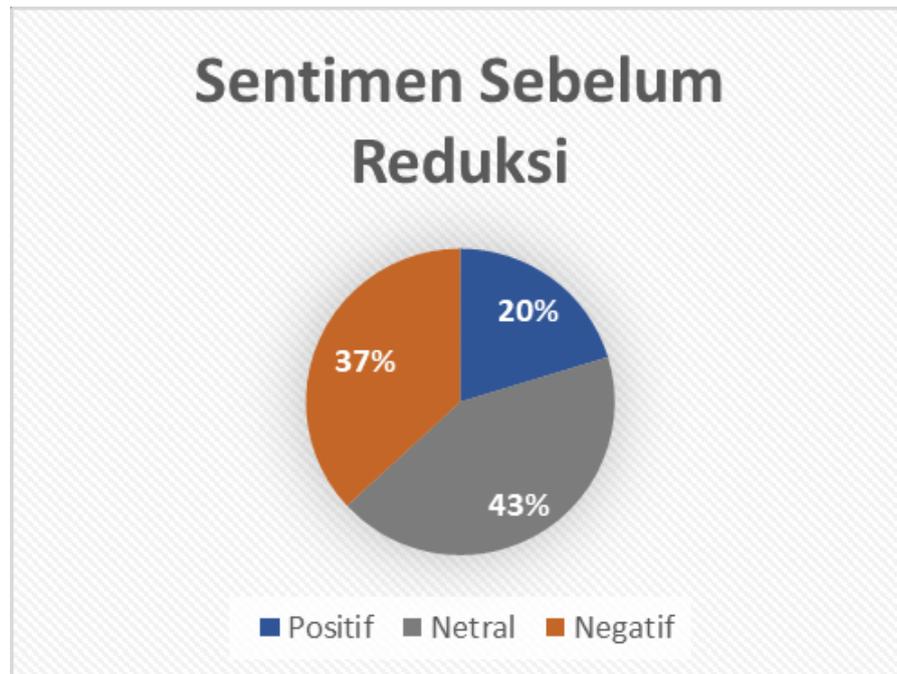
Sentimen negatif: skor < 0

Proses pelabelan pada data ulasan ini masih menggunakan bantuan *software Python*. Penerapan pelabelan pada analisis sentimen ini ditunjukkan pada tabel berikut.

Tabel 4. 8 Pelabelan Data Ulasan

No	Ulasan	Score	Sentiment
1	"['terima', 'kasih', 'aplikasi']"	1	Positif
2	"['vaksinasi', 'boster', 'tanggal', 'sertifikat', 'vaksinasi', 'muncul', 'smpai']"	0	Netral
3	"['update', 'aplikasi', 'baik', 'bug', 'scan', 'qr', 'gagal']"	-2	Negatif
4	"['aplikasi', 'update', 'tuh', 'susah', 'klaim', 'sertifikat', 'vaksin']"	-2	Negatif
5	"['aplikasi', 'bagus']"	0	Netral

Setelah dilakukan *pre-processing* data dan analisis sentimen pada data ulasan pengguna aplikasi PeduliLindungi yang telah dilakukan, didapatkan bahwa pembagian kelas sentimen untuk masing-masing kelas yaitu kelas sentimen positif sebanyak 747 data ulasan, kelas sentimen netral sebanyak 1.578 data ulasan, dan kelas sentimen negatif sebanyak 1.358 data ulasan. Adapun untuk hasil analisis pelabelan ditunjukkan pada grafik *pie chart* yang berada pada gambar 4.7.

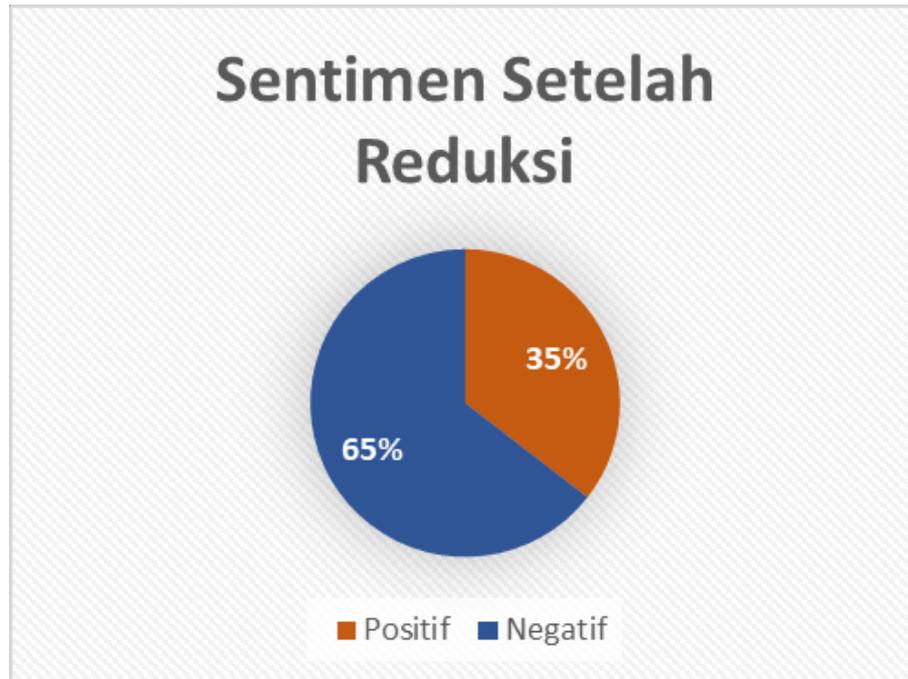


Gambar 4. 7 Visualisasi Hasil Pelabelan Sentimen sebelum reduksi

Setelah didapatkan sentimen pada setiap data, dapat diketahui bahwa pada penelitian ini hanya menggunakan 2 kelas sentimen, yaitu data yang berlabel positif dan data yang berlabel negatif. Hal ini disebabkan karena pada kelas sentimen netral tidak memiliki argument kuat atau tidak memberi masukan yang bermanfaat terhadap perusahaan. Kelas sentimen netral dapat muncul karena terdapat beberapa kemungkinan (Gumilang, 2018), sebagai berikut:

1. Tidak terdapat kata sentimen yang sesuai dengan kamus.
2. Jumlah skor negatif dan skor positif dalam satu kalimat sama.

Berdasarkan hal tersebut, maka dilakukanlah reduksi kelas dengan menghilangkan label sentimen netral, dimana berdasarkan hal tersebut maka awal data yang digunakan 3.683 menjadi 2.105 data ulasan. Dari 2.105 ulasan didapatkan bahwa 747 ulasan merupakan kelas sentimen positif dan 1.358 ulasan merupakan kelas sentimen negatif. Berikut merupakan grafik hasil analisis sentimen yang telah di reduksi.



Gambar 4. 8 Visualisasi Hasil Pelabelan Sentimen sebelum reduksi

4.2.4 Analisis Klasifikasi

Setelah dilakukannya proses sentimen atau pembobotan dan pelabelan kelas, selanjutnya akan dilakukan proses analisis klasifikasi dengan membagi data menjadi *data training* dan *data testing*. Pada penelitian ini digunakan 2 algoritma dalam membantu proses analisis klasifikasi ini, yaitu algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC).

4.2.4.1. Pembagian Data Training dan Data Testing

Data training digunakan dalam membuat suatu model klasifikasi. Dimana, model yang didapatkan akan di representasi menjadi *knowledge* yang akan digunakan dalam melakukan prediksi kelas data baru. Sedangkan *data testing* digunakan untuk mengukur performansi model yang dibangun. Proporsi dalam pembagian *data training* dan *data testing* bersifat subjektif, hal ini mengartikan bahwa pembagian proporsi data tergantung dari peneliti (Aisyah, et al., 2021). Pada penelitian ini akan digunakan tiga nilai perbandingan dalam dalam pembagian *data training* dan *data testing*, yaitu 90:10, 80:20, dan 70:30. Berikut merupakan perbandingan jumlah dari setiap data.

1. Perbandingan *Data Training* sebesar 90% dan *Data Testing* sebesar 10%

Tabel 4. 9 Perbandingan *Data Training* dan *Data Testing* 1

	Persentase	Jumlah	
<i>Data Training</i>	90	1.894,5	1.894
<i>Data Testing</i>	10	210,5	211
Total	100	2105	

Berdasarkan pada tabel 4.9 dapat dilihat bahwa perbandingan pada *data training* dan *data testing* sebesar 90%:10% dari total 2.105 data ulasan, didapatkan bahwa 1.894 data ulasan akan digunakan untuk *data training* dan 211 data ulasan akan digunakan untuk *data testing*.

2. Perbandingan *Data Training* sebesar 80% dan *Data Testing* sebesar 20%

Tabel 4. 10 Perbandingan *Data Training* dan *Data Testing* 2

	Persentase	Jumlah	
<i>Data Training</i>	80	1684	1684
<i>Data Testing</i>	20	421	421
Total	100	2105	

Berdasarkan pada tabel 4.10 dapat dilihat bahwa perbandingan pada *data training* dan *data testing* sebesar 80%:20% dari total 2.105 data ulasan, didapatkan bahwa 1.684 data ulasan akan digunakan untuk data training dan 421 data ulasan akan digunakan untuk *data testing*.

3. Perbandingan *Data Training* sebesar 70% dan *Data Testing* sebesar 30%

Tabel 4. 11 Perbandingan *Data Training* dan *Data Testing* 3

	Persentase	Jumlah	
<i>Data Training</i>	70	1473,5	1473
<i>Data Testing</i>	30	631,5	632
Total	100	2105	

Berdasarkan pada tabel 4.11 dapat dilihat bahwa perbandingan pada *data training* dan *data testing* sebesar 70%:30% dari total 2.105 data ulasan, didapatkan bahwa 1.473 data

ulasan akan digunakan untuk *data training* dan 632 data ulasan akan digunakan untuk *data testing*.

4.2.4.2. Analisis Klasifikasi Pada Algoritma *Naïve Bayes Classifier*

Setelah dilakukan pembagian *data training* dan *data testing*, selanjutnya akan dilakukan pembuatan model prediksi dengan analisis klasifikasi dengan menggunakan bantuan algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* merupakan salah satu algoritma klasifikasi yang terdapat pada machine learning dan memiliki sifat *supervised learning*. analisis klasifikasi ini dilakukan agar mendapatkan nilai akurasi yang akan digunakan untuk memprediksi analisis sentimen. Adapun pengolahan data klasifikasi sentimen adalah sebagai berikut.

1. Pengolahan data dengan perbandingan 90% *data training* dan 10% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 90:10. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.894 dan untuk *data testing* dimulai dari data ke 1.895 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.12.

Tabel 4. 12 Hasil Klasifikasi NBC Dengan Perbandingan 90:10

Output	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	82.46%	87.68%	81.52%
Presisi	63.15%	81.25%	79.76%
<i>Recall</i>	84.21%	78.78%	75.28%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa performansi tertinggi terdapat pada percobaan ke 2. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 87.68%, nilai presisi sebesar 81.25%, dan nilai *recall* sebesar 78.78%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* pada percobaan 2.

Tabel 4. 13 Evaluasi Dengan *Confusion Matrix* Pada Percobaan 2

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i>	<i>False Positif (FP)</i>
	52	12
Negatif	<i>False Negatif (FN)</i>	<i>True Negatif (TN)</i>
	14	133

2. Pengolahan data dengan perbandingan 80% *data training* dan 20% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 80:20. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.684 dan untuk *data testing* dimulai dari data ke 1.685 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.14.

Tabel 4. 14 Hasil Klasifikasi NBC Dengan Perbandingan 80:20

Output	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	85.75%	80.76%	85.04%
Presisi	79.24%	72.78%	77.22%
<i>Recall</i>	82.35%	75.16%	86.33%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa performansi tertinggi terdapat pada percobaan ke 1. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 85.75%, nilai presisi sebesar 79.24%, dan nilai *recall* sebesar 82.35%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* pada percobaan 1.

Tabel 4. 15 Evaluasi Dengan *Confusion Matrix* Pada Percobaan 1

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i>	<i>False Positif (FP)</i>
	139	41
Negatif	<i>False Negatif (FN)</i>	<i>True Negatif (TN)</i>
	22	219

3. Pengolahan data dengan perbandingan 70% *data training* dan 30% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 70:30. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.473 dan untuk *data testing* dimulai dari data ke 1.474 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.16.

Tabel 4. 16 Hasil Klasifikasi NBC Dengan Perbandingan 70:30

<i>Output</i>	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	85.28%	81.65%	84.18%
Presisi	75.59%	72.54%	75.00%
<i>Recall</i>	86.09%	78.31%	85.34%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa performansi tertinggi terdapat pada percobaan ke 1. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 85.28%, nilai presisi sebesar 75.59%, dan nilai *recall* sebesar 86.09%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* pada percobaan 1.

Tabel 4. 17 Evaluasi Dengan *Confusion Matrix* Pada Percobaan 1

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i>	<i>False Positif (FP)</i>
	139	41
Negatif	<i>False Negatif (FN)</i>	<i>True Negatif (TN)</i>
	22	219

4.2.4.3. Analisis Klasifikasi Pada Algoritma *Support Vector Machine*

Algoritma SVM memiliki kernel yang sangat mempengaruhi hasil performansi akurasi yang didapatkan. Pada algoritma ini sendiri memiliki empat kernel yaitu *Linear*, *Polynomial*, *Radial Basis Function (RBF)*, dan *Sigmoid* (Muis & Affandes, 2015). Dari setiap kernel akan dicari performansi akurasi tertinggi untuk dilanjutkan dalam melakukan proses *machine learning*. Berikut merupakan nilai akurasi untuk setiap kernel:

Tabel 4. 18 Perbandingan Akurasi Pada SVM

Kernel	Akurasi
<i>Linear</i>	91.94%
<i>Polynomial</i>	72.51%
<i>Radial Basis Function (RBF)</i>	88.62%
<i>Sigmoid</i>	87.67%

Berdasarkan pada tabel 4.18 dapat dilihat bahwa nilai akurasi tertinggi terdapat pada kernel *Linear* dengan nilai akurasi yang didapatkan sebesar 91.94%. oleh karena itu, maka pada tahap pengolahan data klasifikasi dengan pembagian data 90:10, 80:20, dan 70:30 akan menggunakan bantuan dari kernel *Linear*.

1. Pengolahan data dengan perbandingan 90% *data training* dan 10% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 90:10. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.894 dan untuk *data testing* dimulai dari data ke 1.895 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.19.

Tabel 4. 19 Hasil Klasifikasi SVM Dengan Perbandingan 90:10

Output	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	91.94%	91.94%	91.94%
Presisi	81.03%	81.03%	81.03%
<i>Recall</i>	88.67%	88.67%	88.67%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa nilai performansi pada percobaan ke 1, 2, dan 3 memiliki hasil yang sama. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 91.94%, nilai presisi sebesar 81.03%, dan nilai *recall* sebesar 88.67%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* yang didapatkan.

Tabel 4. 20 Evaluasi Dengan *Confusion Matrix*

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i> 47	<i>False Positif (FP)</i> 11
Negatif	<i>False Negatif (FN)</i> 6	<i>True Negatif (TN)</i> 147

2. Pengolahan data dengan perbandingan 80% *data training* dan 20% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 80:20. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.684 dan untuk *data testing* dimulai dari data ke 1.685 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.21.

Tabel 4. 21 Hasil Klasifikasi SVM Dengan Perbandingan 80:20

Output	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	94.29%	94.77%	95.01%
Presisi	87.71%	89.47%	90.35%
<i>Recall</i>	90.90%	91.07%	91.15%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa performansi tertinggi terdapat pada percobaan ke 3. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 95.01%, nilai presisi sebesar 90.35%, dan nilai *recall* sebesar 91.15%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* pada percobaan 3.

Tabel 4. 22 Evaluasi Dengan *Confusion Matrix* Pada Percobaan 3

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i> 103	<i>False Positif (FP)</i> 11
Negatif	<i>False Negatif (FN)</i> 10	<i>True Negatif (TN)</i> 297

3. Pengolahan data dengan perbandingan 70% *data training* dan 30% *data testing*

Dalam melakukan pengolahan data pada percobaan pertama, pembagian data yang digunakan dengan perbandingan 80:20. Pengolahan data dilakukan dengan 3 kali percobaan, dimana untuk pengolahaya sendiri pada *data training* dimulai dari data ke 1 hingga data ke 1.684 dan untuk *data testing* dimulai dari data ke 1.685 hingga ke 2.105. berikut merupakan hasil analisis klasifikasi sentimen yang diperoleh ditunjukkan pada tabel 4.23.

Tabel 4. 23 Hasil Klasifikasi SVM Dengan Perbandingan 70:30

Output	Percobaan 1	Percobaan 2	Percobaan 3
Akurasi	91.94%	91.94%	91.94%
Presisi	81.03%	81.03%	81.03%
<i>Recall</i>	88.67%	88.67%	88.67%

Berdasarkan dari percobaan yang telah dilakukan, didapatkan bahwa nilai performansi pada percobaan ke 1, 2, dan 3 memiliki hasil yang sama. Dimana didapatkan nilai akurasi pada model klasifikasinya sebesar 91.94%, nilai presisi sebesar 81.03%, dan nilai *recall* sebesar 88.67%. Nilai performansi yang didapatkan pada hasil model prediksi klasifikasi didapatkan dengan menggunakan *confusion matrix*. *Confusion matrix* menjelaskan

jumlah data uji yang benar dan jumlah data uji yang salah. Berikut merupakan tabel *confusion matrix* didapatkan.

Tabel 4. 24 Evaluasi Dengan *Confusion Matrix*

Prediksi	Positif	Negatif
Positif	<i>True Positif (TP)</i> 47	<i>False Positif (FP)</i> 11
Negatif	<i>False Negatif (FN)</i> 6	<i>True Negatif (TN)</i> 148

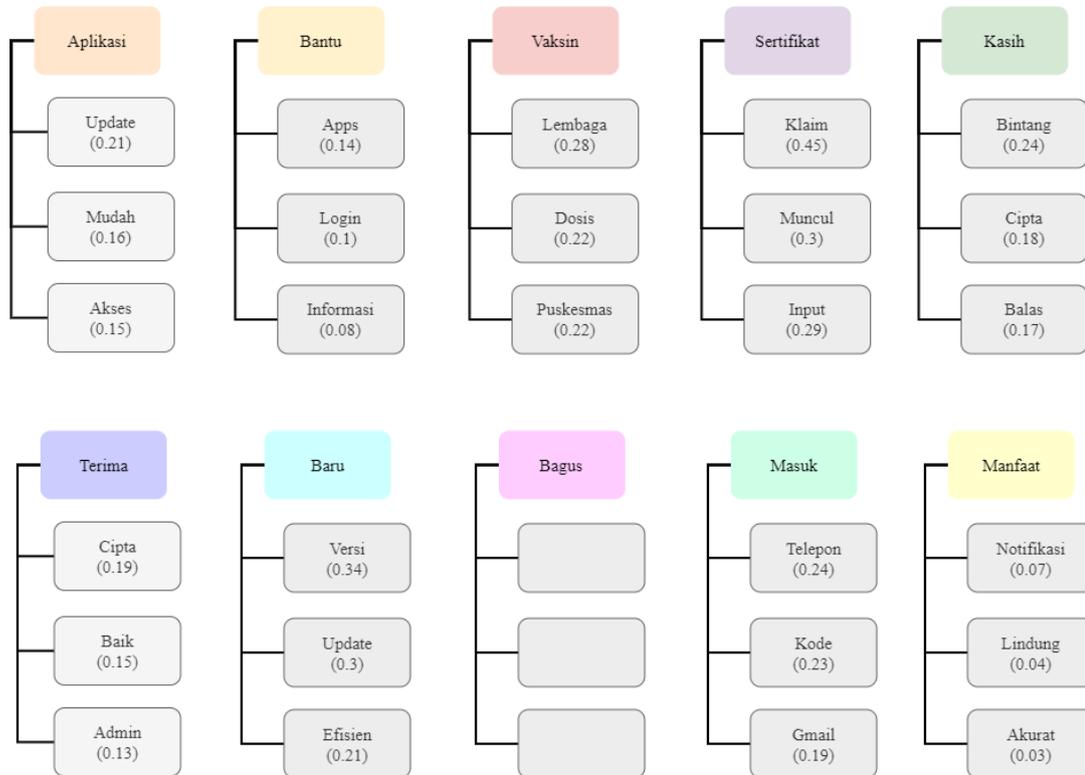
4.2.5 Visualisasi dan Asosiasi

Visualisasi dilakukan untuk setiap klasifikasi dari kelas sentimen. Tujuan dari visualisasi sendiri untuk menampilkan kata-kata ulasan yang sering dibicarakan oleh pengguna. Selain itu juga, dengan mengekstraksi ulasan tersebut, maka dengan dilakukannya visualisasi dapat mengetahui informasi baru dari data ulasan yang dilakukan. Proses ini dilakukan untuk memperkuat informasi yang telah dilakukan dari proses analisis sentimen.

4.2.5.1. Ulasan Positif

Data ulasan positif merupakan data ulasan yang digunakan dari hasil pelabelan dan pembobotan. Nilai pembobotan yang digunakan pada kelas sentimen positif diambil dengan nilai pembobotan lebih dari nol. Untuk mendapatkan informasi terkait ulasan positif yang sering dibicarakan atau diulas, ulasan positif didapatkan berdasarkan frekuensi kata dalam ulasan. Berikut merupakan hasil dari visualisasi ekstraksi ulasan pengguna didalam ulasan positif.

semakin jarang kata-kata tersebut muncul dan semakin jarang digunakan kata-kata tersebut dalam menuliskan ulasan. Dari gambar 4.10 diatas diketahui bahwa kata-kata yang muncul pada *wordcloud* memberikan topik dengan ulasan positif. Setelah diketahui kata yang sering muncul pada ulasan dengan kelas sentimen, selanjutnya akan dilakukan pencarian asosiasi kata yang sering muncul secara bersamaan seperti gambar berikut:



Gambar 4. 11 Asosiasi Kata Pada Ulasan Kelas Sentimen Positif

Berdasarkan pada gambar asosiasi diatas, proses asosiasi merupakan proses ekstraksi informasi yang dilakukan secara berulang dengan cara menyaring kata yang memiliki hubungan dengan kata lain didasarkan pada relevansi kata dengan topik yang dibicarakan.

Asosiasi kata yang berkaitan dengan kata “Aplikasi” didapatkan informasi, bahwa setelah pengguna melakukan update, akses dalam penggunaan aplikasi PeduliLindungi mudah digunakan.

Asosiasi kata yang berkaitan dengan kata “Bantu” didapatkan informasi, bahwa setelah melakukan login aplikasi PeduliLindungi pengguna akan mendapatkan informasi yang sangat membantu untuk memantau masyarakat Indonesia selama masa pandemik Covid-19.

Asosiasi kata yang berkaitan dengan kata “Vaksin” didapatkan informasi, bahwa Lembaga di Indonesia menyediakan vaksin gratis yang membantu masyarakat dalam

membantu melawan virus corona dengan memberikan dosis-dosis yang telah ditetapkan, serta pengguna pada aplikasi yang ini melakukan vaksin dapat dilakukan di Puskesmas.

Asosiasi kata yang berkaitan dengan “Sertifikat” didapatkan informasi, bahwa klaim sertifikat pada aplikasi sangat mudah dan munculnya sertifikat pada aplikasi yang langsung dan tidak perlu menunggu dalam waktu yang lama. Selain itu juga, pengguna aplikasi tidak perlu melakukan input sertifikat kedalam aplikasi.

Asosiasi kata yang berkaitan dengan kata “Kasih” didapatkan informasi, bahwa pengguna memberikan bintang untuk aplikasi PeduliLindungi kepada instansi pemerintah melalui website *google play store*. Selain itu, pihak instansi akan membalas ulasan pengguna melalui wadah website yang sama pula.

Asosiasi kata yang berkaitan dengan kata “Terima” didapatkan informasi, bahwa pihak instansi menerima ulasan yang baik terkait aplikasi ciptaannya. Dimana ulasan-ulasan tersebut dibaca oleh admin yang mengelola yang berguna untuk melakukan improvisasi pada aplikasi.

Asosiasi kata yang berkaitan dengan kata “Baru” didapatkan informasi, bahwa dengan melakukan update pada versi terbaru dapat diketahui bahwa ketika pengguna menjalankan aplikasi PeduliLindungi menjadi lebih efisien.

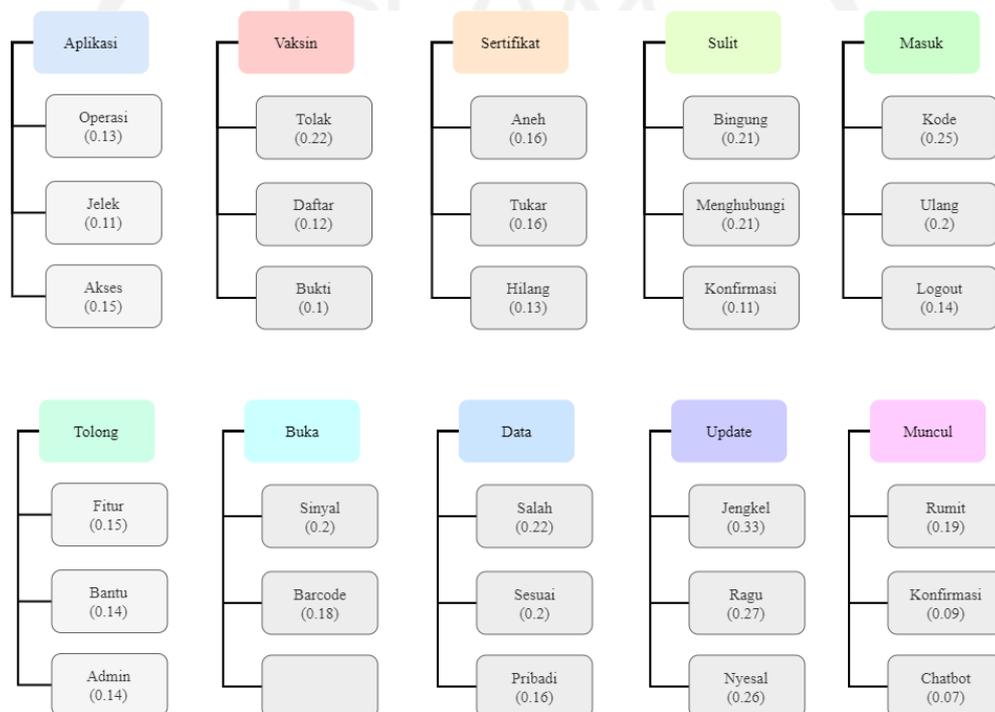
Asosiasi kata yang berkaitan dengan kata “Masuk” didapatkan informasi, bahwa pengguna yang baru menggunakan aplikasi di awal akan mendapatkan kode untuk dapat masuk ke menu utama dari aplikasi, selain itu untuk memudahkan dalam mengakses dokumen-dokumen pengguna akan diminta untuk memasukkan akun gmail untuk melengkapi data diri dan aplikasi ini sering digunakan dengan menggunakan aplikasi.

Asosiasi kata yang berkaitan dengan kata “Manfaat” didapatkan informasi, bahwa notifikasi yang masuk memberikan informasi yang bermanfaat terkait zonasi yang aman bagi masyarakat untuk melakukan aktivitas dalam jangka waktu lama dengan akurat. Selain itu juga dengan hal itu akan dapat melindungi diri dan mencegah dari penyebaran virus Covid-19.

4.2.5.2. Ulasan Negatif

Data ulasan sentimen merupakan data ulasan yang digunakan dari hasil pelabelan dan pembobotan. Nilai pembobotan yang digunakan pada kelas sentimen sentimen merupakan nilai pembobotan yang kurang dari nol dan merupakan kata-kata yang berisi kritikan. Untuk mendapatkan informasi terkait ulasan sentimen yang sering dibicarakan

Wordcloud merupakan salah satu metode visualisasi kata. Semakin besar kata yang ditampilkan pada *wordcloud* maka semakin kata tersebut muncul dan sering digunakan dalam menuliskan ulasan, sebaliknya semakin kecil ukuran kata yang ditampilkan maka semakin jarang kata-kata tersebut muncul dan semakin jarang digunakan kata-kata tersebut dalam menuliskan ulasan. Dari gambar 4.13 di atas diketahui bahwa kata-kata yang muncul pada *wordcloud* memberikan topik dengan ulasan positif. Setelah diketahui kata yang sering muncul pada ulasan dengan kelas sentimen, selanjutnya akan dilakukan pencarian asosiasi kata yang sering muncul secara bersamaan seperti gambar berikut:



Gambar 4. 14 Asosiasi Kata Pada Ulasan Kelas Sentimen Negatif

Asosiasi kata yang berkaitan dengan kata “Aplikasi” didapatkan informasi, bahwa dalam melakukan pengoperasian pada aplikasi masih memiliki akses yang jelek sehingga hal ini membuat pengguna kesulitan dalam menggunakan aplikasi.

Asosiasi kata yang berkaitan dengan kata “Vaksin” didapatkan informasi, bahwa terdapat kesulitan saat ingin melakukan pendaftaran dan sering terjadinya penolakan ketika melakukan vaksin karena keterbatasan sesi dalam melakukan vaksinasi.

Asosiasi kata yang berkaitan dengan kata “Sertifikat” didapatkan informasi, bahwa seringkali terjadi kehilangan file yang ada pada aplikasi PeduliLindungi. Selain itu juga, pada saat melakukan pengisian data diri untuk mengklaim sertifikat akan memunculkan informasi aneh yang tidak diperlukan.

Asosiasi kata yang berkaitan dengan kata “Sulit” didapatkan informasi, bahwa sulitnya dalam menggunakan aplikasi karena kebingungan pengguna. Selain itu juga, ketika ingin menghubungi pihak instansi dalam membantu menggunakan aplikasi masih sulit dihubungi.

Asosiasi kata yang berkaitan dengan kata “Masuk” didapatkan informasi, bahwa ketika pengguna ingin masuk terdapat kode-kode yang tidak terlalu jelas sehingga membuat seringnya masuk secara berulang-ulang. Selain itu juga, pengguna kesulitan dalam melakukan login Kembali ketika sudah melakukan logout.

Asosiasi kata yang berkaitan dengan kata “Tolong” didapatkan bahwa masih terdapat beberapa fitur yang sulit dipahami oleh pengguna sehingga menimbulkan kesulitan dalam menggunakannya.

Asosiasi kata yang berkaitan dengan kata “Buka” didapatkan informasi, bahwa ketika pengguna ingin menggunakan aplikasi PeduliLindungi, didapatkan bahwa diperlukannya sinyal yang memadai, dimana sebelumnya di upgrade ke versi sekarang masih tidak memerlukan sinyal. Sehingga hal ini membuat pengguna tidak dapat mengakses file pada aplikasi.

Asosiasi kata yang terdapat pada kata “Data” didapatkan informasi, bahwa ketika pengguna ingin mengisi data diri untuk dapat mengakses aplikasi, selalu terjadinya kesalahan pengisian data yang seharusnya sudah sesuai. Hal ini membuat pengguna kebingungan saat ingin mengakses aplikasi PeduliLindungi.

Asosiasi kata yang terdapat pada kata “Update” didapatkan informasi, bahwa setelah dilakukannya update ke versi yang sekarang banyak terjadinya kesalahan dalam penggunaannya, sehingga hal ini membuat pengguna yang telah melakukan update ke versi 4.3 hingga 4.4.3.1.

Asosiasi kata yang terdapat pada kata “Muncul” didapatkan informasi, bahwa ketika ingin mengisi data diri pada aplikasi PeduliLindungi akan muncul chatbot untuk melakukan konfirmasi, namun hal ini membuat kebanyakan pengguna menjadi kesulitan dalam melakukan pengaksesan aplikasi.

BAB V PEMBAHASAN

5.1 Gambaran Umum Terkait Persepsi Pengguna Aplikasi PeduliLindungi

30 Maret 2022 hingga 14 Mei 2022 rating yang didapatkan sebesar 4.2, hal ini mengindikasikan rating aplikasi ini mengalami penurunan sebanyak 0.2 dengan awalan 4.4. Aplikasi PeduliLindungi sudah diunduh sebanyak sebanyak 50 juta lebih pengguna dengan 979 ribu ulasan (01/08/2022). Rata-rata rating yang diberikan pengguna dari aplikasi ini sebesar 4.1 dari 5 bintang. Berdasarkan hal tersebut menunjukkan bahwa pelayanan yang diberikan belum sepenuhnya memenuhi harapan pengguna.

Berdasarkan data ulasan pengguna aplikasi PeduliLindungi pada 30 Maret hingga 14 Mei 2022 didapatkan ulasan yang didapatkan sebanyak 5.077 ulasan. Berdasarkan kurun waktu yang telah ditentukan diketahui bahwa pengguna banyak memberikan ulasan pada 5 sampai 11 Mei 2022 seperti yang dapat dilihat pada gambar 4.6. setelah dilakukan scrapping data selanjutnya dilakukan *pre-processing* data. Pada tahap ini data yang awalnya sebanyak 5.077 ulasan menjadi sebanyak 3.683 ulasan. Hal ini terjadi karena adanya ulasan yang hanya berisikan emoticon atau simbol yang mengakibatkan akan menghilangkan data ulasan ketika dilakukannya pembersihan data. Data yang didapatkan setelah dilakukannya pembersihan data selanjutnya dilakukan kelas sentimen dengan melakukan pembagian kategori dan didapatkan bahwa ulasan dengan kelas sentimen positif sebanyak 747 data ulasan, ulasan dengan kelas netral sebanyak 1.578 data ulasan, dan ulasan dengan kelas sentimen negatif sebanyak 1.358 data ulasan. Adanya kelas sentimen netral disebabkan karena adanya kata-kata yang tidak terdeteksi pada dan dinilai kurang memberikan manfaat bagi pihak instansi pemerintah sehingga label kelas sentimen tidak digunakan. Oleh karena itu hanya digunakan dua pelabelan kelas sentimen yaitu sentimen positif dan sentimen negatif. Berdasarkan hasil akhir pelabelan kelas sentimen, maka didapatkan bahwa hanya 2.105 data ulasan yang digunakan dengan 747 data ulasan berlabel positif dan 1.358 data ulasan berlabel negatif hal ini ditunjukkan pada gambar 4.8.

Informasi analisis sentimen yang relevan dan menginterpretasikan rating pada aplikasi PeduliLindungi bisa didapatkan dari penjelasan diatas. Hasil analisis sentimen dan rating aplikasi PeduliLindungi menunjukkan bahwa masih terdapat eror pelayanan.

Oleh karena itu diperlukannya evaluasi dan perbaikan kinerja dari pelayanan aplikasi PeduliLindungi.

5.2 Hasil Klasifikasi Pada *Naïve Bayes Classifier* dan *Support Vector Machine*

Proses analisis klasifikasi pertama kali dilakukan dengan melakukan pembagian *data training* dan *data testing*. Dimana, pada penelitian ini sendiri terdapat tiga perbandingan dengan persentase perbandingan yang berbeda. Setiap bentuk perbandingan dilakukan dengan melakukan percobaan sebanyak 3 kali.

Tabel 5. 1 Perbandingan Akurasi Model Klasifikasi Algoritma NBC dan SVM

Percobaan	Akurasi Model	
	NBC	SVM
Perbandingan 90:10		
Percobaan 1	82.46%	91.94%
Percobaan 2	87.68%	91.94%
Percobaan 3	81.52%	91.94%
Rata-Rata	83.89%	91.94%
Perbandingan 80:20		
Percobaan 1	85.75%	94.29%
Percobaan 2	80.76%	94.77%
Percobaan 3	85.04%	95.01%
Rata-Rata	83.85%	94.69%
Perbandingan 70:30		
Percobaan 1	85.28%	94.77%
Percobaan 2	81.65%	94.93%
Percobaan 3	84.18%	94.93%
Rata-Rata	83.70%	94.88%
Rata-Rata Total	83.81%	93.84%

Pada pengolahan data yang telah dilakukan, pembagian pertama dilakukan dengan melakukan pembagian *data training* dan *data testing* sebesar 90% (1.894 ulasan) dan data testing sebesar 10% (421 ulasan). Pembagian kedua dilakukan dengan melakukan pembagian *data training* dan *data testing* sebesar 80% (1.684 ulasan) dan *data testing*

sebesar 20% (211 ulasan). Serta pembagian ketiga dilakukan dengan melakukan pembagian *data training* dan *data testing* sebesar 70% (1.473 ulasan) dan *data testing* sebesar 30% (632 ulasan).

Pada percobaan pertama dengan perbandingan 90:10 diketahui bahwa pada algoritma *Naïve Bayes Classifier* (NBC) didapatkan nilai rata-rata akurasi sebesar 83.89% sedangkan pada algoritma *Support Vector Machine* (SVM) untuk nilai rata-rata akurasi didapatkan sebesar 91.94%. Pada percobaan kedua dengan perbandingan 80:20 diketahui bahwa pada algoritma *Naïve Bayes Classifier* (NBC) didapatkan nilai rata-rata akurasi sebesar 83.85% sedangkan pada algoritma *Support Vector Machine* (SVM) untuk nilai rata-rata akurasi didapatkan sebesar 94.69%. Sedangkan pada percobaan kedua dengan perbandingan 70:30 diketahui bahwa pada algoritma *Naïve Bayes Classifier* (NBC) didapatkan nilai rata-rata akurasi sebesar 83.70% sedangkan pada algoritma *Support Vector Machine* (SVM) untuk nilai rata-rata akurasi didapatkan sebesar 94.88%. Berdasarkan hal tersebut dapat diketahui bahwa pada proporsi pembagian data disetiap perbandingan algoritma SVM memiliki performa yang lebih baik dibandingkan dengan algoritma NBC. Hasil klasifikasi dengan menggunakan bantuan algoritma NBC dan SVM diketahui bahwa pada algoritma *Naïve Bayes Classifier* (NBC) didapatkan nilai rata-rata total akurasi sebesar 83.81% sedangkan pada algoritma *Support Vector Machine* (SVM) untuk nilai rata-rata akurasi didapatkan sebesar 93.84%. Berdasarkan hal tersebut dapat diketahui bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan dengan algoritma NBC.

Selain itu juga, pada algoritma NBC performansi yang memiliki nilai akurasi tertinggi dari setiap perbandingan terdapat pada proporsi perbandingan 90:10. Hal ini dapat dilihat bahwa nilai akurasi yang didapatkan lebih besar dibandingkan dengan proporsi perbandingan 80:20 dan 70:30. Oleh karena itu, berdasarkan dari pengolahan data yang telah dilakukan dapat diketahui bahwa semakin tinggi proporsi pembagian data untuk *data training* atau data latih, maka semakin tinggi pula akurasi yang dihasilkan. Hal ini juga dapat dilihat dalam jurnal (Gunawan, et al., 2018) yang berjudul “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode *Naïve Bayes*” menggunakan proporsi pembagian data sebesar 80:20 dan 90:10 didapatkan bahwa dengan pembagian data latih lebih besar akan menghasilkan akurasi paling tinggi. Sedangkan, pada algoritma SVM didapatkan bahwa nilai akurasi tertinggi dari setiap perbandingan terdapat pada proporsi perbandingan 70:30. Hal ini dapat dilihat bahwa

nilai akurasi yang didapatkan lebih besar dibandingkan dengan proporsi perbandingan 80:20 dan 90:10. Berdasarkan pengolahan data tersebut dapat diketahui bahwa tidak menjamin semakin tinggi proporsi perbandingan data akan semakin tinggi juga performansi akurasi yang dihasilkan.

Hasil yang telah didapatkan pada pengolahan data dengan menggunakan algoritma NBC dan SVM dapat diketahui bahwa akurasi yang dihasilkan pada algoritma SVM lebih baik dibandingkan algoritma NBC. Berdasarkan pada jurnal (Utami & Masripah, 2021) dan (Giovani, et al., 2020) hasil akurasi tertinggi terdapat pada algoritma SVM. Pada penelitian ini juga dapat diketahui dengan menggunakan proporsi pembagian data yang lebih kecil algoritma SVM juga masih memiliki model klasifikasi dengan nilai akurasi yang baik dibandingkan dengan algoritma NBC. Selain itu juga menggunakan algoritma ini karena kemampuan generalisasi yang baik jika menggunakan data dengan sampel yang berukuran kecil. Algoritma ini relatif baik ketika memiliki pemisah yang jelas antar kelasnya dan algoritma ini juga memiliki performansi lebih baik karena dapat memecahkan permasalahan yang kompleks dengan bantuan kernel.

5.3 Hasil Klasifikasi dan Asosiasi Kata

Berdasarkan dari 747 ulasan positif terdapat beberapa kata yang sering muncul diantaranya adalah kata “aplikasi” muncul paling banyak dengan frekuensi 296 kali, kata “bantu” dengan frekuensi sebesar 220 kali, kata “vaksin” dengan frekuensi sebesar 192 kali, dan seterusnya yang dapat dilihat pada gambar 4.9. kata-kata yang terdapat pada gambar diatas merupakan kata-kata yang memiliki makna positif dan banyak diulas oleh pengguna aplikasi PeduliLindungi.

Dari 1.358 data ulasan yang memiliki makna negatif, kata-kata yang sering muncul diantaranya adalah kata “aplikasi” muncul paling banyak dengan frekuensi 680 kali, kata “vaksin” dengan frekuensi sebesar 671 kali, kata “sertifikat” dengan frekuensi sebesar 489 kali, dan begitu seterusnya yang dapat dilihat pada gambar 4.12. kata-kata yang sering muncul ini selanjutnya akan digunakan untuk analisis asosiasi kata yang bertujuan untuk mendapatkan suatu informasi baru yang jelas.

Hasil asosiasi untuk kelas analisis sentimen positif dapat dilihat pada gambar 4.11 dimana hal ini menunjukkan proses ekstraksi informasi dengan asosiasi yang dilakukan secara berulang dengan cara menyaring kata-kata yang memiliki makna positif dengan

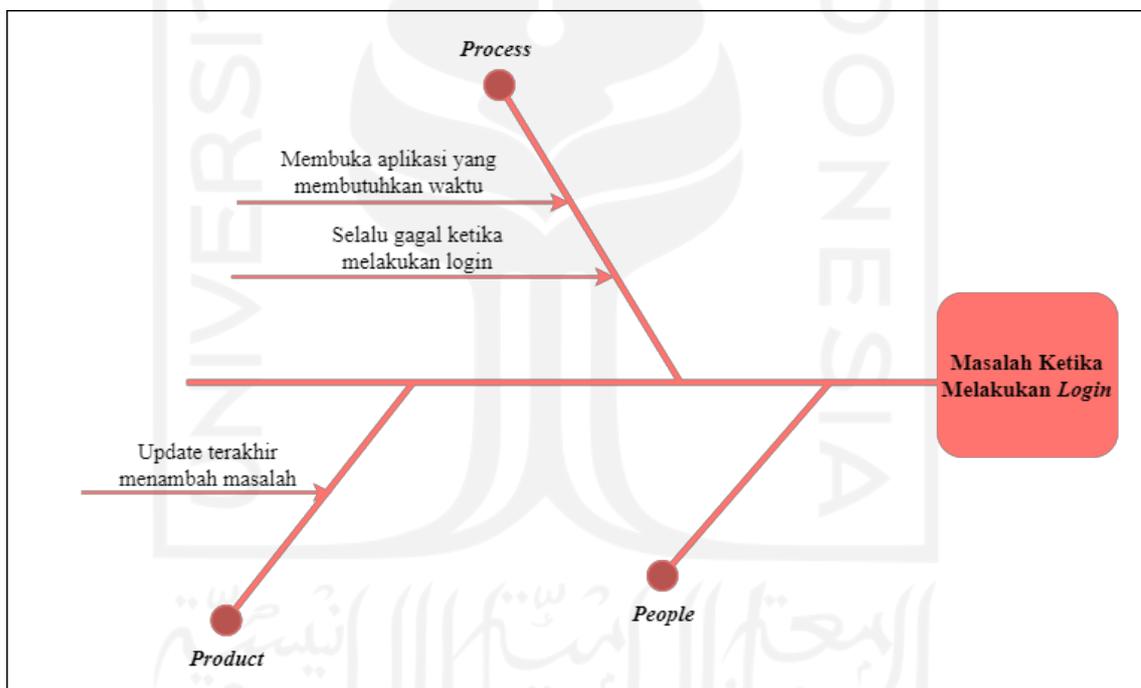
kata lain yang memiliki relevansi dengan kata yang akan diulas. Asosiasi kata yang akan diulas yaitu “aplikasi”, “bantu”, “vaksin”, “sertifikat”, “kasih”, “terima”, “baru”, “bagus”, “masuk” dan “manfaat”. Beberapa pengguna aplikasi PeduliLindungi setelah melakukan update ke versi 4.3 hingga 4.4.3.1 menjadi mudah dalam melakukan pengaksesan. Selain itu juga penggunaan aplikasi ini dapat membantu dalam mendapatkan informasi yang sangat membantu. Aplikasi ini dibuat agar memudahkan masyarakat Indonesia ketika telah melakukan vaksinasi. Sertifikat vaksinasi juga mudah dan langsung bisa disimpan tanpa menunggu dalam waktu yang lama. Selain itu juga, ketika pengguna ingin mengajukan ulasan admin akan membalas ulasan tersebut. Selain itu juga menggunakan versi 4.3 hingga 4.4.3.1 diketahui bahwa ketika menjalankan aplikasi PeduliLindungi menjadi lebih efisien.

Hasil klasifikasi untuk kelas analisis sentimen negatif dapat dilihat pada gambar 4.12 dimana hal ini menunjukkan asosiasi antar kata pada ulasan negatif, yaitu “aplikasi”, “vaksin”, “sertifikat”, “sulit”, “masuk”, “tolong”, “buka”, “data”, “*update*”, “tanggal”, “lahir”, “muncul”. Penggunaan aplikasi PeduliLindungi ketika melakukan pengoperasian pada aplikasi masih memiliki akses yang jelek sehingga hal ini membuat pengguna kesulitan dalam menggunakan aplikasi. Selain itu juga, informasi bahwa terdapat kesulitan saat ingin melakukan pendaftaran dan sering terjadinya penolakan ketika melakukan vaksin karena keterbatasan sesi dalam melakukan vaksinasi. Beberapa pengguna aplikasi yang telah melakukan vaksinasi mendapatkan bahwa seringnya terjadi kehilangan file (sertifikat) yang ada pada aplikasi PeduliLindungi. Pengguna yang melakukan *logout* di aplikasi PeduliLindungi juga mulai kesulitan untuk masuk kembali saat menggunakan versi ini sehingga hal ini membuat pengguna kebingungan dalam menggunakannya. Dalam menggunakannya juga pada versi ini memerlukan sinyal yang memadai untuk dapat digunakan berbeda dengan versi sebelumnya dan itu berdampak bahwa pengguna menyesal menggunakan aplikasi dengan versi ini.

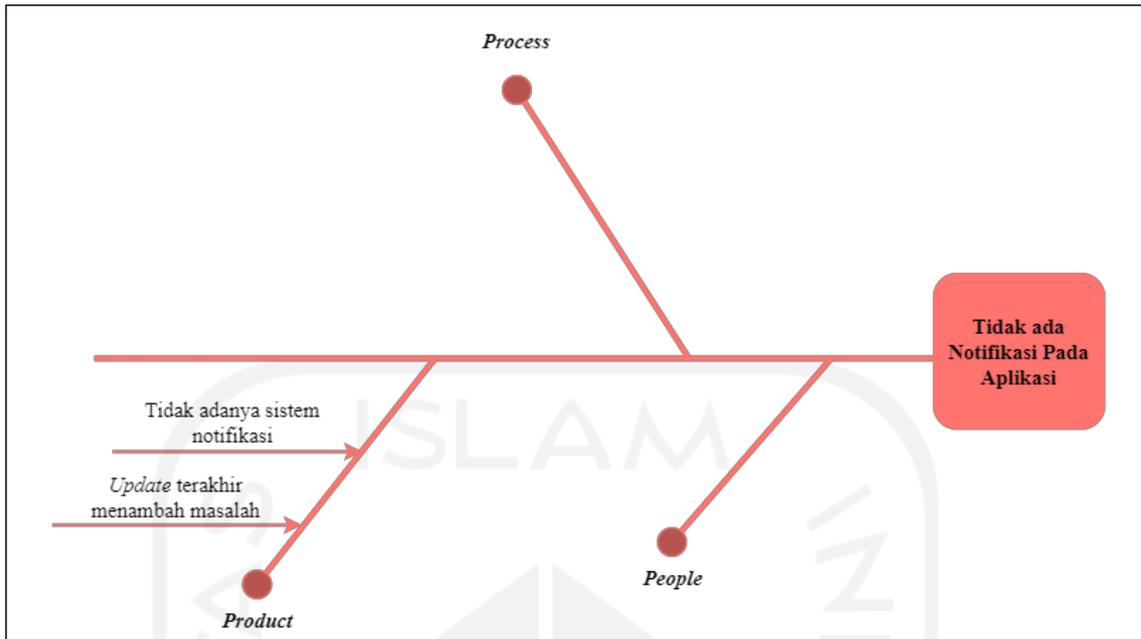
Berdasarkan ulasan dengan kelas sentimen positif dan ulasan dengan kelas sentimen negatif, didapatkan bahwa beberapa pengguna ketika menggunakan aplikasi PeduliLindungi dengan versi 4.3 hingga 4.4.3.1 tidak merasa terjadi kesulitan dan kebingungan, namun beberapa pengguna yang lain yang mengguna aplikasi ini dengan versi yang sama merasa kesulitan, kebingungan, dan ketidaknyamanan dalam penggunaannya. Sehingga banyak pengguna memberikan ulasan negatif dan diperlukannya evaluasi untuk perbaikan kedepannya.

5.4 Rekomendasi Berdasarkan Usulan Negatif

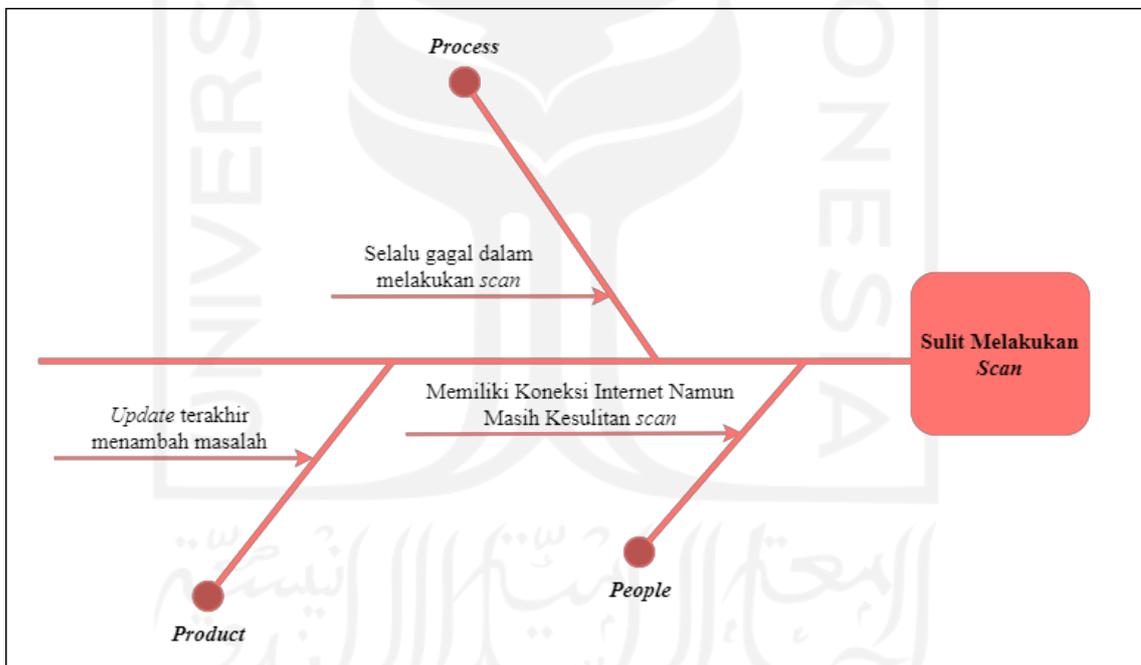
Berdasarkan hasil ulasan kelas sentimen negatif dari pengguna aplikasi PeduliLindungi, maka diperoleh informasi mengenai permasalahan yang tengah dihadapi pengguna. Oleh karena itu, dalam membantu permasalahan yang telah didapatkan, maka diperlukan analisa dengan menggunakan diagram *fishbone*. Diagram fishbone yang digunakan dalam bentuk penggunaan untuk perusahaan jasa yang terdapat 8P faktor, yaitu: *Price, People, Place/Plant, Procedures, Promotion, Processes, Product, dan Policies* (Anderson, 2020). Berdasarkan dari pengolahan data yang telah dilakukan, penentuan permasalahan menggunakan digram fishbone ini hanya menggunakan faktor *product, process, dan people*. Berikut merupakan diagram *fishbone* yang ditampilkan seperti dibawah ini:



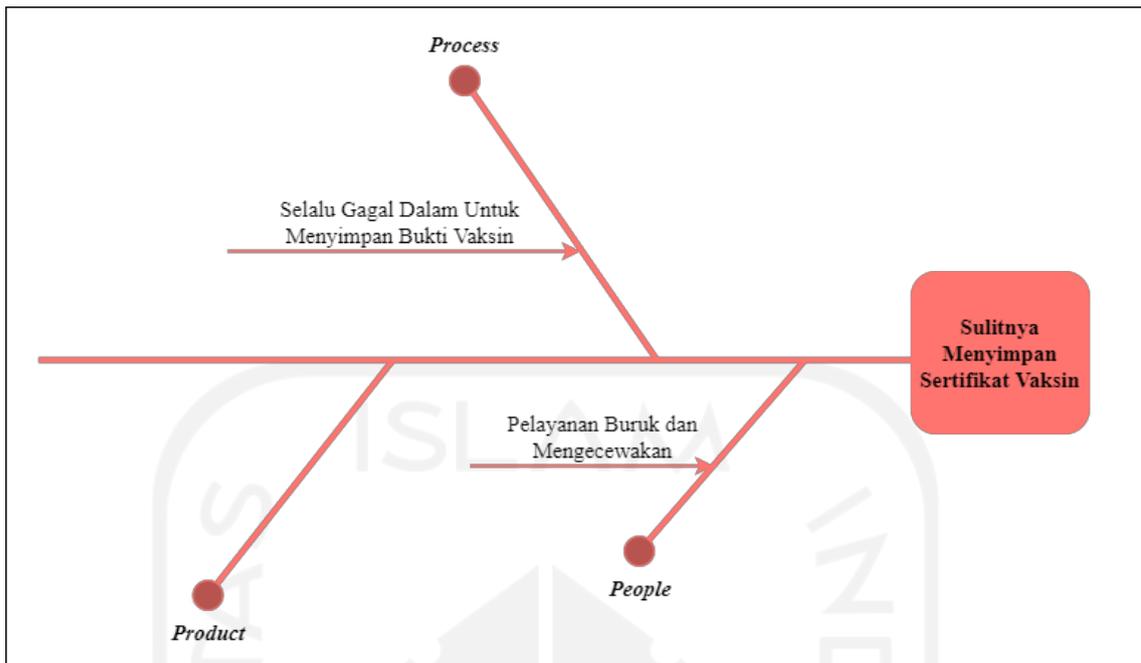
Gambar 5. 1 Diagram *Fishbone* Terkait Permasalahan Login



Gambar 5. 2 Diagram *Fishbone* Terkait Permasalahan Tidak Terdapat Notifikasi



Gambar 5. 3 Diagram *Fishbone* Terkait Permasalahan Sulit Melakukan Scan



Gambar 5. 4 Diagram *Fishbone* Terkait Permasalahan Sulit Menyimpan Bukti Vaksin

Berdasarkan diagram *fishbone* diatas, didapatkan informasi bahwa penyebab aplikasi PeduliLindungi mendapatkan ulasan negatif karena disebabkan oleh beberapa faktor, yaitu *product*, *people*, dan *process*. Setelah diketahui penyebab permasalahan berdasarkan beberapa faktor yang telah didapatkan, selanjutnya akan dilakukan langkah untuk pemecahan masalah. Berikut merupakan tabel dari pemecahan masalah pada layanan aplikasi:

Tabel 5. 2 Usulan Pada Permasalahan Aplikasi PeduliLindungi

No	Faktor	Permasalahan	Usulan Permasalahan
1	<i>People</i>	Pelayanan buruk dan mengecewakan	Memperketat kontrol dan pengawasan terkait keluhan pengguna yang dapat dijadikan evaluasi agar dapat lebih maksimal dalam memberikan pelayanan.
		Tidak adanya koneksi internet	Melakukan perbaikan pada <i>server</i> terkait sistem penggunaan koneksi

No	Faktor	Permasalahan	Usulan Permasalahan
			yang dapat digunakan secara <i>offline</i> agar pengguna lebih mudah dan efisien dalam mengakses aplikasi
2	<i>Product</i>	Tidak ada sistem notifikasi Update terakhir menambah masalah	Menambahkan sistem pengaturan pada aplikasi terkait notifikasi informasi agar memudahkan pengguna untuk mendapatkan informasi. Melakukan evaluasi dan uji coba terkait aplikasi yang telah dilakukan <i>updating</i> terlebih dahulu sebelum digunakan secara massal agar masalah yang ada pada versi sebelumnya tidak terulang kembali.
3	<i>Process</i>	Membuka aplikasi yang membutuhkan waktu Selalu gagal ketika melakukan <i>login</i>	Pihak developer segera mencari masalah dan melakukan perbaikan sehingga dapat meningkatkan efisiensi dalam membuka aplikasi. Pihak developer segera mencari masalah dan melakukan perbaikan sehingga pengguna dapat

No	Faktor	Permasalahan	Usulan Permasalahan
			dengan mudah untuk melakukan <i>login</i> .
		Selalu gagal dalam menyimpan bukti vaksin	Melakukan perbaikan sistem aplikasi terkait <i>tools</i> penyimpanan file pengguna agar masyarakat yang telah melakukan vaksinasi segera mendapatkan sertifikat.
		Selalu gagal dalam melakukan <i>scan</i>	Pihak developer segera mencari masalah terkait proses <i>scanning</i> menggunakan aplikasi PeduliLindungi dan melakukan perbaikan sehingga pengguna dapat dengan mudah melakukan <i>scan</i> ketika menggunakan aplikasi.

BAB VI PENUTUP

6.1 Kesimpulan

Berdasarkan dari analisis yang telah dilakukan, didapatkan beberapa informasi yang disimpulkan sebagai berikut:

1. Ulasan versi 4.3 hingga 4.4.3.1 aplikasi PeduliLindungi yang diambil selama periode 30 Maret 2022 hingga 14 Mei 2022 pada situs *Google Play*. Diketahui bahwa data ulasan yang didapatkan sebanyak 5.077 data ulasan. Setelah dilakukannya *pre-processing* dan analisis sentimen, data yang digunakan sebanyak 2.105 dengan jumlah pada kelas sentimen positif sebanyak 747 ulasan (35%) dan pada kelas sentimen negatif sebanyak 1.358 ulasan (65%). Topik ulasan pada kelas sentimen positif yang sering dibicarakan yaitu aplikasi, bantu, vaksin, sertifikat, kasih, terima, baru, bagus, masuk dan manfaat. Sedangkan untuk topik ulasan pada kelas sentimen negatif yang sering dibicarakan yaitu aplikasi, vaksin, sertifikat, sulit, masuk, tolong, buka, data, *update*, tanggal, lahir, dan muncul. Oleh karena itu, dapat disimpulkan bahwa topik ulasan yang sering dibicarakan pada ulasan kelas sentimen positif dan negatif tidak jauh berbeda karena dapat diketahui bahwa sebagian pengguna dapat menggunakan aplikasi dan sebagian pengguna kesulitan dalam menggunakan aplikasi pada versi ini.
2. Berdasarkan dari pengolahan data yang telah dilakukan dengan pembagian *data training* dan *data testing* sebesar 90:10, 80:20, dan 70:30 dengan dilakukannya percobaan sebanyak 3 kali pengacakan data untuk setiap perbandingan didapatkan hasil performansi akurasi bahwa algoritma *Support Vector Machine* (SVM) memiliki akurasi lebih tinggi pada setiap perbandingan pembagian data dibandingkan dengan menggunakan algoritma *Naïve Bayes Classifier* (NBC).
3. Berdasarkan analisis yang telah digunakan dengan menggunakan diagram *fishbone*, didapatkan bahwa terdapat 8 permasalahan yang teridentifikasi pada ulasan kelas sentimen negatif dari pengguna aplikasi PeduliLindungi yang terdapat pada 3P (*People, Product, dan Process*). Delapan permasalahan tersebut yaitu pelayanan buruk dan mengecewakan, kendala dalam menyimpan bukti vaksinisasi dari aplikasi, update terakhir menambah masalah, selalu gagal dalam menyimpan bukti vaksin, selalu gagal

ketika melakukan login, tidak ada notifikasi pada aplikasi, membuka aplikasi yang membutuhkan waktu, dan sulit melakukan *scan*.

6.2 Saran

Berdasarkan dari pengolahan data dan analisis yang telah dilakukan pada bab sebelumnya, saran yang dapat diberikan pada penelitian ini adalah sebagai berikut:

1. Bagi pihak Lembaga instansi pada aplikasi PeduliLindungi

Diharapkan pada Lembaga instansi yang bersangkutan dalam pembuatan dan pengembangan aplikasi PeduliLindungi berdasarkan hasil ekstraksi informasi dari ulasan-ulasan yang telah diberikan oleh pengguna dalam ulasan yang berbentuk negatif dapat dijadikan sebagai bahan evaluasi. Hal ini bertujuan agar dapat dilakukan perbaikan dan peningkatan dalam pelayanan penggunaan aplikasi PeduliLindungi untuk memberikan pelayanan dengan semaksimal mungkin.

2. Bagi peneliti selanjutnya

Diharapkan bagi peneliti selanjutnya dapat menggunakan algoritma pada *machine learning* yang berbeda dengan peneliti sekarang agar dapat diketahui perbandingan performansi antara algoritma baru dengan performansi pada algoritma yang telah digunakan. Selain itu juga, *dataset* yang digunakan memiliki perbandingan kelas sentiment positif dan kelas sentimen negatif yang seimbang, periode pengumpulan data yang lebih panjang agar informasi yang didapatkan lebih representatif.

DAFTAR PUSTAKA

- A. Yani, D. D., Pratiwi, H. S. and Muhardi, H. 2019. Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 7(4), p. 257. doi: 10.26418/justin.v7i4.30930.
- Afdhal, I., Kurniawan, R., Iskandar, I. & Salambue, R., 2022. Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia. *Jurnal Nasional Komputasi dan Teknologi Informasi*, pp. 122-130.
- Aggarwal, C. & C, Z., 2012. *Mining Text Data Chapter A Survey of Text Classification Algorithms*. London: Kluwer Academics Publisher.
- Anderson, S., 2020. Minitab. [Online] Available at: <https://blog.minitab.com/en/four-types-of-fishbone-diagrams> [Accessed 18 September 2022].
- AnkThon, 2021. *GeeksforGeeks*. [Online] Available at: <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/> [Accessed Sabtu Juni 2022].
- APPKEY, 2019. MARKEY. [Online] Available at: <https://markey.id/blog/technology/ai/deep-learning-adalah#:~:text=Selain%20itu%2C%20sebagai%20kelebihan%20dari,sulit%20untuk%20menyelaraskan%20hasil%20identifikasinya.>
- Arifin, O. & Sasongo, T. B., 2018. Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naive Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. *Seminar Nasional Teknologi Informasi dan Multimedia*, Volume 2, pp. 67-72.
- Arviana, G. N., 2021. *glints*. [Online] Available at: <https://glints.com/id/lowongan/sentimentanalysis/#.Yk6WcchBzIV> [Accessed Senin Maret 2022].
- Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer: Berlin.
- Brownlee, J., 2016. *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. s.l., Jason Brownlee.

- Chan, S., Wicaksono, N. & Chia, E., 2022. algoritma. [Online] Available at: <https://algorit.ma/blog/random-forest-adalah-2022/>
- Deviyanto, A. & Wahyudi, M. D. R., 2018. Penerapan Analisis Sentimen Pada Pengguna Twitter menggunakan Metode K-Nearest Neighbor. *JISKa (Jurnal Informatika Sunan Kalijaga)*, Volume 3, pp. 1-13.
- F. Maria Rosario B, Y. P., 2017. Penerapan Web Scraping Pada Website Company Profile. *Kntia*, Volume 4, pp. 37-43.
- Fang, X. and Zhan, J. 2015. Sentiment analysis using product review data. *Journal of Big Data*. Journal of Big Data, 2(1). doi: 10.1186/s40537-015-0015-2.
- Farach, D. & Nugraha, J., 2019. Implementasi Metode Naïve Bayes Classifier dalam Analisis Sentimen Pada Opini Masyarakat Tentang RUU KUHP. *Jurnal Advance in Social, Education and Humanities Research*, Volume 474.
- Feldman, R. & J, S., 2007. *The Text Mining Handbook: Advanced Approaches in Analysing Unstructured Data*. New York: Cambridge University Press.
- Gie, 2021. accurate. [Online] Available at: https://accurate.id/marketing-manajemen/decision-tree/#Kekurangan_Decision_Tree
- Giovani, A. P. et al., 2020. Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal TEKNOINFO*, Volume 14, pp. 116-124.
- Gorunescu, F., 2011. *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- Gunawan, B., Pratiwi, H. S. & Pratama, E. E., 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika*, Volume 4, pp. 113-118.
- Gumilang, Z.A. 2018. Implementasi Naive-Bayes Classifier dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-commerce Shopee pada Situs Google Play [Skripsi]. Yogyakarta (ID): Program Studi Statistika FMIPA Universitas Islam Indonesia.
- Handayani, F. & Pribadi, F. S., 2015. Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, Volume 7, pp. 19-24.

- Han, J. & Kamber, M., 2006. *Data Mining: Concepts and Techniques Second Edition*. 2nd ed. San Fransisco: Morgan Kauffman.
- Hidayat, R. & Minati, S., 2019. Comparative Analysis of Text Mining Classification Algorithm for English and Indonesia Qur'an Translation. *IJID International Journal on Informatics for Development*, Volume 8, pp. 47-51.
- Himawan, R. D. & Eliyanti, 2021. Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. *JEPIN (Jurnal Edukasi dan Penelitian Informatika*, Volume 7, pp. 58-63.
- Hu, M. & Liu, B., 2004. Mining and Summarizing Customer Reviews. *roceedings of The Tenth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining*, pp. 168-177.
- Hossain, B. e. a., 2019. Surgical Outcome Prediction in Total Knee Arthroplasty Using Machine Learning. *Intell Soft Comput* 25, pp. 105-115.
- Isnain, A. R., Marga, N. S. and Alita, D. 2021. Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(1), p. 55. doi: 10.22146/ijccs.60718.
- Johnson. R, A. W. D. W., 2007. *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Prentice Hall International, Inc.
- Kaswidjanti, W., Himawan, H. & Silitonga, P. D., 2020. The Accuracy Comparison of Social Media Sentiment Analysis Using Lexicon Based and Support Vector Machine on Souvenir Recommendations. *TEST Engineering & Management*, Volume 83, pp. 3953 - 3961.
- Kumar, L. and Bhatia, P. K. 2013. Available Online at www.jgrcs.info TEXT MINING : CONCEPTS , PROCESS AND APPLICATIONS. 4(3), pp. 36–39.
- Kwekha-Rashid, A. S., Abduljabbar, H. N. and Alhayani, B. 2021. Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience (Switzerland)*. Springer International Publishing, (0123456789). doi: 10.1007/s13204-021-01868-7.
- Manning, C., Raghavan, P. & Schütze, H., 2009. *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Maulana, A. A., Susanto, A. and Kusumaningrum, D. P. 2019. Rancang Bangun Web Scraping Pada Marketplace di Indonesia. *JOINS (Journal of Information System)*, 4(1), pp. 41–53. doi: 10.33633/joins.v4i1.2544.

- Mitchell, T. M., 1997. *Machine Learning*. New York: McGraw-Hill.
- Muis, I. A. & Affandes, M., 2015. Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet. *Jurnal Sains, Teknologi dan Industri*, Volume 12, pp. pp. 189-197.
- Murnawan & Sinaga, A., 2017. Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata. *Jurnal Penelitian Pos dan Informatika*, Volume 7, pp. 109-120.
- Muttaqin, N. M. & Kharisnudin, I., 2021. Analisis Sentimen Aplikasi Gojek Menggunakan Support Vector Machine dan K-nearest Neighbor. *UNNES Journal of Mathematics*, Volume 10, pp. 22-27.
- Naufal, M. F. & Kusuma, S. F., 2022. Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, pp. 44-49.
- O'Keefe, T. & K., 2009. Feature Selection and Weighting Methods in Sentiment Analysis. *Proceedings of the 14th Australasian Document Computing Symposium*.
- Pangaribuan, N. I., Simbolon, A. S. & Aruan, N. M., 2021. Analisis Sentimen Aplikasi E-Learning Selama Pandemi COVID-19 Dengan Menggunakan Metode Support Vector Machine dan Convolutional Neural Network. *SEMINASTIKA*, pp. 16-25.
- Parapat, I. M., Furqon, M. T. & S., 2018. Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 2, pp. hlm. 3163-3169.
- Qaiser, S. et al., 2021. A Comparison of Machine Learning Techniques for Sentiment Analysis. *Turkish Journal of Computer and Mathematics Education*, Volume 12, pp. 1738-1744.
- Qiang, W. & Zhongli, Z., 2011. Reinforcement Learning Model, Algorithms and Its Application. *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, pp. 1143-1146.
- Rachmatunnisa, 2021. detikinet. [Online] Available at: <https://inet.detik.com/security/d-5716818/ini-15-masalah-di-aplikasi-pedulilindungi-yang-harus-diperbaiki>

- Ranjan, N. et al., 2015. A Survey On text Mining Analytics and Classification Techniques For text Mining. *International Journal of Development Research*, Volume 5, pp. 5952-5955.
- S., Darmadi, E. A. & Santoso, L. H., 2021. Preferensi Konsumen Dalam Pemilihan Smartphone Berdasarkan Sistem Operasinya. *IKRAITH-EKONOMIKA*, Volume 4, pp. 175-180.
- Saleh, A., 2015. Implementation of Naïve Bayes Classifier Method for Predict Amount of Household Electricity use. *Citec Journa*, pp. 209-210.
- Salton, G. & Buckley, C., 1988. Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, pp. 513-523.
- Samuel, Y., Delima, R. & Rachmat, A., 2014. Implementasi Metode K-Nearest Neighbor dengan Decision Rule Untuk Klasifikasi Subtopik Berita. *Jurnal Informatika*, Volume 10, pp. 1-15.
- Sani, R. R., Zanierza, J. & Luthfiarta, A., 2016. Pengembangan Aplikasi Penentuan Tema Tugas Akhir Berdasarkan Data Abstrak menggunakan Algoritma K-Nearest Neighbor. *Prosiding seminar Nasional Multi Disiplin Ilmu & Call For Papers*, pp. 103-111.
- Santosa, B., 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sari, R., 2020. Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (KNN). *Jurnal Sains dan Manajemen*, Volume 8, pp. 10-17.
- Setiawan, W., 2014. Sistem Pakar Diagnosa Penyakit Mata Menggunakan Naive Bayes Classifier. *Seminar Nasional Sains dan Teknologi*.
- Sidiq, R. P., Dermawan, B. A. & Umaidah, Y., 2020. Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes. *Jurnal Informatika Universitas Pamulang*, Volume 5, pp. 356-363.
- Soepomo, P. 2014. Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes. *Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes*, 2(3), pp. 73–83.
- Thupae, R., Isong, B., Gasela, N. & Abu Mahfouz, A. M., 2018. machine Learning Techniques for Traffic Identification and Classification in SDWSN: A Survey.

IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, pp. 4645-4650.

- Utami, L. U. & Masripah, S., 2021. COMPARATION OF CLASSIFICATION ALGORITHM ON SENTIMENT ANALYSIS OF ONLINE LEARNING REVIEWS AND DISTANCE EDUCATION. *jurnal TECHNO Nusa Mandiri*, Volume 18, pp. 101-110.
- Vapnik, V. & Cortes, C., 1995. Support Vector Machine. *Machine Learning*, Volume 20, pp. 273-297.
- Vapnik, V. N., 1999. An Overview of Statistical Learning Theory. Volume 10(5), pp. 988-999.
- Wahyudi, R. and Kusumawardana, G. 2021. Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), pp. 200–207. doi: 10.31294/ji.v8i2.9681.
- Winanto, A. & Budihartini, C., 2022. Comparison of the Accuracy of Sentiment Analysis on the Twitter of the DKI Jakarta Provincial Government during the COVID-19 Vaccine Time. *Journal of Computer Science an Engineering (JCSE)*, Volume 3, pp. 14-27.
- Yustanti, W., 2012. Algoritma K-Nearest Neighbor untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika, Statistika, & Komputasi*, Volume 9, pp. 57-68.

Lampiran 1. Tahap *Pre-Processing* Menggunakan Python

```

!pip install Sastrawi
! pip install nltk

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import string
import re
import nltk

#import data
import csv

def load_data():
    df = pd.read_csv('Data.csv', delimiter = ';', encoding = 'unicode_escape')
    return df
data_df = load_data()
data_df.head()

df = pd.DataFrame(data_df[['Review', 'Tanggal']])
df

def remove(Review):
    Review = str(Review)
    Review = re.sub('[0-9]+', ' ', Review)
    Review = re.sub('#', ' ', Review)
    Review = re.sub('&', ' ', Review)
    Review = re.sub('\?', ' ', Review)
    Review = re.sub('\.', ' ', Review)
    Review = re.sub(',', ' ', Review)
    Review = re.sub('"', ' ', Review)
    Review = re.sub('\(', ' ', Review)
    Review = re.sub('\)', ' ', Review)
    Review = re.sub('\[', ' ', Review)
    Review = re.sub('\]', ' ', Review)
    Review = re.sub('/', ' ', Review)
    Review = re.sub('@', ' ', Review)
    Review = re.sub('<', ' ', Review)
    Review = re.sub('>', ' ', Review)
    Review = re.sub('%', ' ', Review)
    Review = re.sub('!', ' ', Review)
    Review = re.sub('-', ' ', Review)
    Review = re.sub('_', ' ', Review)
    Review = re.sub('¿', ' ', Review)
    # Review = re.sub('aku', ' ', Review)
    # Review = re.sub('gue', ' ', Review)
    # Review = re.sub('Fitur', ' ', Review)
    # Review = re.sub('gw', ' ', Review)
    # Review = re.sub('ini', ' ', Review)
    # Review = re.sub('semua', ' ', Review)
    # Review = re.sub('buat', ' ', Review)
    # Review = re.sub('kok', ' ', Review)
    # Review = re.sub('aja', ' ', Review)
    # Review = re.sub('nya', ' ', Review)
    # Review = re.sub('mendapatkan', ' ', Review)
    # Review = re.sub('dan', ' ', Review)
    # Review = re.sub('PeduliLindungi', ' ', Review)
    return Review

```

```

df['Review_symbol_cleaned'] = [remove(x) for x in df['Review']]

df.drop_duplicates(subset = 'Review_symbol_cleaned', keep = 'first',
inplace = True)
df

df = df.reset_index(drop = True)
df

df['Review_symbol_cleaned'] = df['Review_symbol_cleaned'].str.lower()

nltk.download('stopwords')
from nltk.corpus import stopwords

stopwords_list = stopwords.words('indonesian')

tempDf = df.copy()
tempDf

#import sastrawi libraries
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

#tokenize
from nltk.tokenize import TweetTokenizer

def tokenize_review(Remove):
    tokenizer = TweetTokenizer(preserve_case = False, reduce_len = True,
                              strip_handles = True)
    review_tokens = tokenizer.tokenize(Remove)
    review_clean = []
    for word in review_tokens:
        if (word not in stopwords_list and
            word not in string.punctuation):
            stem_word = stemmer.stem(word)
            review_clean.append(stem_word)

    return review_clean

tempDf['Review_tokenized'] = [tokenize_review(item) for item
in tempDf['Review_symbol_cleaned']]

tempDf.to_csv('Dataclean1x.csv', index = True)

```

Lampiran 2. Analisis Sentimen Menggunakan RStudio

```

install.packages("tm")
install.packages("wordcloud")
install.packages("RColorBrewer")
install.packages("stringr")
install.packages("ggplot2")
install.packages("plyr")
install.packages("xlsx")

library(tm)
library(wordcloud)
library(RColorBrewer)
library(stringr)
library(ggplot2)
library(plyr)
library(xlsx)

kalimat <- read.csv(file = "D:\\Kiki\\Explore\\TUGAS AKHIR\\Insyallahfix\\Dataclean1x.csv",
header = TRUE)
View(kalimat)

positif <- scan("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data FIX\\positive.txt", what = "character",
comment.char = ";")
negatif <- scan("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data FIX\\negative.txt", what = "character",
comment.char = ";")

kata.positif <- c(positif)
kata.negatif <- c(negatif)

score.sentiment = function(kalimat, kata.positif, kata.negatif,
.progress='none')
{
  require(plyr)
  require(stringr)
  scores = laply(kalimat, function(kalimat, kata.positif,
kata.negatif)
{
  kalimat = gsub('[[:punct:]]', '', kalimat)
  kalimat = gsub('[[:cntrl:]]', '', kalimat)
  kalimat = gsub('\\d+', '', kalimat)
  kalimat = tolower(kalimat)

  list.kata = str_split(kalimat, '\\s+')
  kata2 = unlist(list.kata)
  positif.matches = match(kata2, kata.positif)
  negatif.matches = match(kata2, kata.negatif)
  positif.matches = !is.na(positif.matches)
  negatif.matches = !is.na(negatif.matches)
  score = sum(positif.matches) - (sum(negatif.matches))
  return(score)
}, kata.positif, kata.negatif, .progress= .progress)
scores.df = data.frame(score=scores, Kalimat = kalimat)
return(scores.df)
}

hasil = score.sentiment(kalimat$Review_tokenized, kata.positif, kata.negatif)
head(hasil)
View(hasil)

#Tukar Row
data <- hasil[c(2,1)]
View(data)

write.xlsx(data, "pelabelan.xlsx", append = TRUE)

```

Lampiran 3. Pengolahan Data NBC 90:10 Menggunakan RStudio

```

install.packages("devtools")
install.packages("MIAMaxent")
install.packages("randomForest")
install.packages("RTextTools")
install.packages("doMC", repos="http://R-Forge.Rproject.org")
install.packages("SparseM")
install.packages("e1071")
install.packages("stats")
install.packages("dplyr")
install.packages("NLP")
install.packages("caret")
install.packages("readr")
install.packages("usethis")
install.packages("Rcpp")
install.packages("naivebayes")

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAMaxent)
library(readr)
library(tm)
library(naivebayes)

klasifikasiNbc <- read.csv(file = "D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\pelabelan1.csv", header = TRUE)
View(klasifikasiNbc)

set.seed(5)

#input data
klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
# klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
glimpse(klasifikasiNbc)

klasifikasiNbc$Sentiment <- as.factor(klasifikasiNbc$Sentiment)
corpus <- Corpus(VectorSource(klasifikasiNbc$Kalimat))
corpus
inspect(corpus[1:3])

corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(stripWhitespace)
dtm <- DocumentTermMatrix(corpus)
inspect(dtm[40:50, 10:15])

#Pembagian Training Set 90% dan Testing Set 10%
klasifikasiNbc.train <- klasifikasiNbc[1:1894,]
klasifikasiNbc.test <- klasifikasiNbc[1895:2105,]
dtm.train <- dtm[1:1894,]
dtm.test <- dtm[1895:2105,]
corpus.train <- corpus[1:1894]
corpus.test <- corpus[1895:2105]

```

```

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq

dtm.train.nb <- DocumentTermMatrix(corpus.train, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

dtm.test.nb <- DocumentTermMatrix(corpus.test, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x){
  y <- ifelse(x>0,1,0)
  y <- factor(y, levels = c(0,1), labels = c("No","Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

#Naive Bayes
naive=system.time(classifier <- naiveBayes(trainNB, klasifikasiNbc.train$Sentiment,
laplace = 0))
print(naive)
system.time (pred <- predict(classifier, newdata = testNB))

#Confusion Matrix
conf.mat <- confusionMatrix(table("Perdiction"=pred, "Actual"=
klasifikasiNbc.test$Sentiment))
conf.mat$bySentiment
conf.mat$overall
conf.mat$overall['Accuracy']
conf.mat

```

Lampiran 4. Pengolahan Data NBC 80:20 Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(naivebayes)

klasifikasiNbc <- read.csv(file = "D:\\kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\pelabelan1.csv", header = TRUE)
View(klasifikasiNbc)

set.seed(5)

#input data
klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
glimpse(klasifikasiNbc)

klasifikasiNbc$Sentiment <- as.factor(klasifikasiNbc$Sentiment)
corpus <- Corpus(VectorSource(klasifikasiNbc$Kalimat))
corpus
inspect(corpus[1:3])

corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(stripWhitespace)
dtm <- DocumentTermMatrix(corpus)
inspect(dtm[40:50, 10:15])

#Pembagian Training Set 80% dan Testing Set 20%
klasifikasiNbc.train <- klasifikasiNbc[1:1684,]
klasifikasiNbc.test <- klasifikasiNbc[1685:2105,]
dtm.train <- dtm[1:1684,]
dtm.test <- dtm[1685:2105,]
corpus.train <- corpus[1:1684]
corpus.test <- corpus[1685:2105]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq

dtm.train.nb <- DocumentTermMatrix(corpus.train, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

dtm.test.nb <- DocumentTermMatrix(corpus.test, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

```

```
#Fungsi Pelabelan
convert_count <- function(x){
  y <- ifelse(x>0,1,0)
  y <- factor(y, levels = c(0,1), labels = c("No","Yes"))
  y
}

trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

#Naive Bayes
naive=system.time(classifier <- naiveBayes(trainNB, klasifikasiNbc.train$Sentiment,
laplace = 0))
print(naive)
system.time (pred <- predict(classifier, newdata = testNB))

#Confusion Matrix
conf.mat <- confusionMatrix(table("Perdiction"=pred, "Actual"=
klasifikasiNbc.test$Sentiment))
conf.mat$bySentiment
conf.mat$overall
conf.mat$overall['Accuracy']
conf.mat
```

Lampiran 5. Pengolahan Data NBC 70:30 Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(naivebayes)

klasifikasiNbc <- read.csv(file = "D:\\Kiki\\Explore\\TUGAS AKHIR\\Data FIX\\pelabelan1.csv",
header = TRUE)
View(klasifikasiNbc)

set.seed(5)

#input data
klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
klasifikasiNbc <- klasifikasiNbc[sample(nrow(klasifikasiNbc)),]
glimpse(klasifikasiNbc)

klasifikasiNbc$Sentiment <- as.factor(klasifikasiNbc$Sentiment)
corpus <- Corpus(VectorSource(klasifikasiNbc$Kalimat))
corpus
inspect(corpus[1:3])

corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(stripwhitespace)
dtm <- DocumentTermMatrix(corpus)
inspect(dtm[40:50, 10:15])

#Pembagian Training Set 80% dan Testing Set 20%
klasifikasiNbc.train <- klasifikasiNbc[1:1894,]
klasifikasiNbc.test <- klasifikasiNbc[1895:2105,]
dtm.train <- dtm[1:1894,]
dtm.test <- dtm[1895:2105,]
corpus.train <- corpus[1:1894]
corpus.test <- corpus[1895:2105]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq

dtm.train.nb <- DocumentTermMatrix(corpus.train, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

dtm.test.nb <- DocumentTermMatrix(corpus.test, control = list(dictionary=fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x){
  y <- ifelse(x>0,1,0)
  y <- factor(y, levels = c(0,1), labels = c("No","Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

```

```
#Naive Bayes
naive=system.time(classifier <- naiveBayes(trainNB, klasifikasiNbc.train$Sentiment, laplace =
0))
print(naive)
system.time (pred <- predict(classifier, newdata = testNB))

#Confusion Matrix
conf.mat <- confusionMatrix(table("Perdiction"=pred, "Actual"=
klasifikasiNbc.test$Sentiment))
conf.mat$bySentiment
conf.mat$overall
conf.mat$overall['Accuracy']
conf.mat
```



Lampiran 6. Pengolahan Data SVM 90:10 Pemilihan Kernel Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(RTextTools)

#input data
positiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTrainSemPos.csv")
negatiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTrainSemNeg.csv")
positiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTestSepPos.csv")
negatiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTestSepNeg.csv")

reviewtraining = cbind(c(positiftraining, negatiftraining))
reviewtesting = cbind(c(positiftesting, negatiftesting))
review_all = cbind(c(reviewtraining, reviewtesting))

sentiment_training = c(rep("positiftraining", length(positiftraining)
),rep("negatiftraining", length(negatiftraining)))
sentiment_test = c(rep("positiftraining", length(positiftesting)
),rep("negatiftraining", length(negatiftesting)))
sentiment_all = as.factor(c(sentiment_training, sentiment_test))

mat = create_matrix(review_all, language = "indonesian", removeStopwords
= FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)
mat = as.matrix(mat)
container <- create_container(mat, sentiment_all, trainSize = 1:1894,
testSize = 1895:2105, virgin = FALSE)
model <- train_model(container, 'SVM', kernel = 'polynomial')
result <- classify_model(container, model)
table (as.character(sentiment_all[1895:2105]),
      as.character(result[,"SVM_LABEL"]))

#confusion matrix
akurasi <- recall_accuracy(sentiment_all[1895:2105],
result[,"SVM_LABEL"])
akurasi
create_precisionRecallSummary(container, result)

```

Lampiran 7. Pengolahan Data SVM 90:10 Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(RTextTools)

#input data
positiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTrainSemPos.csv")
negatiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTrainSemNeg.csv")
positiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTestSepPos.csv")
negatiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\SVMorKNN\\DataTestSepNeg.csv")

set.seed(5)

reviewtraining = cbind(c(positiftraining, negatiftraining))
reviewtesting = cbind(c(positiftesting, negatiftesting))
review_all = cbind(c(reviewtraining, reviewtesting))

sentiment_training = c(rep("positiftraining", length(positiftraining)
),rep("negatiftraining", length(negatiftraining)))
sentiment_test = c(rep("positiftraining", length(positiftesting)
),rep("negatiftraining", length(negatiftesting)))
sentiment_all = as.factor(c(sentiment_training, sentiment_test))

mat = create_matrix(review_all, language = "indonesian", removeStopwords
= FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)
mat = as.matrix(mat)
container <- create_container(mat, sentiment_all, trainSize = 1:1894,
testSize = 1895:2105, virgin = FALSE)
model <- train_model(container, 'SVM', kernel = 'linear')
result <- classify_model(container, model)
table (as.character(sentiment_all[1895:2105]),
      as.character(result[,"SVM_LABEL"]))

#confusion matrix
akurasi <- recall_accuracy(sentiment_all[1895:2105],
result[,"SVM_LABEL"])
akurasi
create_precisionRecallSummary(container, result)

```

Lampiran 8. Pengolahan Data SVM 80:20 Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(RTextTools)

#input data
positiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\deldua\\DataTrainDelPos.csv")
negatiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\deldua\\DataTrainDelNeg.csv")
positiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\deldua\\DataTestDuaPos.csv")
negatiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\deldua\\DataTestDuaNeg.csv")

set.seed(10)

reviewtraining = cbind(c(positiftraining, negatiftraining))
reviewtesting = cbind(c(positiftesting, negatiftesting))
review_all = cbind(c(reviewtraining, reviewtesting))

sentiment_training = c(rep("positiftraining", length(positiftraining)
),rep("negatiftraining", length(negatiftraining)))
sentiment_test = c(rep("positiftraining", length(positiftesting)
),rep("negatiftraining", length(negatiftesting)))
sentiment_all = as.factor(c(sentiment_training, sentiment_test))

mat = create_matrix(review_all, language = "indonesian", removeStopwords
= FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)
mat = as.matrix(mat)
container <- create_container(mat, sentiment_all, trainSize = 1:1684,
testSize = 1685:2105, virgin = FALSE)
model <- train_model(container, 'SVM', kernel = 'linear') #dicoba diganti
SVM nya jadi KNN atau yang lain
result <- classify_model(container, model)
table (as.character(sentiment_all[1685:2105]),
      as.character(result[, "SVM_LABEL"]))

#confusion matrix
akurasi <- recall_accuracy(sentiment_all[1685:2105],
result[, "SVM_LABEL"])
akurasi
create_precisionRecallSummary(container, result)

```

Lampiran 9. Pengolahan Data SVM 70:30 Menggunakan RStudio

```

library(Rcpp)
library(SparseM)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAXent)
library(readr)
library(tm)
library(RTextTools)

#input data
positiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\tujtig\\DataTrainTujPos.csv")
negatiftraining = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\tujtig\\DataTrainTujNeg.csv")
positiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\tujtig\\DataTestTigPos.csv")
negatiftesting = readLines("D:\\Kiki\\Explore\\TUGAS AKHIR\\Data
FIX\\tujtig\\DataTestTigNeg.csv")

set.seed(10)

reviewtraining = cbind(c(positiftraining, negatiftraining))
reviewtesting = cbind(c(positiftesting, negatiftesting))
review_all = cbind(c(reviewtraining, reviewtesting))

sentiment_training = c(rep("positiftraining", length(positiftraining)
),rep("negatiftraining", length(negatiftraining)))
sentiment_test = c(rep("positiftraining", length(positiftesting)
),rep("negatiftraining", length(negatiftesting)))
sentiment_all = as.factor(c(sentiment_training, sentiment_test))

mat = create_matrix(review_all, language = "indonesian", removeStopwords
= FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)
mat = as.matrix(mat)
container <- create_container(mat, sentiment_all, trainSize = 1:1473,
testSize = 1474:2105, virgin = FALSE)
model <- train_model(container, 'SVM', kernel = 'linear')
result <- classify_model(container, model)
table (as.character(sentiment_all[1474:2105]),
      as.character(result[, "SVM_LABEL"]))

#confuion matrix
akurasi <- recall_accuracy(sentiment_all[1474:2105],
result[, "SVM_LABEL"])
akurasi
create_precisionRecallSummary(container, result)

```

Lampiran 10. Asosiasi Kata Sentimen Positif Menggunakan RStudio

```

library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(stringr)
library(wordcloud2)

#input data
docs <- read.csv(file = "D:\\Kiki\\Explore\\TUGAS AKHIR\\Data FIX\\positif.csv",
header = TRUE)
View(docs)
docs <- Corpus(VectorSource(docs$kalimat))
docs <- tm_map(docs, removeWords, c("pakai", "gak", "gk", "dah",
"nya", "nga", "yg", "udah", "tgl", "tadi", "kpd", "banget", "close", "dgn", "lot",
"pedulilindungi", "mohon"))
docs <- tm_map(docs, toSpace, "[[:punct:]]")
docs <- tm_map(docs, toSpace, "[[:digit:]]")

docs <- tm_map(docs, gsub, pattern="ngtidak", replacement = "tidak")
docs <- tm_map(docs, gsub, pattern="error", replacement = "eror")

docs <- tm_map(docs, stripwhitespace)
inspect(docs)

dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 100)

set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 5,
max.words=100, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))
findFreqTerms(dtm, lowfreq = 4)

v<-as.list(findAssocs(dtm, terms
=c("aplikasi","bantu","vaksin","sertifikat","kasih",
"terima", "baru","bagus", "masuk", "manfaat", "guna", "mantap"),
corlimit = c(0.13,0.07,0.13,0.07,0.17,0.13,0.17,0.17, 0.02,
0.1)))
View(v)

```

```
k<-barplot(d[1:20,]$freq, las = 2, names.arg =  
  d[1:20,]$word,cex.axis=1.2,cex.names=1.2,  
  main = "Most frequent words",  
  ylab = "Word frequencies",col =topo.colors(20))  
  
k<-barplot(d[1:15,]$freq, las = 2, names.arg =  
  d[1:15,]$word,cex.axis=1.2,cex.names=1.2,  
  main = "Most frequent words",  
  ylab = "Word frequencies",col = terrain.colors(20))  
  
termFrequency<- rowSums(as.matrix(dtm))  
termFrequency<- subset(termFrequency, termFrequency>=48)  
text(k,sort(termFrequency, decreasing = T)-  
  1,labels=sort(termFrequency, decreasing = T),pch = 6, cex = 1)
```



Lampiran 11. Asosiasi Kata Sentimen Negatif Menggunakan RStudio

```

library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(stringr)

#input data
docs <- read.csv(file = "D:\\Kiki\\Explore\\TUGAS AKHIR\\Data FIX\\negatif.csv",
header = TRUE)
View(docs)
docs <- Corpus(VectorSource(docs$Kalimat))
docs <- tm_map(docs, removeWords,c("pakai","gak","gk", "dah",
"nya","nga","yg","udah", "tgl", "tadi", "kpd", "kitir", "gin", "in", "lo", "gua",
"aja", "tanggap", "lahir"))
docs <- tm_map(docs, toSpace, "[[:punct:]]")
docs <- tm_map(docs, toSpace, "[[:digit:]]")

docs <- tm_map(docs, gsub, pattern="login", replacement = "masuk")
docs <- tm_map(docs, gsub, pattern="loading", replacement = "tunggu")
docs <- tm_map(docs, gsub, pattern="bug", replacement = "rusak")
docs <- tm_map(docs, gsub, pattern="susah", replacement = "sulit")
docs <- tm_map(docs, gsub, pattern="limit", replacement = "batas")
docs <- tm_map(docs, gsub, pattern="lelet", replacement = "lambat")

docs <- tm_map(docs, stripWhitespace)
inspect(docs)

dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 100)

set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 5,
max.words=100, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))
findFreqTerms(dtm, lowfreq = 4)

v<-as.list(findAssocs(dtm, terms
=c("aplikasi", "vaksin", "sertifikat", "sulit", "masuk",
"tolong", "buka", "data", "update", "muncul"),
corlimit =
c(0.02,0.01,0.03,0.03,0.07,0.1,0.17,0.09,0.2,0.01)))
View(v)

```

```

k<-barplot(d[1:20,]$freq, las = 2, names.arg =
  d[1:20,]$word,cex.axis=1.2,cex.names=1.2,
  main = "Most frequent words",
  ylab = "Word frequencies",col =topo.colors(20))

k<-barplot(d[1:15,]$freq, las = 2, names.arg =
  d[1:15,]$word,cex.axis=1.2,cex.names=1.2,
  main = "Most frequent words",
  ylab = "Word frequencies",col = terrain.colors(20))

termFrequency<- rowSums(as.matrix(dtm))
termFrequency<- subset(termFrequency, termFrequency>=48)
text(k,sort(termFrequency, decreasing = T)-
  1,labels=sort(termFrequency, decreasing = T),pch = 6, cex = 1)

```

