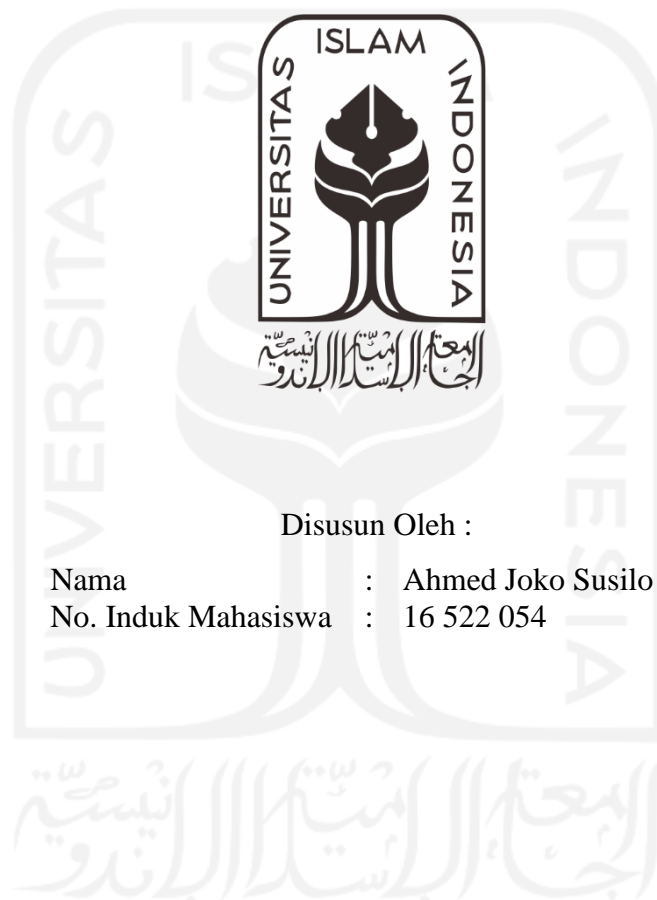


Analisis Sentimen Ulasan Pengguna Aplikasi Kesehatan (*Telemedicine*) Sebagai Pendukung Keputusan Dalam Meningkatkan Kualitas Aplikasi Kesehatan

TUGAS AKHIR

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



Disusun Oleh :

Nama : Ahmed Joko Susilo
No. Induk Mahasiswa : 16 522 054

**PROGRAM STUDI TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2020**

PERNYATAAN KEASLIAN

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

Yogyakarta, Desember 2020



Ahmed Joko Susilo

16522054

LEMBAR PENGESAHAN PEMBIMBING

**Analisis Sentimen Ulasan Pengguna Aplikasi Kesehatan (*Telemedicine*) Sebagai
Pendukung Keputusan Dalam Meningkatkan Kualitas Aplikasi Kesehatan**

TUGAS AKHIR



Disusun Oleh :

Nama : Ahmed Joko Susilo
No. Induk Mahasiswa : 16 522 054

Yogyakarta, Desember 2020

Dosen Pembimbing

Andrie Pasca Hendradewa, S.T., M.T

LEMBAR PENGESAHAN DOSEN PENGUJI

Analisis Sentimen Ulasan Pengguna Aplikasi Kesehatan (*Telemedicine*) Sebagai Pendukung Keputusan Dalam Meningkatkan Kualitas Aplikasi Kesehatan

TUGAS AKHIR

Oleh

Nama : Ahmed Joko Susilo
No. Induk Mahasiswa : 16 522 054

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

Yogyakarta, Januari 2021

Tim Penguji

Andrie Pasca Hendradewa, S.T., M.T.
Ketua

Ir. Ira Promasanti RD, M.Eng
Anggota I

Amarria Dila Sari, S.T., M.Sc.
Anggota II

Mengetahui

Ketua Program Studi Teknik Industri
Fakultas Teknologi Industri
Universitas Islam Indonesia



Laufiq Linnawan, ST., M.M.

HALAMAN PERSEMBAHAN

Alhadmulillahirobbil'alamin

Laporan tugas akhir ini saya persembahkan untuk kedua orang tua saya. Beliau yang telah membesarkan saya selama 22 tahun. Mungkin tidak lah cukup untuk membalas kebaikan dan kasih sayang yang diberikan. Namun, ini adalah langkah awal saya untuk membalas kebaikan Bapak dan Ibu, serta lebih berbakti lagi.



HALAMAN MOTTO

اقرأ

Iqra

Artinya: “Bacalah”

لَا حَوْلَ وَلَا قُوَّةَ إِلَّا بِاللَّهِ الْعَلِيِّ الْعَظِيمِ

Lā haula wa lā quwwata illā billāhil 'aliyyil azhīmi

Artinya ; “Tiada daya dan upaya kecuali dengan kekuatan Allah yang maha tinggi lagi
maha agung”

KATA PENGANTAR

Bismillahirrahmaanirrahiim,

Assalamu'alaikum Warakhmatullahi Wabarakatuh,

Alhamdulillah rabbil'alam, saya ucapkan rasa syukur kepada Allah SWT, yang telah melimpahkan rahmat, dan hidayah-Nya sehingga Tugas Akhir dengan judul “Analisis Sentimen Ulasan Pengguna Aplikasi Kesehatan (*Telemedicine*) Sebagai Pendukung Keputusan Dalam Meningkatkan Kualitas Aplikasi Kesehatan ” ini dapat terselesaikan. Sholawat dan salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW, beserta keluarga, sahabat, dan pengikut beliau hingga akhir zaman yang syafaatnya dinantikan di akhirat kelak.

Tugas akhir merupakan salah satu syarat untuk memperoleh gelar Sarjana di Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Dalam proses menyusun Tugas Akhir, Penulis menyadari tugas akhir yang disusun tidak lepas dari dukungan berbagai pihak. Oleh sebab itu, penulis ingin mengucapkan terima kasih kepada pihak, yang telah memberi dukungan, baik itu dukungan materi maupun non-materi. Penulis ingin menghaturkan ucapan terima kasih kepada:

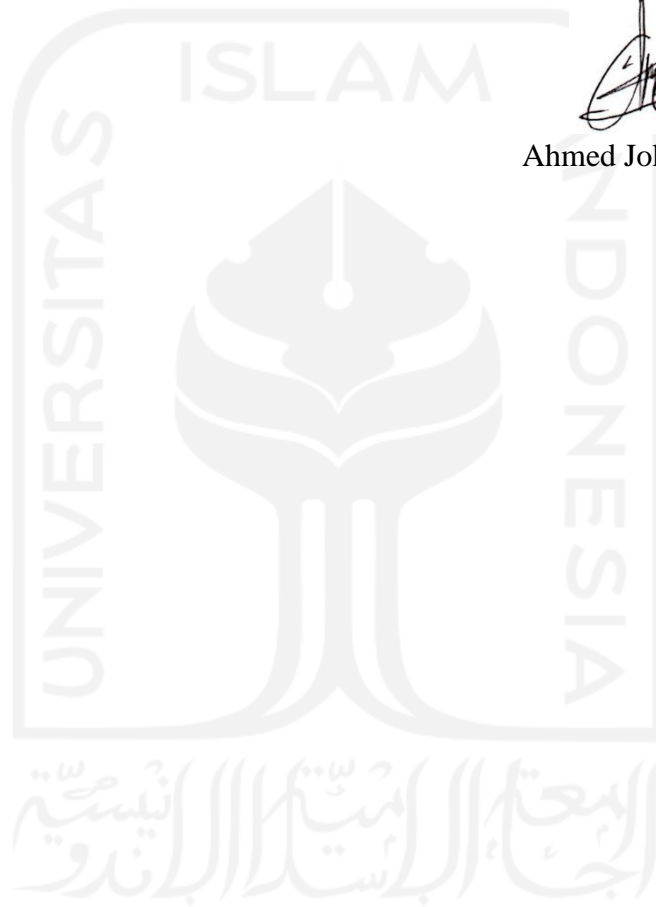
1. Bapak Prof. Dr. Ir. Hari Purnomo, selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia
2. Bapak Muhammad Ridwan Andi Purnomo S.T., M.Sc., Ph.D., selaku Ketua Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia
3. Bapak Dr. Taufiq Immawan, S.T., M.M., selaku Ketua Program Studi S-1 Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.
4. Bapak Andrie Pasca Hendradewa, S.T., M.T., selaku Dosen Pembimbing yang telah memberikan ilmu, bimbingan, saran, serta waktunya dalam penyusunan Tugas Akhir.
5. Bapak dan Ibu tercinta, terima kasih telah memberi dukungan moril dan materi, semangat dan doa, yang selalu diucapkan untuk keberhasilan saya.
6. Seluruh sahabat yang tidak dapat saya sebutkan satu-persatu, terima kasih telah menjadi *support system* saat mengalami kejenuhan, serta saran dan motivasi yang sangat membangun.
7. Teman-teman seluruh angkatan 2016 Teknik Industri Universitas Islam Indonesia

Semoga dukungan dan bantuan dari semua pihak senantiasa mendapat balasan dari Allah SWT. Penulis memahami dalam tugas akhir ini masih terdapat kekurangan oleh sebab itu, penulis mengharapkan kritik dan saran yang membangun. Semoga tugas akhir ini dapat bermanfaat bagi semua pihak. *Aamiin Yaa Robbal 'Aalamiin.*
Wassalamu 'alaikum Warrahmaatullahi Wabarakatuh.

Yogyakarta, 8 Desember 2020



Ahmed Joko Susilo



ABSTRAK

Pengguna internet di Indonesia meningkat setiap tahunnya. Perangkat yang mendukung dan banyak digunakan dalam mengakses internet adalah smartphone. Hal ini menjadikan berkembangnya start-up atau bisnis yang mengembangkan sebuah aplikasi mobile yang tersedia pada smartphone. Pada tahun 2020, terjadi pandemi global yang disebabkan oleh virus covid-19. Adanya pandemi tersebut mengharuskan individu untuk menjaga jarak dalam upaya menghindari penularan virus tersebut. Hal ini mengakibatkan penggunaan aplikasi mobile menjadi sangat dibutuhkan, salah satunya adalah aplikasi kesehatan telemedicine. Pengguna dari aplikasi tersebut mengalami peningkatan jumlahnya, sehingga perlu adanya peningkatan kualitas pada aplikasi telemedicine dengan memanfaatkan kolom komentar/ulasan yang diberikan pengguna melalui google play store. Komentar/ulasan yang tersedia merupakan data tidak terstruktur sehingga pendekatan yang dapat dilakukan untuk mengolah data tersebut adalah dengan analisis sentimen. Pada penelitian ini, menggunakan pendekatan analisis sentimen untuk menunjukkan gambaran umum terkait aplikasi telemedicine dan keluhan yang dialami pengguna. Hasil penelitian menunjukkan pada periode Maret – Juni 2020 adalah sangat suka terhadap aplikasi telemedicine berjumlah 84,3%, suka berjumlah 9%, netral berjumlah 2,51%, tidak suka berjumlah 0,92%, dan sangat tidak suka berjumlah 3,27%. Informasi yang didapatkan pada kelas sentimen positif untuk keseluruhan aplikasi kesehatan (telemedicine) adalah aplikasi sangat membantu pengguna, selain itu aplikasi memiliki respon yang cepat serta pelayanan yang sopan dan ramah. Pada kelas sentimen negatif, informasi yang didapatkan, pengguna memberikan keluhan terkait data pengguna, adanya promosi yang dilakukan dengan melakukan telepon secara berulang kali, kegagalan dalam masuk aplikasi atau dalam melakukan pendaftaran, adanya pelayanan yang diberikan tidak ramah, dan fitur beli obat. Pada ulasan negatif tersebut teridentifikasi menjadi 16 buah keluhan yang kemudian dikelompokkan menjadi 5 kelompok yaitu people, price, process, product, dan promotion.

Kata Kunci: *analisis sentimen, kualitas, pendukung keputusan, telemedicine*

DAFTAR ISI

PERNYATAAN KEASLIAN	II
LEMBAR PENGESAHAN PEMBIMBING.....	III
LEMBAR PENGESAHAN DOSEN PENGUJI.....	IV
HALAMAN PERSEMBAHAN	V
HALAMAN MOTTO.....	VI
KATA PENGANTAR	VII
ABSTRAK.....	IX
DAFTAR ISI.....	X
DAFTAR TABEL.....	XIII
DAFTAR GAMBAR.....	XIV
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	5
1.6 Sistematika Penulisan.....	5
BAB II KAJIAN LITERATUR	7
2.1 Kajian Deduktif.....	7
2.1.1 Startup.....	7
2.1.2 <i>Telemedicine</i> (Telemedis).....	8
2.1.3 Kualitas	9
2.1.4 <i>Data Mining</i>	10
2.1.5 Analisis Sentimen	12
2.1.6 <i>Text Mining</i>	12
2.1.7 Pembobotan Kata.....	14
2.1.8 Klasifikasi	15
2.1.9 Support Vector Machine (SVM).....	16
2.1.10 <i>Association Rule</i>	19
2.1.11 Bauran Pemasaran (<i>Marketing Mix</i>).....	20
2.1.12 Diagram <i>Fishbone</i>	21

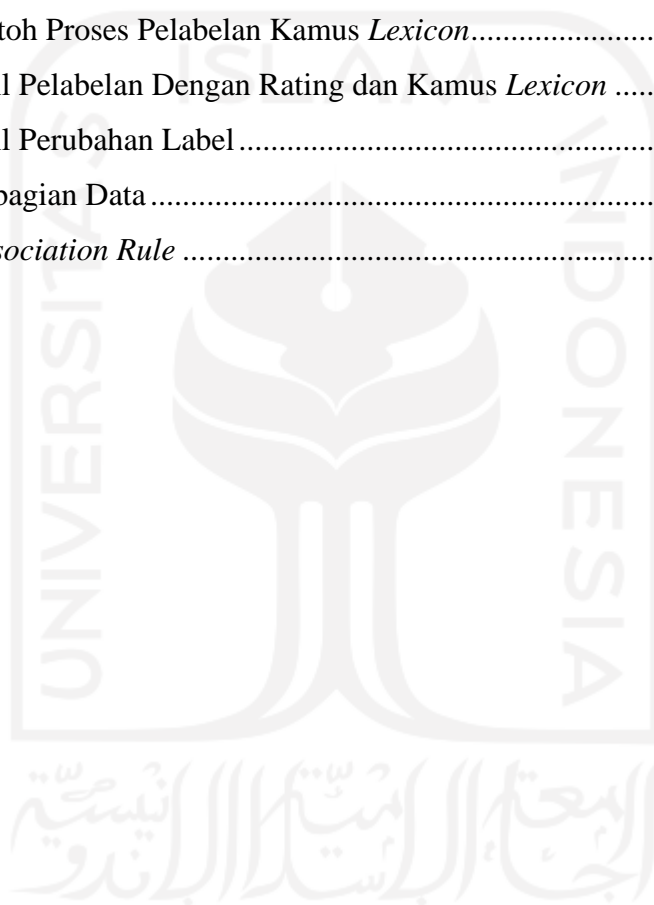
2.2	Kajian Induktif	1
BAB III METODE PENELITIAN		1
3.1	Objek Penelitian	1
3.2	Metode Pengumpulan Data	1
3.3	Diagram Alir Penelitian	2
3.3.1	Menentukan Topik Penelitian	4
3.3.2	Identifikasi Masalah	4
3.3.3	Studi Literatur	4
3.3.4	Pengumpulan Data	4
3.3.5	Pengolahan Data	4
3.3.6	Interpretasi dan Analisis	5
3.3.7	Kesimpulan dan Saran	5
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA		6
4.1	Pengambilan Data	6
4.2	Analisis Deskriptif	7
4.3	Sentimen	9
4.3.1	Data Cleaning dan Case Folding	9
4.3.2	Spelling Normalization	11
4.3.3	Tokenizing	12
4.3.4	<i>Stop Word Removal</i>	12
4.3.5	Stemming	13
4.3.6	Pelabelan	14
4.4	Klasifikasi	17
4.4.1	Pembagian Data Latih dan Data Uji	17
4.4.2	Klasifikasi <i>Support Vector Machine</i>	17
4.4.3	Visualisasi Data dan Asosiasi	18
4.4.4	<i>Fishbone</i> Diagram	25
BAB V ANALISIS DAN PEMBAHASAN		26
5.1	Analisis Deskriptif Pengumpulan Data Awal	26
5.2	Klasifikasi SVM	27
5.3	Visualisasi	28
5.3.1	Sentimen Positif	29
5.3.2	Sentimen Negatif	30

5.4	<i>Fishbone</i> Diagram.....	32
BAB VI PENUTUP		35
6.1	Kesimpulan	35
6.2	Saran.....	36
DAFTAR PUSTAKA		37
LAMPIRAN.....		45



DAFTAR TABEL

Tabel 2.1. Kajian Induktif.....	1
Tabel 4.1. Proses Membersihkan Ulasan.....	10
Tabel 4.2. Proses <i>Spelling Normalization</i>	11
Tabel 4.3. Proses Token.....	12
Tabel 4.4. Proses <i>Stopword Removal</i>	13
Tabel 4.5. Contoh Proses <i>Stemmer</i>	13
Tabel 4.6. Contoh Proses Pelabelan Kamus <i>Lexicon</i>	15
Tabel 4.7. Hasil Pelabelan Dengan Rating dan Kamus <i>Lexicon</i>	16
Tabel 4.8. Hasil Perubahan Label	16
Tabel 4.9. Pembagian Data	17
Tabel 4.10. <i>Association Rule</i>	24



DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses KDD.....	11
Gambar 2.2 <i>Confusion Matrix</i>	16
Gambar 2.3. Konsep SVM Linier	17
Gambar 2.4. SVM Non Linier	19
Gambar 2.5. <i>Fishbone</i> Diagram.....	21
Gambar 3.1. Diagram Alir Penelitian	2
Gambar 3.2. Diagram Alir Penelitian Lanjutan	3
Gambar 4.1. Tampilan Muka Appfollow.io	6
Gambar 4.2. Hasil <i>Crawling</i> dengan Appfollow.io	7
Gambar 4.3. Hasil <i>Cleaning Data</i> 1	8
Gambar 4.4. Jumlah Ulasan Pengguna Aplikasi Telemedicine <i>Google Play</i>	8
Gambar 4.5. Perbandingan Jumlah Ulasan Bulan Maret-Juni.....	9
Gambar 4.6. Proses Menghilangkan Data	10
Gambar 4.7. Proses Tahap Pertama Pelabelan Dengan Rating	14
Gambar 4.8. <i>Confusion Matrix</i> Skenario 3	18
Gambar 4.9. Akurasi Algoritma SVM.....	18
Gambar 4.10. Persebaran Seluruh Kata	19
Gambar 4.11. Persebaran Seluruh Kata (Stopword Removal II).....	20
Gambar 4.12. Persebaran Kata Kelas Sentimen Positif.....	20
Gambar 4.13. <i>Word cloud</i> Kelas Sentimen Positif	21
Gambar 4.14. Persebaran Kata Kelas Positif (Tahap <i>Stopword Removal II</i>)	21
Gambar 4.15. <i>Word cloud</i> Kelas Positif	22
Gambar 4.16. Persebaran Kata Kelas Sentimen Negatif	22
Gambar 4.17 <i>Word cloud</i> Kelas Sentimen Negatif.....	23
Gambar 4.18. Persebaran Kata Kelas Negatif (Tahap <i>Stopword Removal II</i>).....	23
Gambar 4.19. <i>Word cloud</i> Kelas Negatif.....	24
Gambar 4.20. <i>Fishbone</i> Ulasan Negatif.....	25

BAB I

PENDAHULUAN

1.1 Latar Belakang

Revolusi industri menunjukkan perkembangan teknologi yang terjadi di dunia. Saat ini revolusi industri sudah memasuki *industry 4.0* (Shwab, 2016). Pada era ini, perkembangan teknologi semakin pesat, dengan mengarah pada dunia digital. Perkembangan teknologi ini sangat membantu manusia, dengan muncul dan dikembangkannya sistem yang dapat menghubungkan dan mengontrol beberapa perangkat dengan suatu perangkat dari jarak yang jauh atau disebut *Internet of things* (IOT) (Junaidi, 2015). Tujuan dari IOT sendiri adalah untuk memperluas pemanfaatan konektivitas internet yang tersambung secara terus-menerus (Kusumaningrum, Pujiastuti, & Zeny, 2017). Sehingga banyak perangkat yang dikembangkan untuk dapat mengimplementasi sistem tersebut. Selain itu, munculnya perangkat lunak/aplikasi (*mobile app*) yang dapat mempertemukan dua atau lebih individu tanpa harus bertatap muka secara langsung. Sehingga komunikasi dapat dilakukan dengan mudah dimanapun dan kapan pun. Serta aliran informasi yang didapatkan menjadi lebih cepat. Hal tersebut dapat berkembang karena adanya bantuan teknologi internet yang terus berkembang.

Indonesia merupakan salah satu negara yang mengalami peningkatan jumlah pengguna internet. Jumlah penggunaan internet di Indonesia dari tahun ke tahun selalu mengalami peningkatan. Survey pada tahun 2017, jumlah pengguna internet Indonesia mencapai 143,26 juta jiwa. Jumlah tersebut menunjukkan 54,68% jumlah pengguna internet dari total penduduk Indonesia sebesar 262 juta orang. Pada tahun 2018 APJII

melakukan survey kembali dan menunjukkan peningkatan jumlah pengguna internet menjadi 171,17 juta jiwa dari total penduduk sebesar 264,16 juta orang. Alasan utama pengguna internet tersebut adalah komunikasi lewat pesan (APJII, 2018). Selain itu, Indonesia menduduki peringkat tiga dalam pertumbuhan penetrasi internet dunia (Ramadhan, 2020)

Meningkatnya jumlah pengguna internet tiap tahunnya, dan alasan utama pengguna internet adalah mudahnya komunikasi. Sehingga banyaknya individu untuk menggunakan perangkat yang dapat mengakses internet. Salah satu perangkat tersebut adalah kepemilikan dari *smartphone* yang mencapai 50,08% dari populasi, dan perangkat yang banyak dipakai dalam mengakses internet adalah *smartphone* dengan jumlah 44,16% berdasar pengguna internet (APJII, 2017). Hal tersebut menjadikan berkembangnya startup digital dengan mengembangkan *mobile app*. Pada perkembangannya jumlah *mobile app* tidak mencapai satu juta aplikasi di awal peluncuran layanan *google play store*, yaitu di tahun 2008 dengan nama android market. Pada tahun 2018 jumlah *mobile app* yang tersedia mencapai 3,6 juta aplikasi (Teknologi.id, 2018). Berbagai jenis *mobile app* tersedia pada layanan *google play store*, mulai dari hiburan, media sosial, editor, jasa transportasi, perdagangan (*marketplace*), dan kesehatan. Adanya ragam *mobile app* yang tersedia memberikan kemudahan pada pengguna. Salah satunya adalah aplikasi (*mobile app*) kesehatan yang dapat disebut aplikasi *telemedicine*. Pada aplikasi tersebut memberikan layanan kesehatan jarak jauh. Pengguna dapat melakukan konsultasi kesehatan dengan dokter secara daring (Telekonsultasi). Hal ini memberikan kemudahan pada pengguna untuk melakukan konsultasi kesehatan dan menjangkau daerah yang belum memiliki fasilitas pelayanan kesehatan.

Perkembangan aplikasi kesehatan dimulai dari pertemuan WHO (*World Health Organizations*) ke 58 tahun 2005. Pada pertemuan tersebut mengadopsi resolusi WHA 58.28 yang menyatakan setiap negara anggota mulai melakukan perencanaan pembangunan *e-Health* di negara masing-masing (Sunjaya, 2019). Di Indonesia perkembangan aplikasi (*mobile app*) kesehatan yang memberikan layanan *telemedicine* di mulai dari tahun 2015 dan regulasi yang mengatur tentang *telemedicine* dibuat pada tahun 2019 (Sansoko, 2020). Jumlah pengguna aplikasi digital berbasis kesehatan sebesar 10% dari total penduduk Indonesia di tahun 2019 (Petriella, 2019).

Pada tahun 2020, terjadi pandemi global yang ditetapkan pada tanggal 11 Maret 2020 oleh WHO yaitu virus COVID-19 (*Corona Virus Disease*). Adanya pandemi ini memberikan dampak pada semua sektor, mulai dari, kesehatan, sosial, dan ekonomi.. Namun, pada aplikasi *telemedicine* terdapat peningkatan jumlah pengguna sebesar 41 % (Lidwina & Fitra, 2020). Peningkatan tersebut dapat terjadi karena adanya kebijakan untuk masyarakat untuk melakukan *social distancing & physical distancing* yang membuat masyarakat untuk mengurangi interaksi bertemu secara langsung dan menjaga jarak tubuh. Selain itu, Pemerintah memberikan kebijakan dalam pelayanan kesehatan melalui surat edaran Menteri Kesehatan Nomor HK.02.01/MENKES/303/2020 tentang penyelenggaraan pelayanan kesehatan melalui pemanfaatan teknologi informasi dan komunikasi dalam rangka pencegahan penyebaran *corona virus disease* (Covid-19) (Setyowati & Annur, 2020).

Adanya peningkatan pengguna dalam menggunakan aplikasi tersebut maka pendapat dari pengguna selama pandemi ini sangat lah dibutuhkan untuk dapat meningkatkan kualitas dari pelayanan aplikasi *telemedicine*. Pada sebuah aplikasi yang tersedia dalam *website google play*, tersedia kolom komentar / ulasan yang dapat diberikan oleh pengguna aplikasi tersebut untuk memberikan pendapatnya (*feedback*). Ulasan / komentar tersebut termasuk dalam data tidak terstruktur. Hal tersebut karena data yang tersedia tidak berbentuk baris dan kolom dalam database.

Bidang yang berkaitan dengan proses mencari atau mengekstraksi informasi yang berguna dalam kumpulan data (*database*) adalah *data mining* atau *text mining* (Larose, 2005) (Feldman & Sanger, 2007). Perbedaan antara *data mining* dan *text mining* adalah Pada bentuk data yang digunakan. Pada *data mining* bentuk data yang digunakan adalah data terstruktur. Sedangkan bentuk data tidak terstruktur dapat dilakukan proses pengolahan dengan *text mining*.

Salah satu cabang dalam penelitian *text mining* adalah analisis sentimen, yang memiliki fungsi untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen (Rofiqoh, Perdana, & Fauzi, 2017). Menurut Liu (2010), Analisis sentimen merupakan studi komputasi yang menganalisis terkait opini/pendapat, sentimen, dan emosi yang

ditunjukkan melalui text. Maka penelitian ini menggunakan analisis sentimen untuk proses pengolahan data tidak terstruktur dan mendapatkan informasi dari hasil klasifikasi setiap kelas sentimen positif dan negatif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka berikut rumusan masalah dalam penelitian ini :

1. Bagaimana gambaran umum persepsi pengguna aplikasi (*mobile app*) *telemedicine* berdasarkan ulasan persepsi pada aplikasi (*mobile app*) *telemedicine* di website *Google Play*?
2. Informasi apa yang dapat diketahui dari hasil setiap klasifikasi kelas sentimen positif dan negatif yang telah dilakukan?
3. Faktor apa saja yang perlu dilakukan untuk memperbaiki atau meningkatkan kualitas layanan dari ulasan negatif yang di dapat?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengidentifikasi gambaran umum persepsi pengguna aplikasi (*mobile app*) *telemedicine* berdasarkan ulasan persepsi pada aplikasi (*mobile app*) *telemedicine* di website *Google Play*.
2. Memperoleh informasi penting dan bermanfaat dari hasil setiap klasifikasi yang telah dilakukan.
3. Mengidentifikasi faktor-faktor yang harus dilakukan untuk meningkatkan atau memperbaiki kualitas layanan dari hasil ulasan negatif.

1.4 Manfaat Penelitian

Manfaat pada penelitian ini sebagai berikut :

1. Mengetahui gambaran umum mengenai persepsi pengguna Alodokter, Klikdokter, Halodoc, SehatQ, dan *Good Doctor* berdasarkan data ulasan pada *website Google Play*.

2. Pihak yang memiliki kepentingan dapat melihat informasi yang ada dalam kumpulan ulasan atau komentar yang sangat banyak sehingga dapat fokus melakukan penanganan serta evaluasi ke arah yang lebih baik.
3. Memberikan khazanah ilmu pengetahuan mengenai analisis sentimen yang diterapkan pada industri kesehatan dan dapat dijadikan acuan bagi penelitian berikutnya.

1.5 Batasan Penelitian

Adapun batasan masalah yang ditentukan dalam penelitian ini yaitu sebagai berikut :

1. Pada penelitian ini menggunakan data ulasan mengenai aplikasi *Telemedicine* Alodokter, Klikdokter, Halodoc, SehatQ, dan *Good Doctor* pada *Google Play* yang diambil bulan Maret – Juni 2020
2. Ulasan yang digunakan adalah ulasan yang berbahasa Indonesia
3. Menggunakan data ulasan yang berasal dari *Google Play Store*

1.6 Sistematika Penulisan

Penelitian ini menggunakan kaidah penelitian ilmiah dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Pada Bab ini menguraikan latar belakang dari penelitian ini, serta memuat rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

Bagian ini memuat kajian deduktif berupa teori-teori pendukung dan menjadi dasar penelitian ini serta kajian induktif berupa penelitian-penelitian terdahulu.

BAB III METODE PENELITIAN

Pada bagian ini memuat objek penelitian, subjek penelitian, jenis data, metode pengumpulan data, metode pengolahan data, dan alur penelitian.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Pada bab ini menguraikan pengumpulan dan pengolahan data yang dilakukan dalam penelitian, serta menunjukkan hasil dari pengolahan data.

BAB V PEMBAHASAN

Bab pembahasan memuat tentang analisis secara detail mengenai hasil pengolahan data pada bab sebelumnya.

BAB VI PENUTUP

Pada bagian penutup memuat kesimpulan yang menjawab rumusan masalah secara singkat dan jelas serta saran mengenai pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA**LAMPIRAN**

BAB II

KAJIAN LITERATUR

2.1 Kajian Deduktif

2.1.1 Startup

Startup merupakan kata serapan dari bahasa inggris yaitu *start-up*. Secara bahasa pengertian tersebut jika diterjemahkan adalah “memulai”. Istilah ini pada mulanya digunakan untuk merepresentasikan suatu tindakan atau proses memulai suatu organisasi atau usaha bisnis, yang kemudian merujuk pada perusahaan yang belum lama beroperasi. Perusahaan-perusahaan ini merupakan perusahaan yang baru memulai operasinya atau didirikan dan masih dalam tahap pengembangan dan penelitian untuk menemukan pasar yang tepat. Hal tersebut lah yang kemudian startup dikenal atau disebut juga sebagai perusahaan rintisan (Liputan6.com, 2019).

Pada tahun 1990-2000 teknologi internet mulai banyak diketahui masyarakat dunia. Teknologi ini kemudian dimanfaatkan oleh perusahaan untuk memulai bisnis dengan membuka website pribadi. Pada tahun 1998-2000 banyak perusahaan dot-com yang mulai didirikan secara bersamaan. Kemudian istilah startup menjadi populer secara internasional ditahun tersebut yang dikenal dengan masa “*bubble dot-com*”. Tinggi nya pertumbuhan startup yang memanfaatkan teknologi tersebut menjadikan istilah startup dikenal dan diketahui memiliki hubungan yang erat dengan bidang teknologi (Kurnianti, 2017).

Saat ini istilah startup memiliki kaitan erat dengan suatu perusahaan yang memiliki inovasi dan pemanfaatan terhadap teknologi informasi. Secara umum perusahaan startup memiliki beberapa karakteristik sebagai berikut (Karsandi, 2017);

1. Usia perusahaan kurang dari 3 tahun
2. Jumlah pegawai kurang dari 20 orang
3. Pendapatan kurang dari 100.000 Dollar /tahun.
4. Produk yang dibuat berupa aplikasi dalam bentuk digital
5. Beroperasi dalam website
6. Posisi berada tahap berkembang dan beroperasi dalam bidang teknologi

2.1.2 *Telemedicine* (Telemedis)

Telemedis dalam bahasa Indonesia berasal dari kata “tele” yang memiliki arti “jarak jauh” dan “medis” yang berarti “bersifat kedokteran”. Secara internasional dan dalam bahasa inggris lebih dikenal dengan *telemedicine* (Prawiroharjo, Pratama, & Librianty, 2019). Secara singkat layanan telemedis berarti “layanan yang menggunakan fasilitas komunikasi elektronik untuk memberikan layanan atau dukungan medis dari jarak yang terpisah (tidak bertatap muka) (Field, 1996). Pada sejarahnya *telemedicine* sudah digunakan selama 20-30 tahun terakhir bersamaan dengan berkembangnya teknologi informasi dan komunikasi. Jika *telemedicine* dianggap sebagai aktivitas medis yang dilakukan dari jarak jauh maka memiliki sejarah yang lebih lama. Sebagai contoh adalah awal pengobatan jarak jauh saat adanya informasi terkait penyakit PES yang ditularkan ke seluruh eropa pada abad pertengahan. Dengan adanya perkembangan layanan pos nasional pada pertengahan abad ke-19, sarana untuk melakukan pengiriman layanan kesehatan, praktik dokter yang memberikan diagnosis dan arahan untuk penyembuhan, telah ditetapkan. Perkembangan signifikan selanjutnya pada akhir abad 19, dengan adanya komunikasi radio sarana dalam memberikan saran medis menjadi lebih cepat. Pada tahun 1920 Institut Gereja Pelaut New York menjadi salah satu organisasi pertama yang menyediakan perawatan medis menggunakan radio (Wootton, Craig, & Patterson, 2017).

Di Indonesia penggunaan *telemedicine* sudah dilakukan sejak tahun 90an, dan terus mengalami perkembangan. Beragam aplikasi sistem *telemedicine* telah dikembangkan

seperti telenursing, sistem telemedika puskesmas, sistem telemedika untuk pengelolaan wabah, sistem resep elektronik, sistem *e-health* untuk manajemen penyakit TBC, sistem *m-health*, sistem *e-health* berbasis pengolahan citra, sistem telemedika bergerak, sistem *e-psychology*, open EHR dan MRS sistem. Pengembangan tersebut masih dalam tahap pendidikan dan belum diterapkan di bidang medis pemerintahan (Santoso, Rahmah, Setiasari, & Sularsih, 2015). Saat ini perkembangan *telemedicine* mulai diterapkan di bidang medis pemerintahan. Dengan adanya regulasi dari pemerintah yaitu Peraturan Menteri Kesehatan Nomor 20 Tahun 2019 terkait penyelenggara pelayanan *telemedicine* antar fasilitas pelayanan kesehatan.

2.1.3 Kualitas

Kualitas merupakan suatu nilai dapat berupa sifat atau atribut yang diberikan oleh suatu produk atau jasa yang dapat memenuhi keinginan dan kepuasan dari pengguna atau konsumen (Ahyari, 2000). Definisi lainnya, kualitas dibagi menjadi dua yaitu, kualitas produk dan kualitas layanan/jasa. Kemampuan sebuah produk yang dapat menjalankan fungsinya, seperti durabilitas, reliabilitas, ketepatan, kemudahan penggunaan dan reparasi produk disebut sebagai kualitas produk (Kotler & Amstrong, 2012). Sedangkan kualitas layanan adalah tingkat keunggulan yang diharapkan dan pengendalian atas keunggulan tersebut untuk memenuhi keinginan pelanggan (Tjiptono, 2007). Dengan berkembangnya teknologi, pelayanan dapat dilakukan secara online yang kemudian kualitas dari layanan *online* ini dikenal dengan *web servqual* (*web service quality*) atau *e-servqual* (*elektronik service quality*) (Santoso J. , 2003). Model dari *e-servqual* kemudian dikembangkan dan memiliki dimensi sebagai berikut (Kim & Oh, 2011):

1. *Usability*

Usability merupakan pengukuran kualitas pengalaman pengguna yang berkaitan dengan pemanfaatan barang atau sistem termasuk Situs Web, perangkat lunak, dan komunikasi seluler (Pearson, Pearson, & Green, 2007).

2. *Aesthetics*

Pada dimensi ini, menyiratkan atribut-atribut atau fitur dengan penampilan yang terdapat di dalam perangkat lunak, dan situs web. Hal ini mencerminkan bagaimana informasi dikirimkan pada pengguna melalui kejelasan dan

keterbacaan campuran warna, font, animasi, gambar, efek suara, dan teks (Gummerus, Liljander, Pura, & van Riel, 2004).

3. *Security* (Keamanan)

Dimensi ini menjelaskan terkait tingkat kepercayaan pengguna terhadap keselamatan dan perlindungan informasi pribadi di situs Web, dan perangkat lunak (Zeithaml, Parasuraman, & Malhotra, 2000).

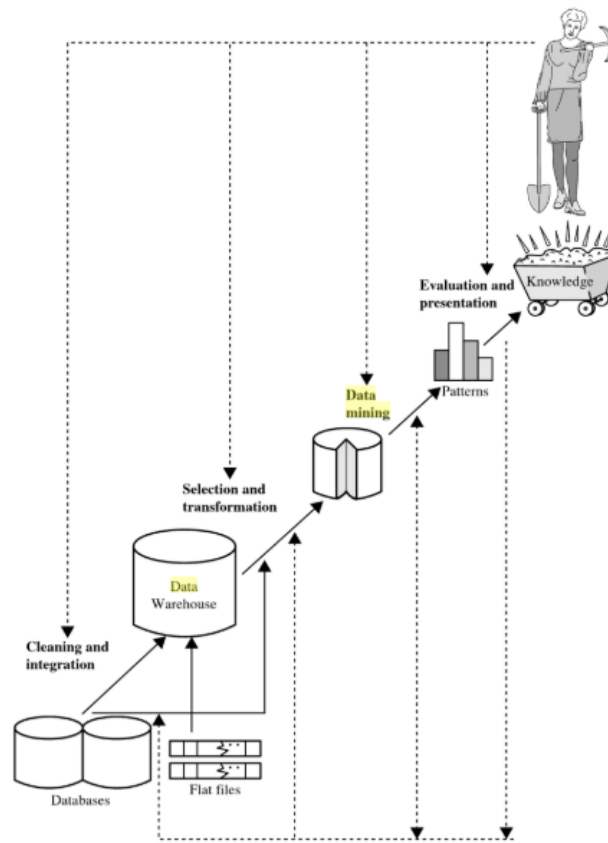
4. *Responsiveness*

Responsiveness adalah cepat dan tanggapnya penyedia layanan saat pelanggan/konsumen membutuhkan (Parasuraman, Zeithaml, & Berry, 1998).

2.1.4 *Data Mining*

Data mining adalah suatu proses untuk mendapatkan pengetahuan dari kumpulan data (*database*) dengan melakukan ekstraksi data sehingga dapat memberikan manfaat (Yanto & Kesuma, 2017). Menurut Gartner grup dalam (Larose, 2005), *Data mining* adalah proses menemukan korelasi, pola, dan tren baru yang memiliki makna atau arti dengan menyaring sejumlah data besar yang tersimpan, menggunakan teknologi pengenalan pola serta statistik dan matematika. *Data mining* merupakan inti dalam proses KDD (*Knowledge Discovery in Database*) yang melibatkan penyimpulan algoritma dalam ekstraksi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui (Maimon & Rokach, 2005).

Data mining terkadang dijadikan sinonim dari KDD (*Knowledge Discovery in Database*), sementara sebaliknya *data mining* tersebut merupakan salah satu langkah dalam proses KDD. Tahapan dalam proses KDD seperti gambar berikut;



Gambar 2.1 Tahapan Proses KDD
 Sumber : (Han, Kamber, & Pei, 2011)

Menurut Han et al (2011) proses KDD dapat dijelaskan sebagai berikut:

1. *Data Cleaning*
 Pada tahap ini proses menghilangkan data yang tidak konsisten dan *noise*
2. *Data Integration*
 Di tahap ini merupakan proses menggabungkan data, jika data yang diperoleh memiliki beberapa sumber data
3. *Data Selection*
 Pada tahapan ini adalah proses memilih data yang memiliki kaitan dengan tugas analisis
4. *Data Transformation*
 Proses mengubah ke bentuk yang dapat digunakan dalam proses *data mining*
5. *Data Mining*
 Tahap ini merupakan proses inti dalam melakukan ekstraksi pola data
6. *Pattern Evaluation*

Pada tahap ini adalah proses untuk mengidentifikasi pola menarik yang mewakili *Knowledge* di dalam data.

7. *Knowledge Presentation*

Tahap ini merupakan proses penggambaran dan teknik merepresentasikan pengetahuan yang telah dilakukan untuk menunjukkan hasil atau interpretasi dari pengetahuan yang didapatkan.

2.1.5 Analisis Sentimen

Analisis sentimen merupakan studi komputasi yang menganalisis terkait opini/pendapat, sentimen, dan emosi yang diekspresikan melalui text (Liu, 2010). Penelitian terkait hal tersebut mulai populer pada tahun 2002 dan terus berkembang. Analisis sentimen memberikan luaran yaitu informasi yang dikategorikan menjadi nilai positif dan negatif. Pada prosesnya untuk mendapatkan informasi tersebut, analisis sentimen menciptakan sebuah sistem yang kemudian dapat melakukan klasifikasi terhadap text dalam suatu dokumen. Pemanfaatan analisis sentimen dilakukan untuk memeriksa pendapat terhadap suatu produk atau suatu kejadian (Hidayat, 2015).

2.1.6 Text Mining

Menurut Feldman & Sanger (2007), *text mining* adalah bidang penelitian dari ilmu komputer yang mencoba memecahkan krisis kelebihan informasi dengan menggabungkan beberapa teknik, seperti *data mining*, *machine learning*, *natural language processing*, pencarian informasi, dan *knowledge management*. *Text mining* adalah proses mencari atau mengekstraksi informasi yang berguna dari data tekstual (Vijayarani, Ilamthi, & Nithya, 2015).

Text mining dan *data mining* memiliki proses yang sama, namun memiliki perbedaan yaitu pada bentuk data nya. Pada *data mining* memiliki bentuk data yang terstruktur, sedangkan pada *text mining* memiliki bentuk data yang tidak terstruktur atau semi struktur seperti pada email, HTML, file, dan keseluruhan dokumen tekstual (Gupta & Lehal, 2009). Data terstruktur adalah data yang sudah tersedia dalam data historis dalam bentuk catatan, seperti yang sudah tersedia di dalam database relasional dan *spreadsheet*. Data

tidak terstruktur merupakan lawan dari data terstruktur yang belum memiliki baris kolom dalam database. Data semi struktur merupakan data yang diketik dalam sistem basis data konvensional atau bukan data mentah (Vijayarani, Ilamthi, & Nithya, 2015).

2.1.6.1 *Text Preprocessing*

Text preprocessing merupakan tahapan untuk membersihkan dan mempersiapkan data sebelum dilakukan pengolahan lebih lanjut (Haddi, et al, 2013). Tujuan dari *text preprocessing* adalah untuk menyeragamkan bentuk kata, menghilangkan *noise*, dan mengurangi volume kata (Berry & KOGan, 2010). Menurut Feldman & Sanger dalam (Fatmawati & Affandes, 2017) proses dalam melakukan *text preprocessing* adalah sebagai berikut:

1. *Cleaning*

Tahap *cleaning* merupakan tahap untuk menghilangkan data yang tidak dibutuhkan atau membersihkan data yang tidak penting dan tidak mempengaruhi proses klasifikasi

2. *Tokenizing (Tokenization)*

Tokenizing merupakan tahap memecah kata dalam sebuah kalimat menjadi kata tunggal.

3. *Case Folding*

Case Folding adalah proses mengubah kata dalam bentuk huruf besar yang terdapat dalam sebuah kalimat menjadi bentuk huruf kecil.

4. *Spelling Normalization*

Spelling Normalization adalah proses dalam memperbaiki kata yang tidak sesuai dengan penulisan kata yang sebenarnya.

5. *Filtering*

Filtering adalah proses untuk menghilangkan kata-kata yang tidak penting. Proses filtering ini lebih dikenal dengan disebut *stopword removal*.

6. *Stemming*

Stemming adalah proses untuk mengubah kata-kata yang memiliki imbuhan atau kata majemuk menjadi kata dalam bentuk kata dasar yang sesuai dengan KBBI (Kamus Besar Bahasa Indonesia)

7. *Pembobotan (Term Weighting)*

Pembobotan merupakan proses merubah kata menjadi bentuk vektor, yang mana setiap kata dihitung menjadi satu fitur.

2.1.7 Pembobotan Kata

Pembobotan kata (*term weighting*) merupakan proses yang bertujuan untuk memberikan bobot pada setiap kata (*term*) yang terdapat pada dokumen teks yang akan diproses (Nurjanah, Perdana, & Fauzi, 2017). *Term* dapat berupa kata, frasa atau unit hasil *indexing* lainnya dalam suatu dokumen yang dapat digunakan untuk mengetahui isi atau konteks dari dokumen tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata diberikan sebuah indikator, yaitu *term weight* (Zafikri, 2008).

Dalam melakukan pembobotan kata terdapat berbagai metode yang dapat digunakan, dimana tahapan dalam pembobotan kata yang banyak digunakan (Nurjanah, Perdana, & Fauzi, 2017) (Sebastiani, 2002) sebagai berikut :

1. *Term Frequency* (TF)

Term frequency adalah frekuensi munculnya kata pada suatu dokumen teks. Persamaan *Term frequency* dapat dinotasikan sebagai berikut

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{lainnya} \end{cases} \quad (2.1)$$

2. *Document Frequency* (DF)

Document Frequency adalah frekuensi atau jumlah dokumen yang memiliki kata di dalamnya.

3. *Inverse Document Frequency* (IDF)

Inverse Document Frequency merupakan frekuensi kemunculan kata (*term*) pada keseluruhan dokumen teks. Persamaan *Inverse Document Frequency* dapat dinotasikan sebagai berikut :

$$idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right) \quad (2.2)$$

Keterangan :

N = Jumlah dokumen teks

$df_{(t)}$ = Jumlah dokumen yang mengandung kata (*term*) t

4. *Term Frequency – Inverse Document Frequency* (TF-IDF)

Pada tahap ini merupakan kombinasi dari nilai *tf* dan nilai *idf* dalam perhitungan bobot. Persamaan dapat dinotasikan sebagai berikut :

$$W_{t,d} = W_{tf_{t,d}} \times idf_t \quad (2.3)$$

Keterangan :

$W_{tf_{t,d}}$ = *Term frequency*

idf_t = *Inverse Document Frequency*

2.1.8 Klasifikasi

Klasifikasi merupakan salah satu bagian dalam tugas *data mining*. Tugas tersebut adalah melakukan pelabelan terhadap data yang belum memiliki label atau kelas data (Larose, 2005). Klasifikasi merupakan bentuk analisis data yang menghasilkan model prediksi untuk mendeskripsikan label atau kelas data (Han & Kamber, 2006). Pada prosesnya untuk menghasilkan model prediksi dan dapat mendeskripsikan label, terdapat empat komponen dasarnya yaitu (Gorunescu, 2011):

1. Kelas (*Class*) yaitu variabel dependen dari model yang berupa kategori dan menjelaskan sebuah label pada objek setelah proses klasifikasi
2. Prediktor (*Predictor*) yaitu variabel independen dari model yang diwakili oleh karakteristik (atribut) data yang akan diklasifikasikan dan berdasarkan klasifikasi yang dibuat.
3. *Training Set* yaitu kumpulan data yang berisi nilai-nilai atau *record* untuk dua komponen sebelumnya (kelas dan prediktor) dapat berupa variabel kontinyu maupun kategori, dan digunakan untuk melakukan pelatihan dalam membangun model prediksi.
4. *Testing Set* yaitu kumpulan data baru yang akan diklasifikasikan oleh model prediksi yang telah dibuat untuk mengukur tingkat akurasi klasifikasi (kinerja model) sehingga performansi model klasifikasi dapat dievaluasi.

Evaluasi kinerja model klasifikasi sangat penting dilakukan. Hal tersebut menunjukkan model klasifikasi atau prediksi yang telah dibuat dapat dengan tepat untuk melakukan pelabelan terhadap data baru. Untuk dapat menentukan nilai tersebut maka dibutuhkan yang namanya tabulasi *confusion matrix*. Berikut merupakan gambar *confusion matrix*:

CLASSIFICATION	PREDICTED CLASS		
	Class = YES	Class = NO	
OBSERVED CLASS	Class = YES	a <i>(true positive-TP)</i>	b <i>(false negative-FN)</i>
	Class = NO	c <i>(false positive-FP)</i>	d <i>(true negative-TN)</i>

Gambar 2.2 *Confusion Matrix*
 Sumber : (Gorunescu, 2011)

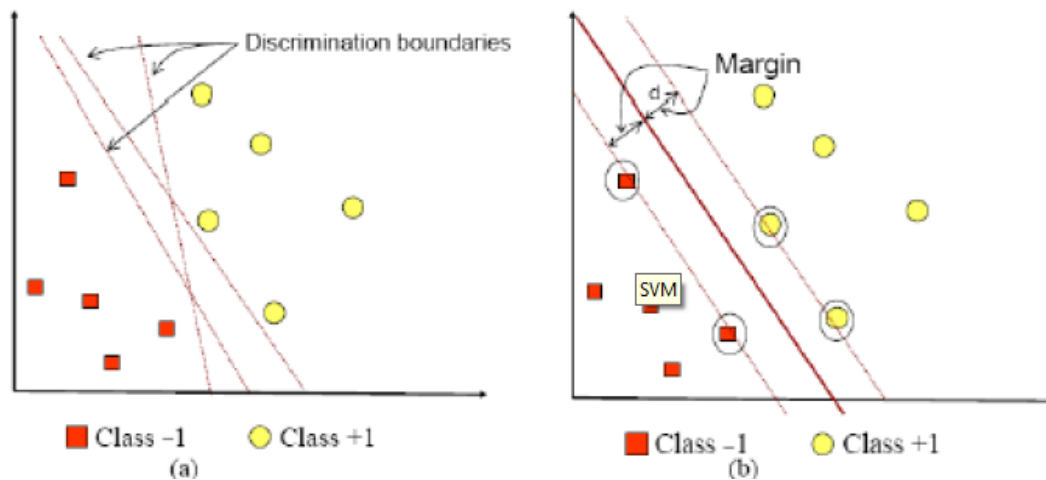
confusion matrix menunjukkan rincian detail dari kesalahan prediksi. Kelas prediksi dapat dilihat pada bagian kolom sedangkan kelas observasi pada bagian baris sebelah kiri. Setiap sel yang terdapat pada tabulasi tersebut berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas observasi yang ditugaskan oleh model kelas prediksi yang diberikan. Kemudian untuk menentukan akurasi dari model prediksi yang didapatkan, dapat menggunakan rumus sebagai berikut:

$$Akurasi = \frac{a + d}{a + b + c + d} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

2.1.9 Support Vector Machine (SVM)

Support vector machine merupakan suatu metode untuk melakukan prediksi dalam kasus klasifikasi maupun regresi (Santosa B. , 2007). *Support vector machine* termasuk dalam *supervised learning* yang berarti model atau mesin mempelajari terlebih dahulu untuk melakukan klasifikasi, dengan membagi data menjadi dua set data, yaitu data *training* dan data *testing*. Metode ini diperkenalkan pertama kali pada tahun 1992 oleh Vapnik. *Support vector machine* dapat menemukan fungsi pemisah yang bisa memisahkan dua set data dari dua kelas yang berbeda. Konsepnya secara sederhana dapat dijelaskan bahwa SVM berusaha untuk mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada input *space* dengan memaksimalkan jarak antar kelas (Praptiwi, 2018). SVM dapat digunakan pada jenis set data yang berbeda, yaitu set data yang terpisah secara linier, dan non linier.

2.1.9.1 Support Vector Machine (SVM) Linier



Gambar 2.3. Konsep SVM Linier

Sumber : (Buntoro, 2017)

Gambar 2.3. menunjukkan konsep dari *Support Vector Machine* (SVM) bekerja, dimana terdapat dua kelompok objek yang akan diklasifikasikan. Objek pertama adalah berbentuk persegi dengan warna merah, dan objek kedua adalah lingkaran berwarna kuning. Pada gambar (b) dapat diketahui untuk *hyperplane* terbaik yang dapat memisahkan dua kelas dengan jarak yang maksimum, dimana Objek yang diberi tanda lingkaran disebut *support vector*. Kemudian pada objek tersebut yang bertumbukan atau bersinggungan dengan garis, maka garis tersebut yang disebut sebagai *hyperplane*. Garis yang berada di antara dua *hyperplane* disebut sebagai margin.

Secara matematika konsep SVM dapat dijelaskan sebagai berikut, data yang tersedia dinotasikan $x \in R^d$ sedangkan label setiap kelas dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, 3, \dots, n$. Diasumsikan kedua kelas objek dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan (Praptiwi, 2018):

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.4)$$

Pola \vec{x} , yang termasuk dalam kelas +1 (sampel positif) dapat di definisikan sebagai berikut :

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (2.5)$$

Pada pola \vec{x} , yang termasuk kelas -1 (sampel negatif) dinotasikan sebagai berikut :

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2.6)$$

Dengan keterangan sebagai berikut :

R^d = ruang vektor

\vec{w} = vektor bobot

\vec{x} = vektor data (input)

b = bias

Pada pemisahan *hyperplane* dengan bentuk canonical harus memenuhi *constraint* atau batasan yang disebut sebagai fungsi kendala, dengan notasi sebagai berikut :

$$y_i[(\vec{w} \cdot \vec{x}) + b] \geq 1, i = 1, \dots, l \quad (2.7)$$

Kemudian *hyperplane* yang memisahkan data harus meminimalkan,

$$\min_w \tau(w) = \frac{1}{2} \|w\|^2 \quad (2.8)$$

2.1.9.2 Support Vector Machine (SVM) Non Linier

Pada SVM (*Support Vector Machine*) dengan jenis set data yang tidak terpisah secara linier (non linier) maka penerapannya yang digunakan adalah dengan memanfaatkan fungsi dari kernel (Puspitasari, Ratnawati, & Widodo, 2018). Fungsi tersebut adalah menerapkan set data yang diterapkan pada ruang berdimensi tinggi (*feature space*). Pada dasarnya klasifikasi data non linier memiliki notasi persamaan (Rachman & Purnami, 2012):

$$\bar{\alpha} = \operatorname{argmin} \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^l \alpha_i \quad (2.9)$$

Nilai $K(x_i, x_j)$ merupakan fungsi kernel yang menunjukkan pemetaan non linier pada *feature space*. Persamaan ini memberikan *hard classifier* pada pemisahan *hyperplane* di *feature space* dengan persamaan :

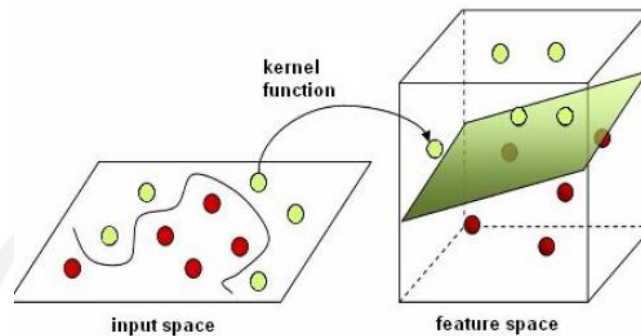
$$f(x) = \operatorname{sign} \left(\sum_{SVs} \alpha_i y_i K(x_i, x) + \hat{b} \right) \quad (2.10)$$

Dengan nilai :

$$w^T \cdot x = \sum_{SVs} \hat{\alpha}_i y_i K(x_i, x) \quad (2.11)$$

$$b = -\frac{1}{2} \sum \bar{\alpha}_i y_i [K(x_r, x) + K(x_s, x)] \quad (2.12)$$

Kemudian pada fungsi kernel yang dapat dipakai pada SVM non linier adalah fungsi kernel linier, *Polynomial*, sigmoid, dan *Radial basic function*. Dengan menggunakan fungsi kernel maka klasifikasi dapat dilakukan.



Gambar 2.4. SVM Non Linier
Sumber : (Zararsiz, Elmali, & Ozturk, 2012)

2.1.10 Association Rule

Association rule merupakan salah satu dari tugas atau teknik yang terdapat dalam *Data mining*. Tugas *Association rule* adalah untuk menemukan atribut yang muncul secara bersamaan atau aturan asosiatif antara suatu kombinasi item. Pada dunia bisnis, teknik ini lebih dikenal dengan analisis afinitas atau analisis keranjang belanja (*market basket analysis*) *Association rule* adalah bentuk jika “kejadian sebelumnya” kemudian “konsekuensinya” (*IF antecedent, THEN consequent*) dengan ukuran *support* dan *confidence* terkait dengan aturan tersebut (Larose & Larose, 2014). *Support* dan *confidence* merupakan bagian dalam pengukuran aturan ketertarikan (*rule interestingness*), karena kedua ukuran tersebut mencerminkan kegunaan dan kepastian aturan yang ditemukan (Han, Kamber, & Pei, 2012). Hal tersebut menjadikan parameter yang digunakan dalam *association rule* (Sheikh, Tanveer, & Hamdani, 2004).

2.1.10.1 Support

Support didefinisikan pada kumpulan item dan memberikan proporsi transaksi yang mengandung *Z* dan oleh karena itu digunakan sebagai ukuran signifikansi (pentingnya) kumpulan *item* (Agrawal, Imielinski, & Swami, 1993). Nilai *support* ini merupakan persentase suatu kombinasi *item* yang muncul pada sekumpulan transaksi atau dalam

database. Perhitungan dalam menentukan nilai *support* dapat didefinisikan seperti pada (2.13)

$$Support = P(A \cap B) = \frac{\text{Jumlah transaksi memuat } A \text{ dan } B}{\text{Total jumlah transaksi}} \quad (2.13)$$

2.1.10.2 Confidence

Confidence didefinisikan sebagai probabilitas untuk melihat aturan *consequent* dengan syarat bahwa transaksi juga mengandung *antecedent* (Agrawal, Imielinski, & Swami, 1993). Nilai *confidence* ini merupakan kuat hubungan antar *item* dalam aturan asosiasi (*association rule*) yang terbentuk. Perhitungan dalam menentukan nilai *confidence* dapat didefinisikan pada (2.14)

$$Confidence = P(B|A) = \frac{Support(A \cap B)}{P(A)} \quad (2.14)$$

2.1.11 Bauran Pemasaran (*Marketing Mix*)

Bauran pemasaran pertama kali dicetuskan oleh Jerome McCarthy. Menurut Jerome McCarthy bauran pemasaran merupakan bagian dari strategi pemasaran yang digunakan untuk merincikan empat faktor keputusan pemasaran yang terdiri dari 4P: *product*, *place*, *price*, dan *promotion* (Damanik & Wibowo, 2018). Bauran pemasaran rangkaian sarana pemasaran taktis terpadu yang dapat dikendalikan (harga, produk, tempat, dan promosi) untuk mengetahui respon pasar sasaran yang diinginkan oleh perusahaan (Machfoedz, 2005). Bauran pemasaran merupakan seperangkat alat pemasaran taktis yang dapat dikendalikan dan dipadukan oleh perusahaan untuk menghasilkan tanggapan yang diinginkan pasar sasaran (Armstrong, Adam, Denize, & Kotler, 2015). Seperangkat alat dalam bauran pemasaran yang dimaksudkan adalah 4P yang digunakan pada bauran pemasaran produk (Marlius, 2016). Sedangkan menurut kotler (2000) dalam (Amilia, 2016) menyampaikan bahwa komponen yang tercakup dalam kegiatan bauran pemasaran (*marketing mix*) pada bidang jasa bertambah yaitu *process*, *people*, dan *physical evidence*. Bauran pemasaran pada bidang jasa tersebut dikenal dengan sebutan 7P.

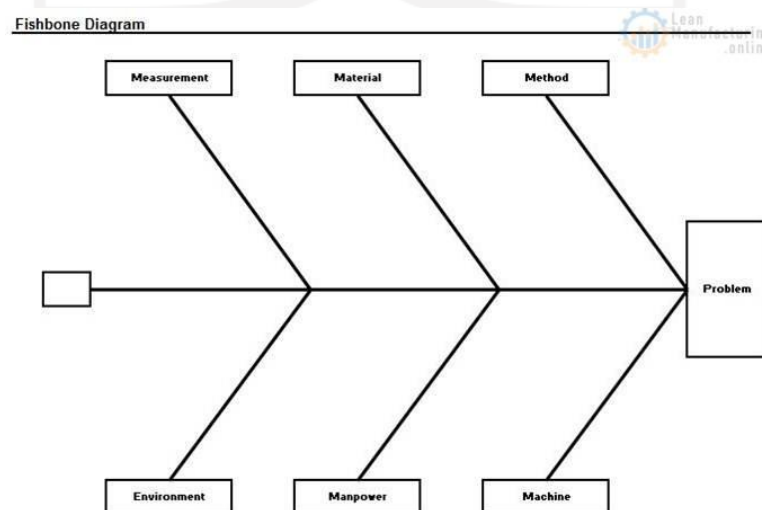
2.1.12 Diagram *Fishbone*

Diagram *fishbone* atau istilah lainnya ishikawa diagram, diagram sebab-akibat, merupakan alat untuk mengidentifikasi akar penyebab dari suatu permasalahan dengan mengelompokkan penyebabnya (ILIE & CIOCOIU, 2010). Nama Ishikawa diagram diambil dari nama orang yang pertama kali memperkenalkan dan menggunakan diagram ini, yaitu Kaoru Ishikawa ditahun 1960 (Juran, 1999). Diagram *fishbone* menunjukkan pada bagian kepala merupakan permasalahan yang terjadi, pada bagian struktur atau tulang ikan menunjukkan penyebab dan sub penyebab dari masalah tersebut.

Pada pembuatan Diagram *fishbone* terdapat beberapa langkah yang perlu dilakukan, yaitu:

1. Identifikasi permasalahan
2. Mencari faktor utama yang terlibat
3. Identifikasi kemungkinan penyebabnya
4. Analisis Diagram

Pada umumnya faktor utama yang terlibat terbagi menjadi beberapa kategori seperti manusia (*people*), mesin (*machine*), material, metode (*method*), pengukuran (*measurement*), dan lingkungan (*environment*). Namun faktor utama yang menjadi penyebab tidak selalu sama (Liliana, 2016).



Gambar 2.5. *Fishbone* Diagram

Sumber : <https://leanmanufacturing.online/the-fishbone-diagram/>

2.2 Kajian Induktif

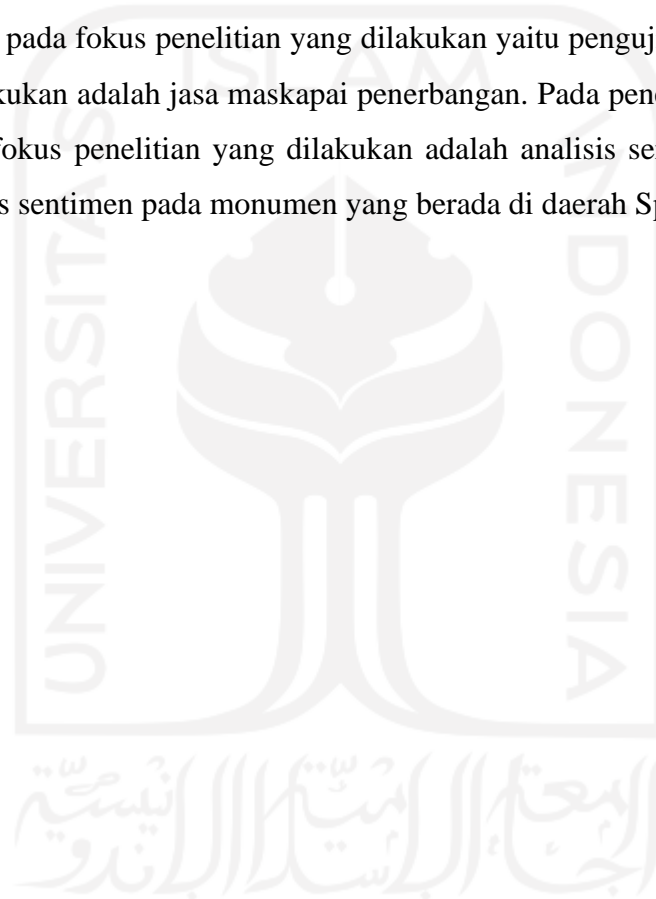
Kajian induktif merupakan hasil penelitian-penelitian yang telah dilakukan sebelumnya, serta perkembangannya dan memiliki kaitan dengan topik yang menjadi penelitian ini. Penelitian-penelitian tersebut dipilih berdasarkan rentang waktu antara tahun 2015-2020 yang diambil melalui media internet. Kata kunci pencarian yang digunakan yaitu analisis sentimen. Kajian induktif akan disajikan pada Tabel 2.1

Tabel 2.1. Kajian Induktif

No.	Judul	Penulis	Tahun	Nama Jurnal	Metode	Hasil Kajian
1.	Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine dan Naive Bayes	Andi Nurul Hidayat	2015	Jurnal Elektronik Sistem Informasi dan Komputer (JESIK)	SVM dan Naive Bayes	Hasil pengujian algoritma terhadap wacana politik pada media masa online diketahui performa dari <i>support vector machine</i> memiliki tingkat akurasi sebesar 90,50%, dan <i>naive bayes</i> dengan tingkat akurasi sebesar 59,98%
2.	Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen <i>Review</i> Jasa Maskapai Penerbangan Menggunakan Naive Bayes	Risa Wati	2016	Jurnal Evolusi	Naive Bayes	Hasil penelitian menunjukkan analisis sentimen yang menggunakan algoritma klasifikasi <i>naive bayes</i> mendapat tingkat akurasi sebesar 60%, dan algoritma <i>naive bayes</i> dengan seleksi fitur algoritma genetik tingkat akurasi yang didapatkan sebesar 89,5%
3.	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter	Ghulam Asrofi Buntoro	2017	Integer Journal		Pada hasil penelitian menunjukkan dari 100 tweet untuk setiap calon, AHY dan Anies mendapat sentimen tertinggi yaitu 62 positif dan Ahok mendapat 33 sentimen positif.

No.	Judul	Penulis	Tahun	Nama Jurnal	Metode	Hasil Kajian
						Kemudian dari hasil klasifikasi setiap algoritma, akurasi tertinggi diperoleh <i>naïve bayes</i>
4.	<i>Sentimen Analysis in Tripadvisor</i>	Ana Valdivia , M. Victoria Luzón , Francisco Herrera	2017	IEEE Intelligent System	SentiStrength , Bing, Syuzhet, dan Core NLP	Hasil penelitian menunjukkan monumen Alhambra, Mezquita Cordoba, dan Sagrada Familia. mendapat sentimen positif dengan nilai diatas 90%
5.	<i>The Evolution of Sentimen Analysis-A Review of Research Topics, Venues, and Top Cited Papers</i>	Mika V. Mantyala, Daniel Graziotin, Miikka Kuutila	2017	Elsevier	<i>Searching Literature</i>	Hasil penelitian menunjukkan dalam beberapa tahun terakhir analisis sentimen mulai bergeser dari menganalisis ulasan produk online ke teks media sosial dari <i>twitter</i> dan <i>facebook</i> . Selain itu jurnal yang paling sering disitasi yang berjudul <i>Opinion mining and sentimen analysis</i> dengan total sitasi 5520 pada google scholar dan 2487 di Scopus
6.	Komparasi Algoritma Support Vector Machine dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur	Deni Gunawan, Dwiza Riana, Dian Ardiansyah, Fajar Akbar, & Salman Alfarizi	2020	Jurnal Teknik Komputer AMIK BSI	SVM + GA dan Naive Bayes + GA	Hasil penelitian menunjukkan bahwa <i>support vector machine</i> memiliki tingkat akurasi yang meningkat dengan penambahan fitur seleksi algoritma genetika yaitu 92,61% menjadi 93,03%. Sedangkan pada <i>naïve bayes</i> tingkat akurasi mengalami penurunan yaitu dari 93,29% menjadi 92,85%

Pada Tabel 2.1 menunjukkan penelitian yang telah dilakukan dan memiliki keterkaitan dengan analisis sentimen. Penelitian dengan pendekatan analisis sentimen telah dilakukan di beberapa bidang atau subjek penelitian. Pada penelitian Hidayat (2015), dan Gunawan, et al (2020) bidang yang dilakukan penelitian adalah pada bidang politik dengan berfokus pada pengujian algoritma klasifikasi pada analisis sentimen. Selain itu, penelitian Wati (2016) memiliki kemiripan pada fokus penelitian yang dilakukan yaitu pengujian algoritma klasifikasi pada analisis sentimen, namun dengan bidang penelitian yang dilakukan adalah jasa maskapai penerbangan. Pada penelitian Buntoro (2017) memiliki kemiripan pada bidang yang diteliti yaitu politik, namun fokus penelitian yang dilakukan adalah analisis sentimen. Pada bidang pariwisata dilakukan oleh Valdivia (2017) yang berfokus pada analisis sentimen pada monumen yang berada di daerah Spanyol



BAB III

METODE PENELITIAN

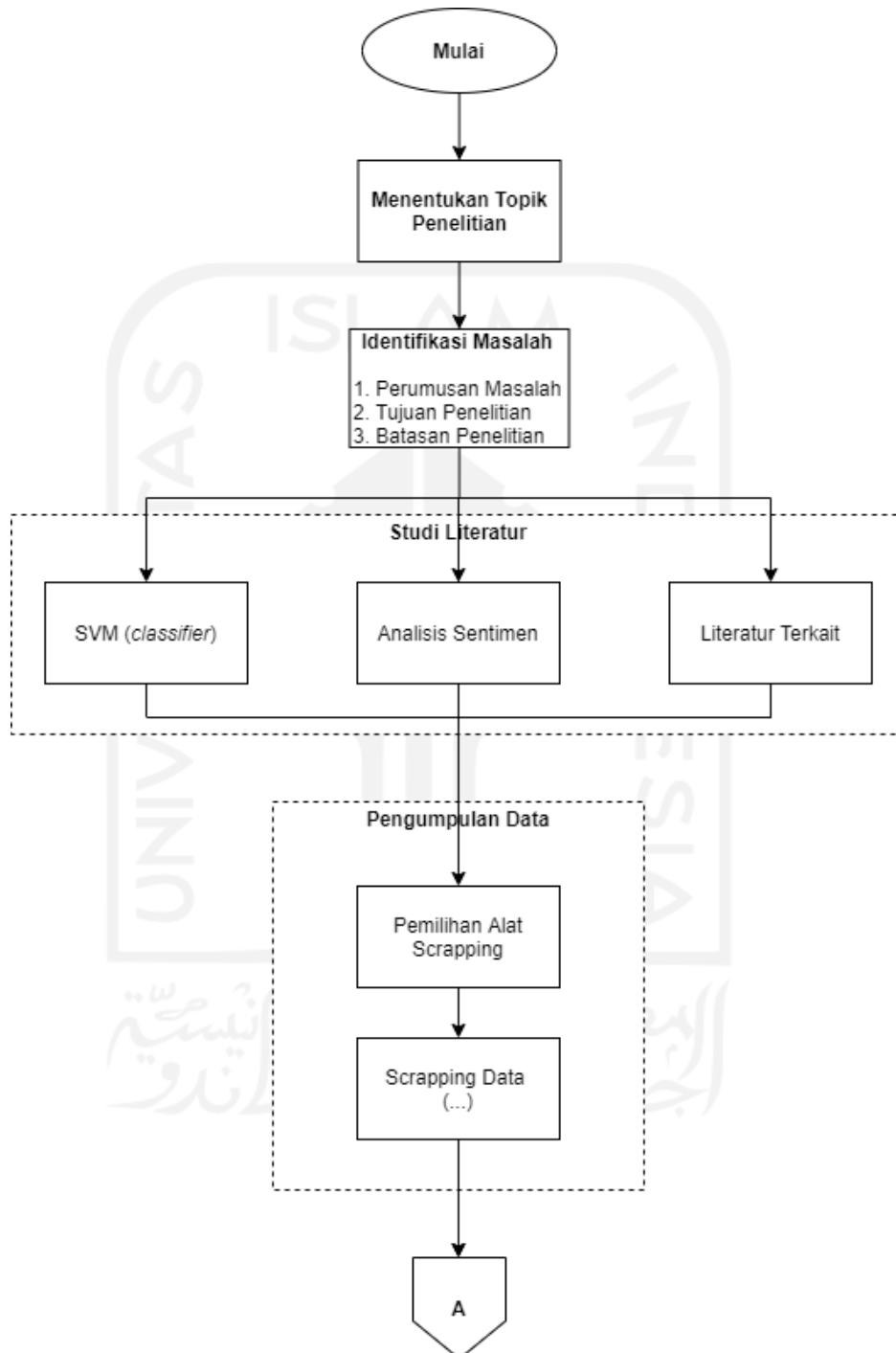
3.1 Objek Penelitian

Objek pada penelitian ini adalah aplikasi kesehatan (*telemedicine*) yang tersedia pada website *google play store*. Aplikasi tersebut terdiri dari Alodokter, Halodoc, Sehatq, Klik Dokter, dan Good doctor.

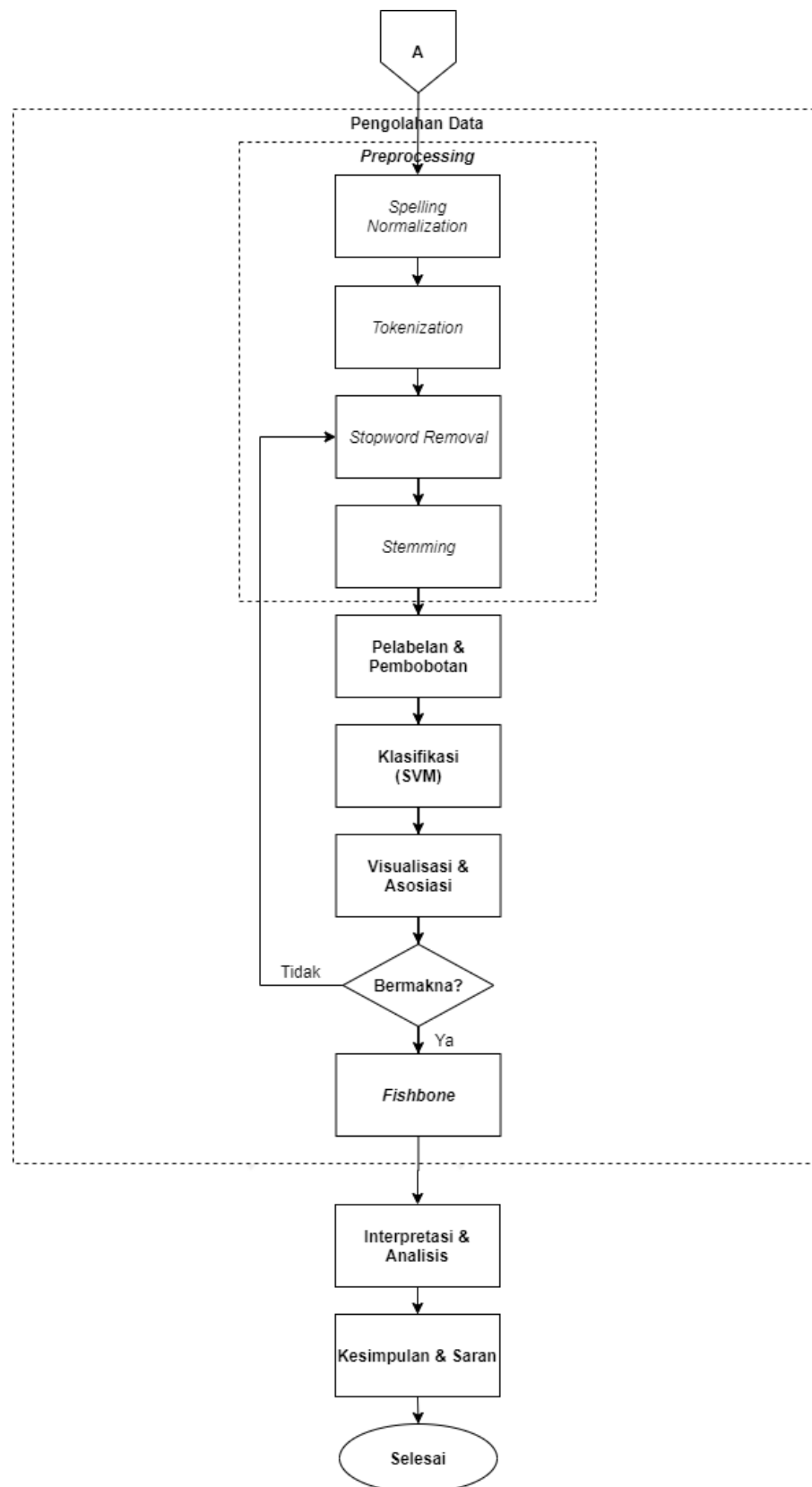
3.2 Metode Pengumpulan Data

Pengumpulan data yang dilakukan pada penelitian ini menggunakan metode *web scraping*. Metode tersebut merupakan proses pengambilan data dari sebuah website di dalam internet. Pengambilan data tersebut dilakukan dengan bantuan dari pihak ketiga yaitu sebuah *website* *appfollow.io*. Website tersebut menyediakan fasilitas *scraping* data

3.3 Diagram Alir Penelitian



Gambar 3.1. Diagram Alir Penelitian



Gambar 3.2. Diagram Alir Penelitian Lanjutan

3.3.1 Menentukan Topik Penelitian

Pada tahap ini, dilakukan penentuan topik penelitian yang ingin dilakukan sebagai fokus dalam penelitian. Penentuan topik penelitian tersebut di dasarkan pada 4 bidang keilmuan yang terdapat dalam Teknik Industri. Bidang keilmuan tersebut terdiri dari *operational research*, sistem produksi, sistem manajemen, dan ergonomi.

3.3.2 Identifikasi Masalah

Pada bagian identifikasi masalah merupakan proses awal dalam mengenalkan permasalahan yang memiliki keterkaitan dengan topik penelitian *operational research* dan berhubungan dengan *big data*.

3.3.3 Studi Literatur

Studi literatur merupakan proses mengetahui dan mempelajari penelitian sebelumnya yang memiliki kesesuaian dengan studi kasus penelitian. Studi literatur terdiri dari landasan teoritis yang memuat teori yang dibutuhkan dalam penelitian dan penelitian yang memiliki kesesuaian dengan studi kasus penelitian. Pada tahap ini, studi literatur digunakan sebagai referensi atau acuan dalam penelitian.

3.3.4 Pengumpulan Data

Pengumpulan data merupakan proses menghimpun atau mengumpulkan data yang dibutuhkan dalam penelitian. Pada bagian ini, metode yang digunakan dalam melakukan pengumpulan data adalah teknik *web scraping* data.

3.3.5 Pengolahan Data

Tahapan pengolahan data diawali dengan melakukan *preprocessing*. Tahapan tersebut merupakan pengolahan awal data yang berupa teks. Tahapan tersebut terdiri dari *spelling normalization*, *tokenization*, *stopword removal*, dan *stemming*.

Setelah data melalui tahap preprocessing, data selanjutnya dilakukan pelabelan & pembobotan. Pelabelan dilakukan untuk memberikan kategori / kelas pada setiap ulasan, sedangkan pembobotan untuk mengubah data tersebut ke dalam bentuk angka. Pengubahan data tersebut bertujuan untuk membantu algoritma klasifikasi membaca data tersebut. Selanjutnya data akan di proses dalam algoritma klasifikasi, yang kemudian dilanjutkan dalam tahapan visualisasi & asosiasi, dan membuat diagram *fishbone*.

3.3.6 Interpretasi dan Analisis

Interpretasi dan analisis merupakan tahapan dalam melakukan identifikasi hasil dari pengolahan data yang telah dilakukan. Pada bagian ini akan dilakukan pembahasan terkait hasil pengolahan data.

3.3.7 Kesimpulan dan Saran

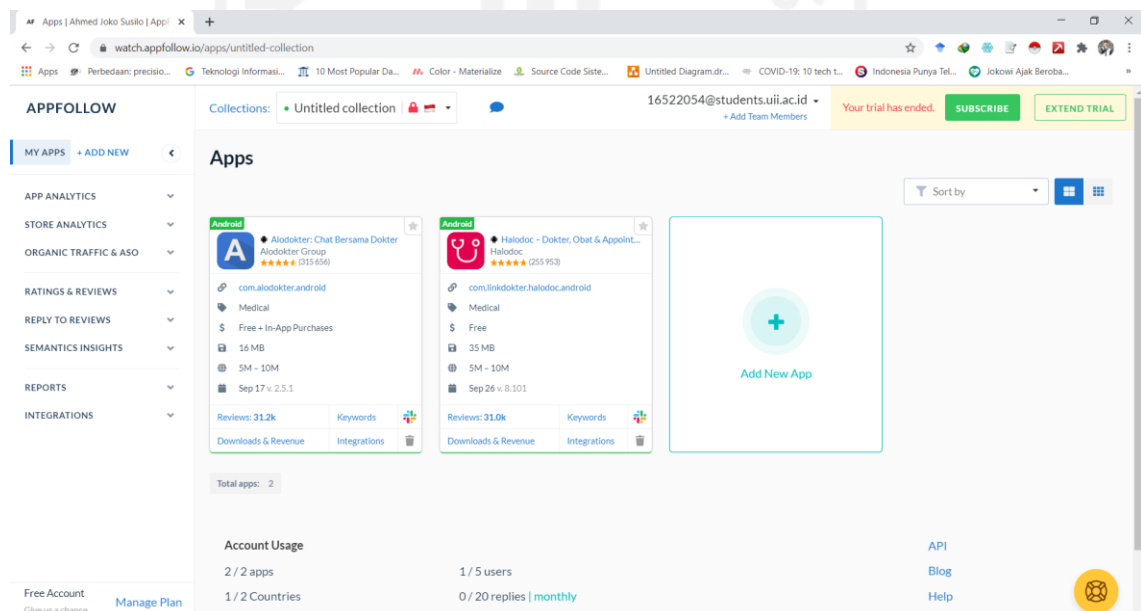
Pada bagian kesimpulan dan saran menguraikan hasil pembahasan pada sub bab sebelumnya dalam uraian singkat untuk menjawab rumusan masalah. Serta memberikan saran terkait penelitian yang telah dilakukan, untuk perkembangan penelitian selanjutnya.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengambilan Data

Data yang digunakan pada penelitian ini adalah data ulasan atau komentar dari pengguna aplikasi kesehatan / *Telemedicine* yang tersedia pada layanan *google play store*. Ulasan tersebut diambil dari 5 aplikasi yaitu, Alodokter, Halodoc, Sehatq, Klikdokter, dan Good Doctor. Proses pengambilan data dilakukan dengan teknik *scraping* yang tersedia pada *website* *appfollow.io*.



Gambar 4.1. Tampilan Muka Appfollow.io

Gambar 4.1 menunjukkan tampilan muka dari website Appfollow.io. Pengambilan data diawali dengan memilih tanda *plus (+)* atau *add new app* pada tampilan muka appfollow.io untuk menambahkan aplikasi yang akan dilakukan *scraping data*. Setelah aplikasi ditambahkan kemudian memilih / klik pada bagian *reviews* di aplikasi yang telah ditambahkan. Sebagai contoh, pada aplikasi alodokter yang ada pada Gambar 4.1 kemudian memilih / klik pada bagian *reviews*. Pengambilan data ulasan dilakukan untuk ulasan pada tanggal 1 maret – 30 juni 2020. Pengambilan tersebut dilakukan untuk 5 aplikasi yang kemudian disimpan dalam bentuk *file excel / csv*. Jumlah total ulasan yang terkumpul sejumlah 19.148 komentar yang terdiri dari Alodokter 1.839 komentar, Halodoc 16.110 komentar, Klikdokter 472 komentar, Sehatq 727 komentar, dan Good doctor 0 komentar. Pada proses pengambilan data, dilakukan secara bertahap untuk setiap aplikasi. Pengambilan data dilakukan dengan memilih waktu perbulan untuk menghindari kegagalan dalam proses *scraping data*. Hal ini karena jumlah data yang terlalu besar untuk di proses jika dilakukan dalam satu waktu untuk keseluruhan ulasan dari bulan Maret - Juni.

Date	AppID	AppName	Language	Version	VersionCo OS	Author	Rating	Title	Review	Translated	Developer	User	Device	Tyt	Tags	Categories	Notes	Likes	Dislikes	Lir
2020-04-1	com.alodc	Alodokter	id			Mawahda	5		dokternya ramah dan cepat responnya											
2020-04-1	com.alodc	Alodokter	id			Eki Vitriani	5		Alodokter emng mantap...											
2020-04-1	com.alodc	Alodokter	id			Emiya Nar	5		Bagus											
2020-04-1	com.alodc	Alodokter	id			Aisyah Ger	5		Sangat puas bisa membantu konsultasi kesehatan dan hasilnya benar. Terima kasih											
2020-04-1	com.alodc	Alodokter	id			risa ristias	5		Respon dokter cepat sekali Terimakasihhh semoga aps ini makin sukses											
2020-04-1	com.alodc	Alodokter	id			Tommy Ar	5		sangat sangat membantu											
2020-04-1	com.alodc	Alodokter	id			Ira Buamb	5		Sangat membantu											
2020-04-1	com.alodc	Alodokter	id			mukmin ki	5		Api ini bagus untuk menanya tentang masalah kesehatan											
2020-04-1	com.alodc	Alodokter	id			Antony Pur	4		Bagus											
2020-04-1	com.alodc	Alodokter	id			REZA RIDZ	5		Jawaban ny lengkap dan detail											
2020-04-1	com.alodc	Alodokter	id			Ijak Sahab	5		Sangat membantu											
2020-04-1	com.alodc	Alodokter	id			RATNA RA	2		Maaf bintangnya di turunkan, Aplikasi nya bagus respon jg cepat, tp selalu ada yg melonop atas nama aldokter dengan mengatakan ttg perindu											
2020-04-1	com.alodc	Alodokter	id			Nadiatul K	5		Aplikasi ini sangat membantu											
2020-04-1	com.alodc	Alodokter	id			Rika sintia	5		Good terbantu											
2020-04-1	com.alodc	Alodokter	id			Anwar Anu	5		Sangat membantu											
2020-04-1	com.alodc	Alodokter	id			Bobby Sim	5		Sangat membantu ...											
2020-04-1	com.alodc	Alodokter	id			Rudi Lubisi	5		Sangat membantu											
2020-04-1	com.alodc	Alodokter	id			Figri Athah	5		Ok											
2020-04-1	com.alodc	Alodokter	id			faiz 1	5		Informasi yg disampaikan cukup jelas. Terimakasih alodokter ????											
2020-04-1	com.alodc	Alodokter	id			moh ali	5		ok											
2020-04-1	com.alodc	Alodokter	id			Kania Octa	2		Kalau bisa yg dokter psikolog coba gratis sekali											
2020-04-1	com.alodc	Alodokter	id			Andri Decis	5		Sungguh sangat membantu sekali...terima kasih alodokter ????											
2020-04-1	com.alodc	Alodokter	id			Andi Prada	5		Respon cepat											
2020-04-1	com.alodc	Alodokter	id			Auval Yusu	4		Mmbantu											

Gambar 4.2. Hasil *Crawling* dengan Appfollow.io

4.2 Analisis Deskriptif

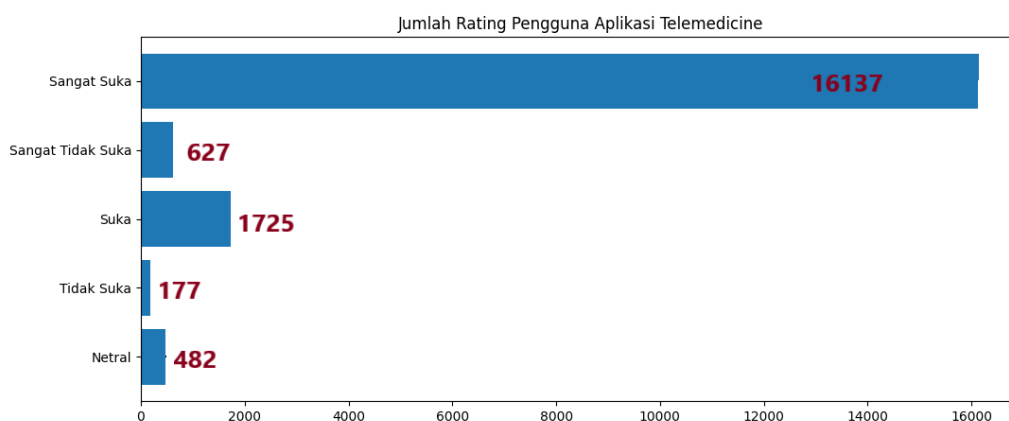
Data yang telah didapatkan, kemudian dilakukan proses *cleaning* / pembersihan. Proses ini adalah menghilangkan data yang tidak dibutuhkan dari hasil *scraping* sebelumnya.

Data yang tetap disimpan atau digunakan adalah data pada kolom *Date* (tanggal), *Rating*, dan *Review*.

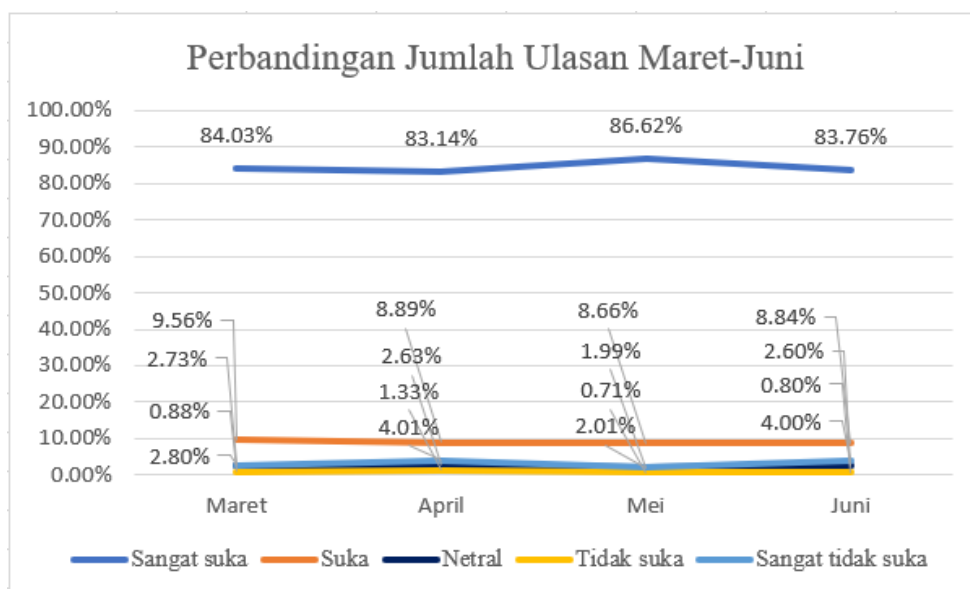
Date	Rating	Review
4/12/2020	5	docternya ramah dan cepat responnya
4/12/2020	5	Alodokter emng mantap...
4/12/2020	5	Bagus
4/12/2020	5	Sangat puas bisa membantu konsultasi kesehatan dan hasilnya benar. Terima kasih
4/12/2020	5	Respond dokter cepat sekali Terimakasihhh semoga aps ini makin sukses
4/12/2020	5	sangat sangat membantu
4/12/2020	5	Sangat membantu
4/12/2020	5	Apl ini bagus untuk menanya tentang masalah kesehatan
4/12/2020	4	Bagus
4/12/2020	5	Jawaban ny lengkap dan detail
4/12/2020	5	Sangat membantu
4/12/2020	2	Maaf bintangnya di turunkan, Aplikasi nya bagus respon jg cepat, tp selalu ada yg melonjol atas nama alodokter dengan mengatakan ttg perlindungan dan asuransi kesehatan.Minta data diri jg dan rekening bank. Makasi yah alodokter, . . .
4/12/2020	5	Aplikasi ini sangat membantu
4/12/2020	5	Good terbanutu
4/12/2020	5	Sangat membantu
4/12/2020	5	Sangattt membantu ...
4/12/2020	5	Sangat membantu
4/12/2020	5	Ok
4/12/2020	5	Informasi yg disampaikan cukup jelas. Terimakasih alodokter ????
4/12/2020	5	ok
4/12/2020	2	Kalau bisa yg dokter psikolog coba gratis sekali
4/12/2020	5	Sungguh sangat membantu sekali, terima kasih alodokter ????
4/12/2020	5	Respon cepat
4/12/2020	4	Membantu

Gambar 4.3. Hasil *Cleaning Data 1*

Setelah melakukan *Cleaning data 1*, kemudian data dibuatkan visualisasi dengan tujuan memudahkan dalam melakukan analisis deskriptif. Visualisasi yang dilakukan adalah mengubah data tersebut ke dalam bentuk diagram batang. Visualisasi ini bertujuan untuk mengetahui gambaran umum informasi dari ulasan yang diberikan oleh pengguna aplikasi kesehatan / *telemedicine* yang terdapat pada *google play*. Visualisasi yang dilakukan dengan menunjukkan aspek jumlah ulasan yang telah diberikan oleh pengguna berdasarkan urutan waktu, dan *rating* yang diberikan oleh pengguna.



Gambar 4.4. Jumlah Ulasan Pengguna Aplikasi *Telemedicine Google Play*



Gambar 4.5. Perbandingan Jumlah Ulasan Bulan Maret-Juni

4.3 Sentimen

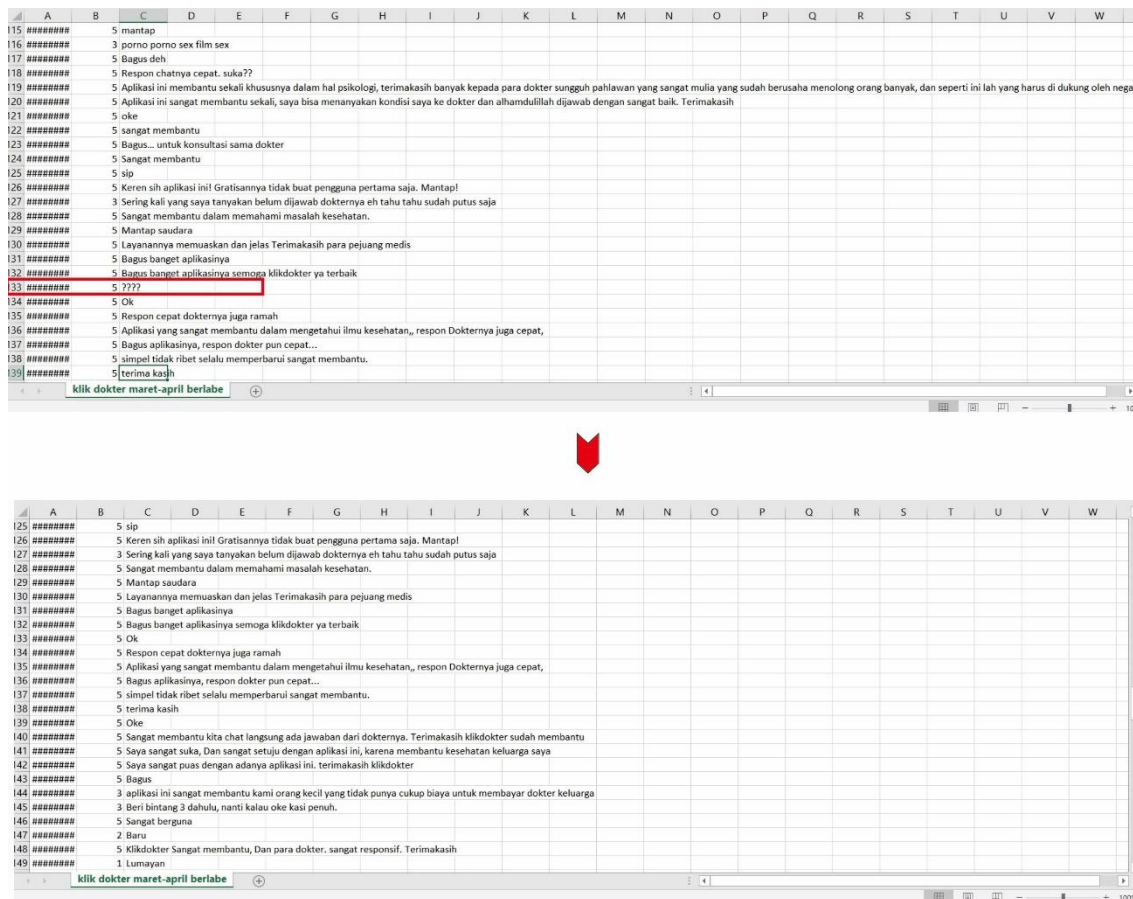
Data ulasan yang tersedia pada situs *google play store*, merupakan data *text*. Data tersebut masih memiliki kekurangan yaitu tidak terstruktur seperti memiliki angka, tanda baca, *emoticon*, dan kata-kata yang tidak memiliki makna. Oleh karena itu, sebelum dilakukan klasifikasi dokumen *text* yang sesuai dengan kelas sentimen perlu dilakukan tahapan *preprocessing*. Tahapan ini bertujuan untuk menyeragamkan bentuk kata atau *text* yang terdapat dalam setiap ulasan pengguna dan menghilangkan karakter-karakter selain huruf.

Pada tahapan *preprocessing*, terdapat beberapa tahapan yang dapat dilakukan. Pada penelitian ini tahapan yang dilakukan adalah data *cleaning & case folding*, *spell normalization*, *tokenizing*, *stop word removal*, dan *stemming*

4.3.1 Data Cleaning dan Case Folding

Tahapan *cleaning* data bertujuan untuk menghilangkan data yang tidak memiliki *text* dalam ulasannya, dan membersihkan ulasan / komentar dari hal yang tidak dibutuhkan. Ulasan yang harus dibersihkan adalah dokumen *text* yang memiliki tanda baca, angka, dan *emoticon*. Pada tahap *case folding* setiap huruf dalam dokumen *text* diubah ke bentuk

huruf kecil. Pada Gambar 4.6 menunjukkan proses menghilangkan data yang tidak memiliki *text*



Gambar 4.6. Proses Menghilangkan Data

Setelah proses menghilangkan data, jumlah data yang akan dilanjutkan pada proses berikutnya menjadi tereduksi. Jumlah data yang akan digunakan adalah 18.913 ulasan. Ulasan / komentar yang masih memiliki angka, tanda baca, dan *emoticon*, kemudian dibersihkan. Proses membersihkan dibantu dengan *library* yang terdapat dalam bahasa pemrograman *python*. Pada *library* tersebut juga disediakan *library* untuk mengubah huruf besar menjadi huruf kecil. proses tersebut dapat diketahui pada Tabel 4.1.

Tabel 4.1. Proses Membersihkan Ulasan

No.	Sebelum	Sesudah
1.	Sangat membantu sekali alodokter ini	sangat membantu sekali alodokter ini
2.	Sangat sangat membantu.. Saya suka alodokter	sangat sangat membantu saya suka alodokter

No.	Sebelum	Sesudah
3.	jawabannya sangat membantu.. makasi dok	jawabannya sangat membantu makasi dok
4.	bagus sangat membantu aplikasi ini.....	bagus sangat membantu aplikasi ini
5.	Sangat membantu????	sangat membantu
6.	puas sekali dengan layanan aplikasinya ???? PENJUAL DATA... tidak HERAN BANYAK BINTANG 1... nanti KALIAN BAKAL DI teleponin DI TAWARIN produk-produk	puas sekali dengan layanan aplikasinya penjual data tidak heran banyak bintang 1 nanti kalian bakal di teleponin di tawarin produk produk

4.3.2 Spelling Normalization

Setelah data melalui proses *cleaning*, ulasan dalam setiap dokumen *text* dilakukan tahap *spelling normalization*. pada tahap ini dilakukan untuk mendapat kualitas dokumen *text* yang baik. Hal tersebut karena kata yang terdapat pada dokumen *text* akan diperbaiki ejaannya ataupun memperbaiki kata yang disingkat. Contoh pada tahap ini dapat dilihat pada Tabel 4.2. Proses ini dilakukan dengan bantuan software *microsoft excel* dan bahasa pemrograman *python*.

Tabel 4.2. Proses *Spelling Normalization*

No.	Sebelum	Sesudah
1.	dokternya ramh dan cepat responnya	dokternya ramah dan cepat responnya
2.	app bgus sya suka tapi sayang hanya bebrapa alodok yg gratis	aplikasi bagus saya suka tapi sayang hanya beberapa alodokter yang gratis
3.	dijawab dgn baik	dijawab dengan baik
4.	sangat membantu di saat qt jauh dari klinik atw pun rmh sakit	sangat membantu di saat kita jauh dari klinik atau pun rumah sakit
5.	cepat banget balesnya penjelasan sangat panjang dan detail thx ada yg konsul gratis dan berbayar	cepat banget balasnya penjelasan sangat panjang dan detail terimakasih ada yang konsultasi gratis dan berbayar
6.	trmksh krn sya dpt bnyak masukan	terimakasih karena saya dapat banyak masukan
7.	Sy suka apliksinya dokter ny ramahh baik sopan dn cpat tanggap	saya suka aplikasinya dokter nya ramah baik sopan dan cepat tanggap

4.3.3 Tokenizing

Tokenizing adalah proses untuk memisahkan kata yang terdapat dalam dokumen *text* menjadi kata yang independen atau tidak saling berhubungan (Praptiwi, 2018). Potongan kata atau kata yang sudah terpisah disebut sebagai token. Pemisahan kata ini akan dipisahkan dengan tanda koma (,) untuk setiap katanya. Contoh *tokenizing* dapat dilihat pada Tabel 4.3.

Tabel 4.3. Proses Token

No.	Sebelum	Sesudah
1.	dokternya ramah dan cepat responnya	dokternya, ramah, dan, cepat, responnya
2.	aplikasi bagus saya suka tapi sayang hanya beberapa alodokter yang gratis	aplikasi, bagus, saya, suka, tapi, sayang, hanya, beberapa, alodokter, yang, gratis
3.	dijawab dengan baik	dijawab, dengan, baik
4.	sangat membantu di saat kita jauh dari klinik atau pun rumah sakit	sangat, membantu, di, saat, kita, jauh, dari, klinik, atau, pun, rumah, sakit
5.	cepat banget balasnya penjelasan sangat panjang dan detail	cepat, banget, balasnya, penjelasan, sangat, panjang, dan, detail,
6.	terimakasih ada yang konsultasi gratis dan membayar	terimakasih, ada, yang, konsultasi, gratis, dan, membayar
7.	terimakasih karena saya dapat banyak masukan saya suka aplikasinya dokter nya ramah baik sopan dan cepat tanggap	terimakasih, karena, saya, dapat, banyak, masukan saya, suka, aplikasinya, dokter, nya, ramah, baik, sopan, dan, cepat, tanggap

4.3.4 Stop Word Removal

Pada tahap *stop word removal* bertujuan untuk menghilangkan kata-kata yang tidak dibutuhkan dan tidak memberi pengaruh terhadap sentimen. Proses untuk menghilangkan kata-kata tersebut dilakukan dengan *library* sastrawi, sebuah *library* yang tersedia dalam bahasa pemrograman *python*. *Library* tersebut memiliki kumpulan kata-kata *stop word* dalam bahasa indonesia yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).

Setelah data melalui proses tokenisasi (*tokenizing*) kemudian data dapat dilakukan proses *stop word removal*. Hal ini karena data yang digunakan terdiri dari kumpulan dokumen *text* (ulasan per baris) yang besar. Sehingga akan memudahkan *library* sastrawi

dalam memeriksa setiap kata karena sudah dalam bentuk token. Contoh proses *stop word removal* dapat dilihat pada Tabel 4.4.

Tabel 4.4. Proses *Stopword Removal*

No.	Sebelum	Sesudah
1.	dokternya, ramah, dan , cepat, responnya	dokternya, ramah, cepat, responnya
2.	aplikasi, bagus, saya , suka, tapi , sayang, hanya , beberapa, alodokter, yang , gratis	aplikasi, bagus, suka, sayang, beberapa, alodokter, gratis
3.	dijawab, dengan , baik	dijawab, baik
4.	sangat, membantu, di , saat , kita , jauh, dari , klinik, atau , pun , rumah, sakit	sangat, membantu, jauh, klinik, rumah, sakit
5.	cepat, banget, balasnya, penjelasan, sangat, panjang, dan , detail, terimakasih, ada , yang , konsultasi, gratis, dan , membayar	cepat, banget, balasnya, penjelasan, sangat, panjang, detail, terimakasih, konsultasi, gratis, membayar
6.	terimakasih, karena , saya , dapat , banyak, masukan	terimakasih, banyak, masukan
7.	saya , suka, aplikasinya, dokter, nya, ramah, baik, sopan, dan , cepat, tanggap	suka, aplikasinya, dokter, nya, ramah, baik, sopan, cepat, tanggap

4.3.5 Stemming

Tahap *stemming* merupakan proses untuk mengubah bentuk kata majemuk atau kata yang memiliki imbuhan menjadi bentuk kata dasar. Pada penelitian ini penggunaan *stemming* tidak dilakukan pada keseluruhan kata, namun penggunaannya pada saat melakukan pelabelan. Fungsi ini diintegrasikan dalam *syntax / code* dalam pelabelan. Jika suatu kata dalam ulasan tidak dikenali oleh kamus maka kata tersebut akan diubah ke dalam bentuk kata dasar yang kemudian dicocokkan kembali pada kamus. Contoh proses ini dapat dilihat pada Tabel 4.5.

Tabel 4.5. Contoh Proses *Stemmer*

No.	Sebelum	Sesudah
1.	dokternya , ramah, cepat, responnya	dokter , ramah, cepat, responnya
2.	aplikasi, bagus, suka, sayang, beberapa, alodokter, gratis	aplikasi, bagus, suka, sayang, beberapa, alodokter, gratis
3.	dijawab , baik	jawab , baik

No.	Sebelum	Sesudah
4.	sangat, membantu , jauh, klinik, rumah, sakit	sangat, bantu , jauh, klinik, rumah, sakit
5.	cepat, banget, balasnya , penjelasan , sangat, panjang, detail, terimakasih, konsultasi, gratis, membayar	cepat, banget, balas , jelas , sangat, panjang, detail, terimakasih, konsultasi, gratis, membayar
6.	terimakasih, banyak, masukan	terimakasih, banyak, masuk
7.	suka, aplikasinya , dokter, nya, ramah, baik, sopan, cepat, tanggap	suka, aplikasi , dokter, nya, ramah, baik, sopan, cepat, tanggap

4.3.6 Pelabelan

Pelabelan bertujuan untuk membantu dalam proses klasifikasi. Pada prosesnya, klasifikasi yang dilakukan termasuk dalam *supervised learning*. Sehingga data ulasan harus diberikan label terlebih dahulu sebelum masuk dalam tahap klasifikasi. Proses pelabelan yang dilakukan melalui 3 tahapan. Tahap pertama adalah melakukan pelabelan dengan mengubah rating yang telah diberikan oleh pengguna menjadi sentimen positif, netral dan negatif. Perubahan tersebut dengan berdasarkan rating yang diberikan, jika rating yang diberikan bernilai 4 dan 5 maka ulasan tersebut diberi label positif. Ulasan yang diberi label 1 dan 2 maka akan diberi label negatif. Pada ulasan yang memiliki nilai rating 3 akan diberi label netral. Proses tersebut dapat dilihat pada Gambar 4.7.

text	rating	pelabelan_rating
dokternya ramah dan respon cepatnya	5	positif
alodokter emang mantap...	5	positif
Bagus	5	positif
Sangat puas bisa membantu konsultasi kesehatan dan hasilnya benar. Terimakasih	5	positif
Respon dokter cepat sekali Terimakasih semoga aplikasi ini makin sukses	5	positif
sangat sangat membantu	5	positif
Sangat membantu	5	positif
aplikasi ini bagus untuk menanya tentang masalah kesehatan	5	positif
Bagus	4	positif
Jawaban nya lengkap dan detail	5	positif
Sangat membantu	5	positif

Gambar 4.7. Proses Tahap Pertama Pelabelan Dengan Rating

Tahap kedua dalam pelabelan adalah dengan menggunakan kamus *lexicon* untuk melakukan pelabelan. Pada kamus *lexicon* terdapat bobot kata yang menunjukkan kata

tersebut termasuk dalam kategori positif atau negatif, dengan melihat nilai pada bobot tersebut (Koto & Rahamaningtyas, 2017). jika bobot kata bernilai positif (+) maka kata termasuk dalam kategori sentimen positif dan jika bernilai (-) maka kata tersebut dikategorikan sebagai kata sentimen negatif Perhitungan untuk skor sentimen adalah sebagai berikut :

$$\text{Skor} = (\text{Jumlah Bobot Kata Positif}) + (\text{Jumlah bobot Kata Negatif}) \quad (4.1)$$

Jika sebuah ulasan memiliki bobot lebih besar dari nilai 0 maka ulasan tersebut dikategorikan sebagai positif dan sebaliknya . Proses ini dapat dilihat pada Tabel 4.6, dan jumlah untuk hasil pelabelan untuk setiap kategori dengan berdasarkan *rating* dan kamus *lexicon* dapat dilihat pada Tabel 4.7.

Tabel 4.6. Contoh Proses Pelabelan Kamus *Lexicon*

Ulasan	Kata Positif (bobot)	Kata Negatif (bobot)	Skor
Aplikasi ini bagus untuk menanya tentang masalah kesehatan	Bagus (2) Sehat(4)	Aplikasi (-4) Masalah (-4)	-2
Respon dokter cepat	Respon (2) Cepat (3)	-	5

Pada Tabel 4.6 menunjukkan proses pelabelan yang dilakukan dengan kamus *lexicon*. Sebagai contoh pada ulasan “Aplikasi ini bagus untuk menanya tentang masalah kesehatan” yang terdapat pada Tabel 4.6. Pada ulasan tersebut terdapat empat kata yang terdapat dalam kamus *lexicon* yaitu kata ‘bagus’, ‘sehat’, ‘aplikasi’, dan ‘masalah’. Kata ‘bagus’, dan ‘sehat’ memiliki bobot masing-masing 2 dan 4, bobot tersebut terkelompok dalam kata positif yang berada di kamus *lexicon*. Hal tersebut dapat dilihat pada nilainya yang bernilai positif (+). Pada kata ‘aplikasi’, dan ‘masalah’ memiliki bobot sama yaitu negatif 4 (-4). Kata tersebut terkelompok dalam kata negatif yang berada dalam kamus *lexicon* dengan melihat nilai dari angka tersebut yang bernilai negatif (-). Bobot dari setiap kelompok kata dilakukan penjumlahan. Pada kata positif berjumlah enam (6), pada kata negatif berjumlah negatif 8 (-8), dan kemudian dijumlah total antara kata positif dan negatif dengan nilai akhir atau skor berjumlah negatif 2 (-2).

Pada prosesnya, pelabelan menggunakan kamus lexicon sangat bergantung dengan bobot kata yang tersedia. Menentukan sebuah kalimat termasuk dalam sentimen positif atau negatif dilakukan dengan menganalisis setiap kata yang terdapat dalam kalimat tersebut dengan melihat bobot dari setiap kata. Sehingga pada kalimat yang memiliki sifat ambigu atau multipolar belum dapat mengidentifikasi dengan tepat untuk kelas sentimen positif, negatif, dan netral.

Tabel 4.7. Hasil Pelabelan Dengan Rating dan Kamus *Lexicon*

Sentimen	Rating	Kamus <i>Lexicon</i>
Positif	17.645	17.099
Negatif	792	1.201
Netral	476	613
Total	18.913	18.913

Pada Tabel 4.7 menunjukkan hasil pelabelan yang dilakukan dengan dua pendekatan yaitu *rating* dan kamus *lexicon*. Kolom *rating* menunjukkan hasil pelabelan yang telah dilakukan dengan ulasan yang berlabel positif berjumlah 17.645, ulasan negatif berjumlah 792, dan ulasan netral berjumlah 18.913. Pada kolom kamus *lexicon* ulasan positif berjumlah 17.099, ulasan negatif berjumlah 1.201, dan ulasan netral berjumlah 613.

Tahap ketiga ini berfokus pada ulasan yang berkategori netral yang ada pada pelabelan *rating*, karena pada penelitian ini bertujuan dan berfokus pada sentimen positif dan negatif. Untuk mengetahui ulasan tersebut termasuk dalam kategori positif atau negatif maka dilakukan perubahan label dengan didasarkan hasil label kamus *lexicon*. Namun Selama proses pelabelan terdapat ulasan yang tetap berkategori netral atau memiliki nilai 0, Sehingga untuk ulasan yang tetap berkategori netral tidak akan digunakan pada tahap selanjutnya. Jumlah total untuk perubahan label yang telah dilakukan dapat dilihat pada Tabel 4.8.

Tabel 4.8. Hasil Perubahan Label

Sentimen	Ulasan
Positif	18.009
Negatif	874
Total	18.883

4.4 Klasifikasi

4.4.1 Pembagian Data Latih dan Data Uji

Setelah data diberikan label, kemudian dilanjutkan dengan membagi data menjadi data latih (*training*) dan data uji (*testing*). Hal ini dibutuhkan karena dalam algoritma klasifikasi yang digunakan merupakan *supervised learning*. Sehingga dibutuhkan sebuah data yang digunakan untuk membuat model pembelajaran dan selanjutnya akan dievaluasi oleh data uji. Pada penelitian ini pembagian data yang digunakan berdasarkan pada beberapa skenario yang digunakan yaitu data dibagi menjadi 80% : 20%, 75% : 25%, 70% : 30%, dan 65% : 35%. Sehingga jumlah untuk data latih yang digunakan seperti pada Tabel 4.9

Tabel 4.9. Pembagian Data

Skenario	Pembagian Data		Total
	Data Latih	Data Uji	
Skenario 1 (80 : 20)	15.106	3.777	18.883
Skenario 2 (75 : 25)	14.162	4.721	18.883
Skenario 3 (70 : 30)	13.218	5.665	18.883
Skenario 4 (65 : 35)	12.273	6.610	18.883

4.4.2 Klasifikasi *Support Vector Machine*

Pada tahap klasifikasi algoritma yang digunakan adalah *support vector machine* (SVM) dengan jenis set data yang berbeda, yaitu linier dan non-linier. Klasifikasi untuk non-linier menggunakan *kernel linier*, *polynomial*, *sigmoid* dan *radial basis function* (RBF). Berikut Gambar 4.8 merupakan *confusion matrix* dari algoritma SVM:

CONFUSION MATRIX

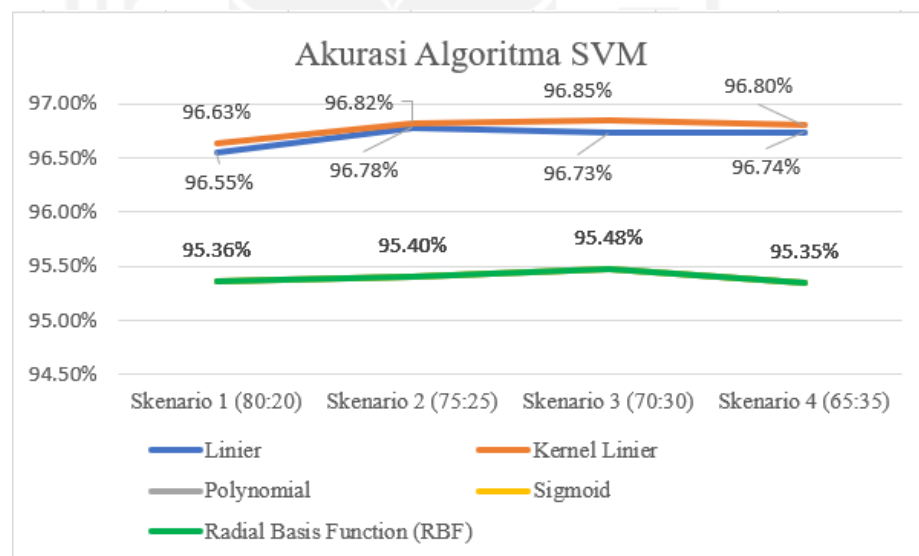
Kernel Linier	Positif	Negatif	Polynomial	Positif	Negatif
Positif	118	138	Positif	0	256
Negatif	40	5369	Negatif	0	5409

Sigmoid	Positif	Negatif	RBF	Positif	Negatif
Positif	0	256	Positif	0	256
Negatif	0	5409	Negatif	0	5409

Linier	Positif	Negatif
Positif	127	129
Negatif	56	5353

Gambar 4.8. *Confusion Matrix* Skenario 3

Hasil dari Gambar 4.8 dapat dilanjutkan untuk mengetahui akurasi dari setiap algoritma. Berikut ini adalah hasil perbandingan akurasi klasifikasi dari algoritma SVM untuk setiap skenario pada Gambar 4.9



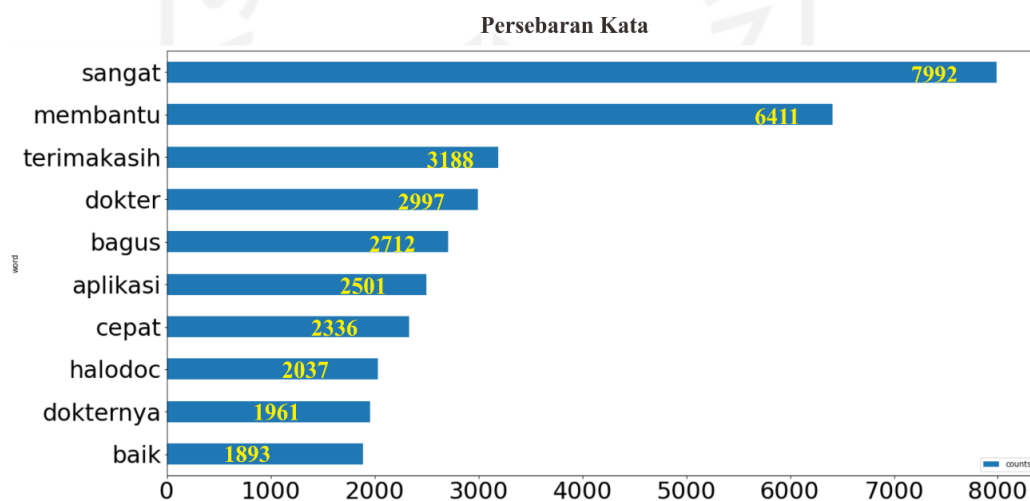
Gambar 4.9. Akurasi Algoritma SVM

4.4.3 Visualisasi Data dan Asosiasi

Pada tahap ini hasil klasifikasi yang telah dilakukan ditahap sebelumnya kemudian dilakukan visualisasi dengan menggunakan diagram batang dan *library word cloud*. Visualisasi ini untuk memberikan tampilan yang menarik dan memudahkan dalam

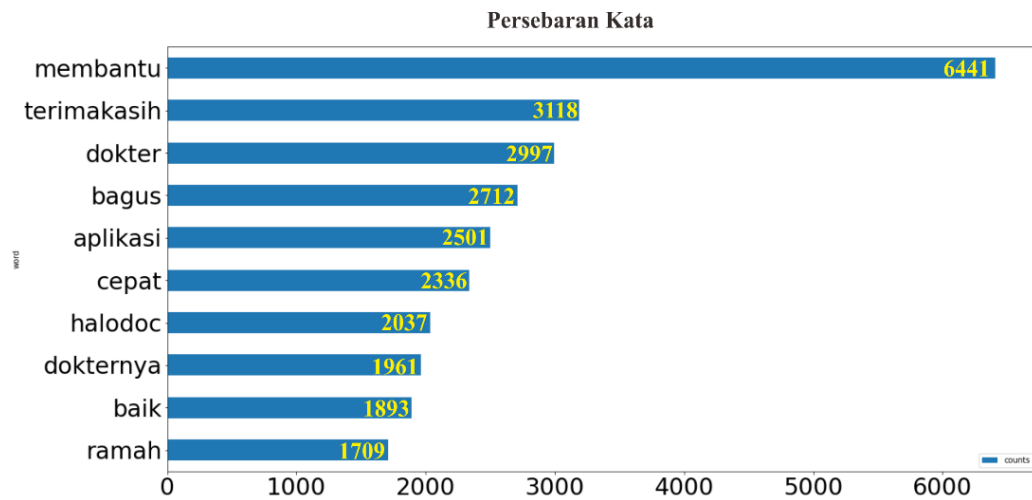
melakukan analisis. Tahap ini dapat memberikan gambaran kata yang lebih detail tentang sentimen dari pengguna aplikasi. Kemudian dilakukan tahapan asosiasi pada sentimen negatif untuk membantu dalam menunjukkan kata yang memiliki frekuensi muncul secara bersamaan.

Hal pertama yang dilakukan adalah dengan menunjukkan persebaran kata yang paling sering muncul dengan menggunakan diagram batang. Persebaran yang ditunjukkan adalah jumlah 10 kata tertinggi yang paling sering muncul. Berikut visualisasi untuk persebaran keseluruhan kata yang paling sering muncul dalam ulasan ada pada Gambar 4.10.



Gambar 4.10. Persebaran Seluruh Kata

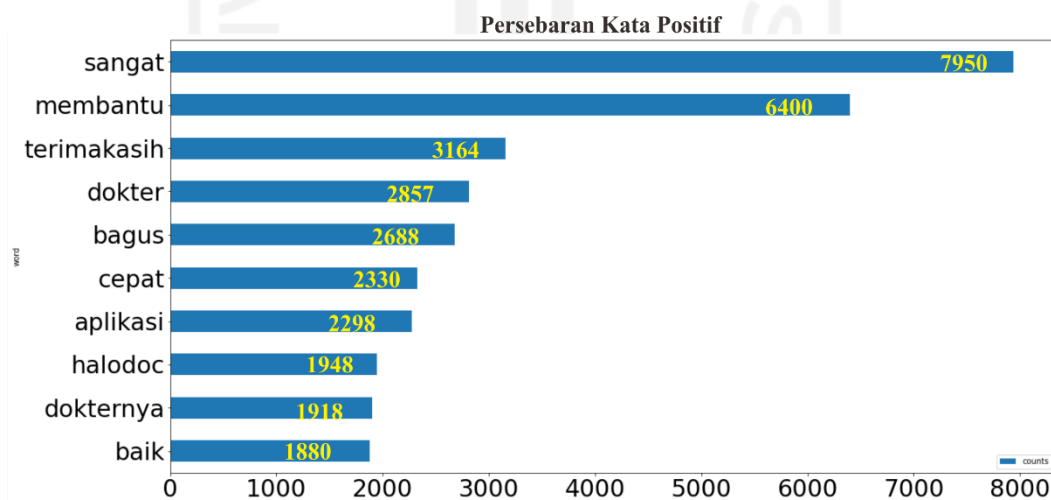
Pada gambar 4.10. masih terdapat kata yang belum memiliki makna sehingga dilakukan *stopword removal* tahap II. Sebagai contoh adalah kata ‘sangat’, sehingga dilakukan proses *stopword* kembali. Berikut merupakan gambar hasil *stopword removal* tahap II



Gambar 4.11. Persebaran Seluruh Kata (Stopword Removal II)

Selanjutnya adalah membuat visualisasi untuk persebaran kata yang paling sering muncul setiap kelas sentimen dengan menggunakan diagram batang dan melakukan *association rule* pada kata kelas sentimen negatif yang memiliki kemunculan secara bersamaan.

4.4.3.1 Sentimen Positif



Gambar 4.12. Persebaran Kata Kelas Sentimen Positif

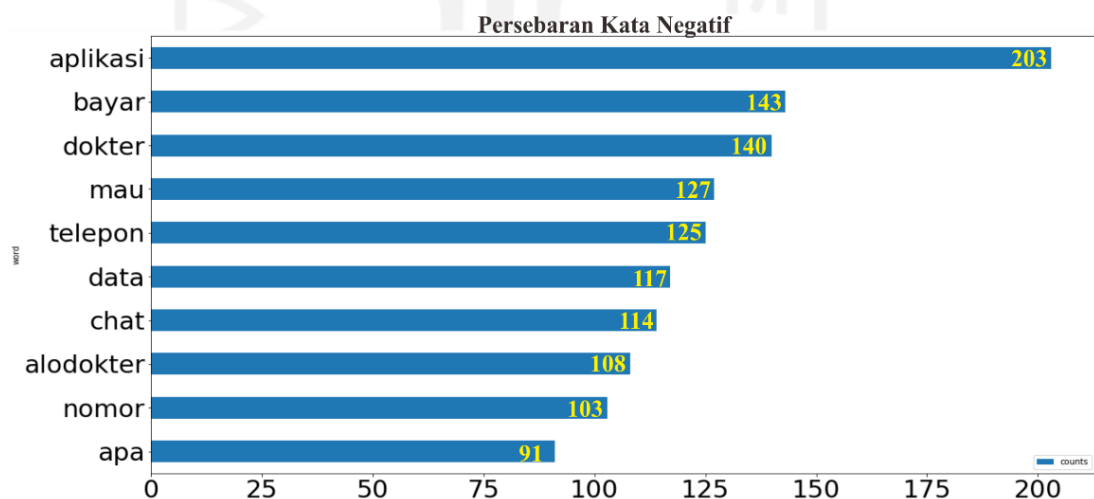
Gambar 4.11. merupakan hasil persebaran kata kelas sentimen positif. Setelah dilakukan visualisasi dengan diagram batang, visualisasi yang selanjutnya dilakukan adalah dengan menggunakan *word cloud* untuk data sentimen positif. Berikut Gambar 4.12. merupakan *word cloud* untuk kelas sentimen positif.



Gambar 4.15. *Word cloud* Kelas Positif (Stopword Removal II)

4.4.3.2 Sentimen Negatif

Setelah melakukan visualisasi terhadap kelas sentimen positif, selanjutnya dilakukan visualisasi terhadap kelas sentimen negatif, dan melakukan tahapan *association rule*. Berikut Gambar 4.15. merupakan 10 kata paling banyak dalam ulasan negatif.



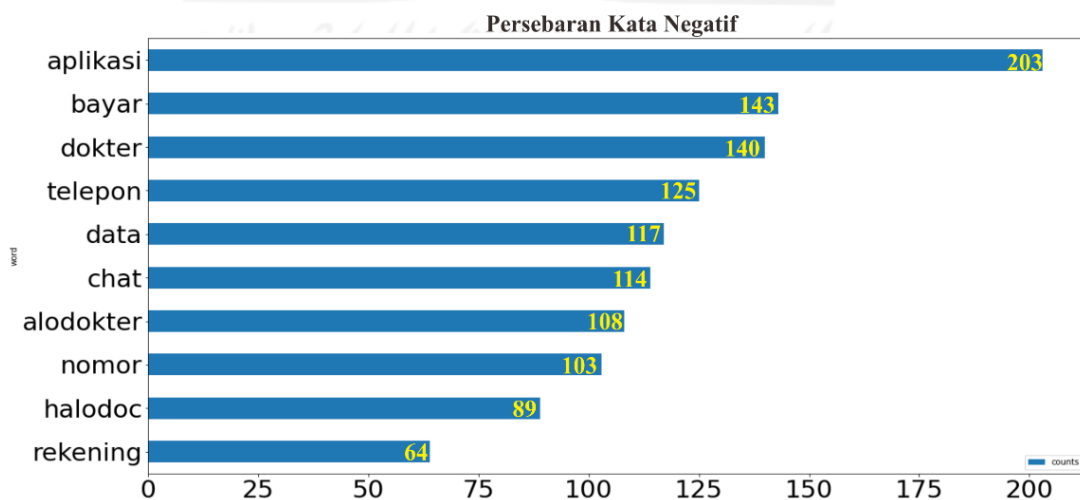
Gambar 4.16. Persebaran Kata Kelas Sentimen Negatif

Setelah menampilkan visualisasi dengan diagram batang, selanjutnya dilakukan visualisasi dengan *word cloud*. Berikut Gambar 4.16 merupakan visualisasi yang telah dilakukan dengan *word cloud*.



Gambar 4.17 *Word cloud* Kelas Sentimen Negatif

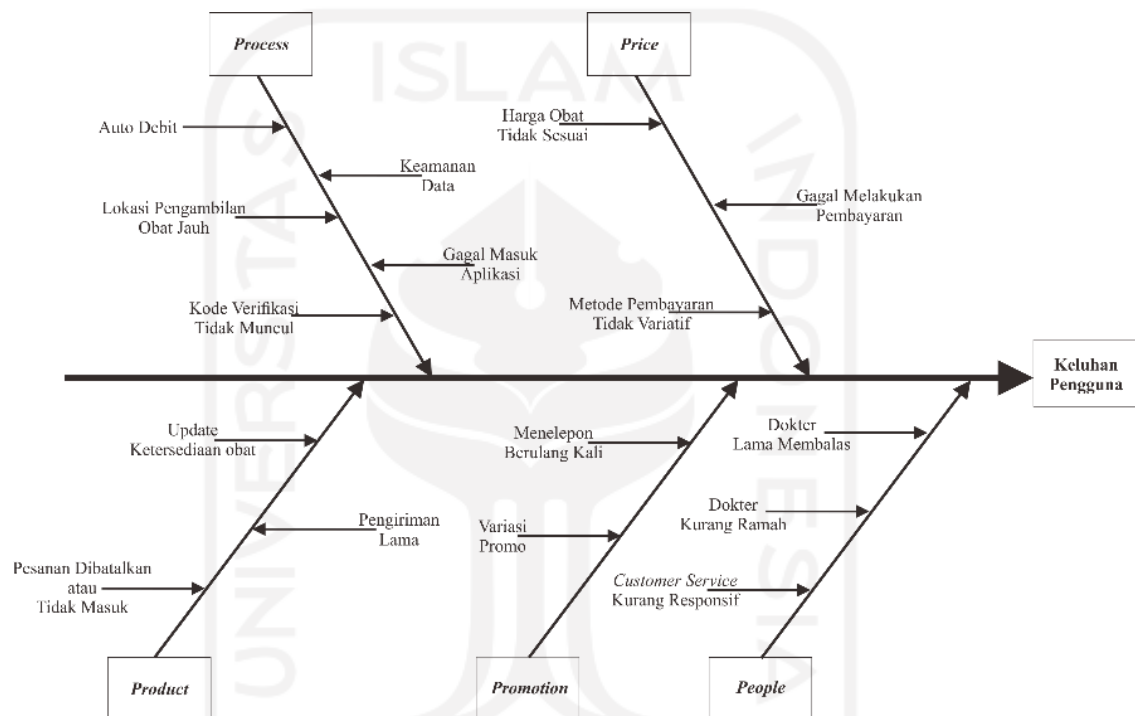
Hasil yang telah ditunjukkan pada Gambar 4.15 dan 4.16 terdapat kata-kata yang belum memiliki arti / makna terhadap penelitian. Sehingga perlu dilakukan penyaringan atau *filter*. Hal ini dapat dilakukan dengan menerapkan *stopword removal* Tahap II. Pada ulasan negatif kata-kata yang akan dilakukan penyaringan adalah kata “sangat”, “kelebihan”, “mau”, “kalau”, “sama”, “apa”, “nya”, dan “padahal”. Kata-kata tersebut kemudian akan digabungkan dengan *library* sastrawi yang tersedia. Berikut Gambar 4.17 merupakan visualisasi persebaran kata negatif setelah melalui *stopword removal* Tahap II dan Gambar 4.18 merupakan *word cloud* sentimen negatif setelah melalui *stopword removal* Tahap II.



Gambar 4.18. Persebaran Kata Kelas Negatif (Tahap *Stopword Removal* II)

4.4.4 Fishbone Diagram

Berdasarkan hasil ulasan negatif, maka dapat diketahui informasi dari keluhan yang didapatkan oleh pengguna aplikasi. Keluhan tersebut terbagi menjadi 5 bagian yaitu, *people, price, product, process, promotion*. Pembagian tersebut dengan memperhatikan variabel yang terdapat di bauran pemasaran (*marketing mix*) Informasi tersebut dapat diketahui pada Gambar 4.19.



Gambar 4.20. Fishbone Ulasan Negatif

BAB V

ANALISIS DAN PEMBAHASAN

5.1 Analisis Deskriptif Pengumpulan Data Awal

Analisis deskriptif merupakan salah satu metode dalam melakukan analisis data. Analisis ini menggunakan metode numerik dan grafik untuk mencari pola data, meringkas informasi yang ada pada kumpulan data serta menyampaikan informasi dalam bentuk sesuai dengan apa yang diinginkan (Santosa R. G., 2004). Menurut Sugiyono (2009), analisis deskriptif adalah statistik yang digunakan untuk menganalisa data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi.

Pada penelitian ini, diketahui dari hasil awal proses *scraping* data yang dilakukan dari bulan Maret - Juni di dapatkan jumlah data sebesar 19.148 ulasan. Jumlah ulasan tersebut menunjukkan pada bulan Maret - Juni pengguna aplikasi *telemedicine* memberikan *feedback* (umpan balik) yaitu sangat suka dengan jumlah sebesar 16.137, pada kategori suka dengan jumlah 1.725, kategori sangat tidak suka 627, kategori tidak suka 177, dan kategori netral 482.

Jumlah ulasan yang telah didapatkan kemudian dilakukan analisis dari segi waktu atau perbulan. Diketahui jumlah ulasan yang didapatkan mengalami penurunan dari bulan Maret – Mei. Pada bulan Mei jumlah ulasan yang didapatkan sebesar 5.134 ulasan, bulan April sejumlah 4.442 ulasan, bulan Mei berjumlah 3.925. Namun, pada bulan Juni diketahui jumlah ulasan yang didapatkan mengalami peningkatan yang signifikan yaitu

5.647 ulasan. Selain itu, pada setiap bulannya *feedback* yang paling sering diberikan adalah sangat suka. Kemudian, diikuti kategori suka, sangat tidak suka, netral, dan tidak suka. Pada bulan Maret jumlah ulasan dengan *feedback* sangat suka mencapai 84.03%, *feedback* suka mencapai 9.56%, *feedback* sangat tidak suka 2.80%, *feedback* netral 2.73%, dan *feedback* tidak suka 0.88%. Pada bulan April *feedback* sangat suka jumlah persentase sebesar 83.14%, *feedback* suka 8.89%, *feedback* sangat tidak suka 4.01%ulasan, *feedback* netral 2.63%, dan *feedback* tidak suka 1.33%. Pada bulan Mei *feedback* sangat suka mencapai 86.62%,*feedback* suka 8.66%,*feedback* sangat tidak suka 2.01%,*feedback* netral 1.99%, dan *feedback* tidak suka 0.71%. Pada bulan Juni *feedback* sangat suka mencapai 83.76%, *feedback* suka 8.84%, *feedback* sangat tidak suka 4%, *feedback* netral 2.60%, dan *feedback* tidak suka 0.80%.

Diketahui pada bulan Maret – Mei jumlah ulasan yang diberikan mengalami penurunan. Akan tetapi, *feedback* yang diberikan pada bulan Maret – April untuk kategori sangat tidak suka mengalami peningkatan. Pada bulan Maret persentase yang di dapatkan sejumlah 2.80%, di bulan April sebesar 4.01%. Selain itu, pada kategori suka, grafik yang diberikan mengalami penurunan persentase pada bulan Maret – Mei.

5.2 Klasifikasi SVM

Diketahui berdasarkan Gambar 4.9 akurasi yang didapatkan untuk setiap algoritma SVM linier dan non-linier akurasi tertinggi dimiliki oleh algoritma SVM kernel linier (non-linier). Pada algoritma SVM kernel linier (non-linier) dengan 4 skenario yang telah dilakukan nilai akurasi yang diberikan selalu mengalami kenaikan dan menurun pada skenario ke 4. Skenario 1 dengan pembagian jumlah data latih 80% dan data uji 20% didapatkan akurasi sebesar 96,63%. Skenario 2 dengan pembagian jumlah data latih 75% dan data uji 25% didapatkan akurasi sebesar 96,82%. Skenario 3 dengan pembagian data latih 70% dan data uji 30% didapatkan akurasi sebesar 96,85%. Skenario 4 dengan pembagian data latih 65% dan data uji 35% didapatkan akurasi sebesar 96,80%. Pada algoritma SVM linier (linier) diketahui hasil dari 4 skenario yang telah dilakukan akurasi yang dimiliki mengalami kenaikan dan penurunan. Hal tersebut dapat dilihat pada skenario 1 yang memiliki nilai akurasi sebesar 96,55%, kemudian pada skenario 2 mengalami peningkatan menjadi 96,78%. Pada skenario 3 akurasi yang didapatkan

mengalami penurunan menjadi 96,73% dan pada skenario 4 mengalami peningkatan menjadi 96,74%. Pada algoritma SVM non linier seperti polynomial, sigmoid, dan RBF memiliki nilai akurasi yang sama, dan grafik yang ditunjukkan mengalami peningkatan dan menurun pada skenario 4. Pada skenario 1 didapatkan nilai akurasi sebesar 95,36%, skenario 2 nilai akurasi yang didapat sebesar 95,40%, skenario 3 akurasi sebesar 95,48%, dan skenario 4 menjadi 95,35%.

Pada Gambar 4.9. dapat diketahui cara kerja atau performa setiap algoritma SVM dalam melakukan klasifikasi. Hal tersebut dapat dijelaskan dengan *confusion matrix* Skenario 3. Berdasarkan Gambar 4.8 diketahui kemampuan dari algoritma SVM kernel linier (non-linier) untuk melakukan klasifikasi kelas positif yang tepat benar (*true positive*) berjumlah 118 ulasan dan kelas negatif yang tepat benar (*true negative*) berjumlah 5369 ulasan. Berbeda dengan algoritma SVM non linier yang lain seperti polynomial, sigmoid, dan RBF yang berjumlah *null* pada ketepatan kelas positif dan untuk kelas negatifnya berjumlah 5409 ulasan. Sedangkan pada algoritma SVM linier kemampuan dalam klasifikasi kelas positif yang tepat benar (*true positive*) berjumlah 127 dan kelas negatif yang tepat benar (*true negative*) berjumlah 5353.

Pada setiap algoritma klasifikasi terdapat kesalahan dalam melakukan klasifikasi / prediksi terhadap ulasan. hal tersebut dapat diketahui dari tabel *confusion matrix* yang ditampilkan pada Gambar 4.8. Pada algoritma klasifikasi SVM kernel linier (non-linier) terdapat kesalahan dalam memprediksi kelas positif (*false negative*) berjumlah 138 ulasan, dan kesalahan dalam melakukan prediksi kelas negatif (*false positive*) berjumlah 40 ulasan. Pada algoritma SVM non-linier yang lain seperti polynomial, sigmoid, dan RBF memiliki kemampuan yang sama, kesalahan dalam memprediksi kelas positif (*false negative*) berjumlah 256 ulasan, dan untuk kesalahan dalam memprediksi kelas negatif (*false positive*) berjumlah *null*. Pada algoritma SVM linier kesalahan yang dilakukan dalam prediksi kelas positif (*false negative*) berjumlah 129 ulasan, dan kesalahan prediksi pada kelas negatif (*false positive*) berjumlah 56 ulasan.

5.3 Visualisasi

Visualisasi ini merupakan bagian analisis deskriptif yang membantu untuk menampilkan informasi yang didapatkan (Wardhani, et al, 2019). Diketahui berdasarkan Gambar 4.10. persebaran kata yang paling sering muncul pada keseluruhan kelas sentimen adalah ‘sangat’ (7992 kata), ‘membantu’(6411 kata), ‘terimakasih’ (3188 kata), ‘dokter’ (2997 kata), ‘bagus’ (2712 kata), ‘aplikasi’ (2501 kata), ‘cepat’ (2336 kata), ‘halodoc’ (2037 kata), ‘dokternya’ (1961 kata), dan ‘baik’ (1893 kata). Setelah dilakukan proses *stopword removal* tahap II, diketahui Gambar 4.11 menunjukkan keseluruhan kata kelas sentimen sebagai berikut, ‘membantu’(6411 kata), ‘terimakasih’ (3188 kata), ‘dokter’ (2997 kata), ‘bagus’ (2712 kata), ‘aplikasi’ (2501 kata), ‘cepat’ (2336 kata), ‘halodoc’ (2037 kata), ‘dokternya’ (1961 kata), ‘baik’ (1893 kata) dan ‘ramah’ (1709 kata). berdasarkan 10 kata tertinggi yang paling banyak muncul menunjukkan aplikasi telemedicine memberikan manfaat yang baik untuk pengguna. Hal ini dapat dilihat pada kata ‘membantu’ yang memiliki jumlah kemunculan kata tertinggi, dan juga pada kata ‘baik’ dan ‘ramah’ yang muncul pada peringkat terakhir kata paling banyak muncul.

5.3.1 Sentimen Positif

Berdasarkan Gambar 4.12 dapat diketahui persebaran kata yang paling banyak muncul dalam ulasan positif adalah ‘sangat’ (7950 kata), ‘membantu’(6400 kata), ‘terimakasih’ (3164 kata), ‘dokter’ (2857 kata), ‘bagus’ (2688 kata), ‘cepat’ (2330 kata), ‘aplikasi’ (2298 kata), ‘halodoc’ (1948 kata), ‘dokternya’ (1918 kata), dan ‘baik’ (1880 kata). Pada ulasan positif persebaran kata yang paling banyak muncul untuk 10 kata tertinggi memiliki kesamaan dengan persebaran kata untuk seluruh kelas sentimen. Hal ini dapat dikatakan bahwa kata yang paling banyak muncul untuk keseluruhan kelas sentimen termasuk dalam kelas sentimen positif dan memiliki manfaat yang baik pada pengguna. Kemudian dikarenakan masih terdapat kata yang tidak memiliki arti atau mengandung makna pada penelitian maka dilakukan proses *stopword removal* tahap II. Pada Gambar 4.14. menunjukkan 10 kata terbanyak muncul dalam ulasan positif setelah melalui *stopword removal* tahap II. Pada Gambar 4.14. menampilkan kata “ramah” yang masuk dalam 10 kata terbanyak muncul dalam ulasan positif, setelah kata “sangat” dihilangkan. Pada ulasan positif peringkat tertinggi pada kata yang paling banyak muncul ada pada kata “membantu”, dan “terimakasih”. Kata ‘baik’ mengalami perubahan peringkat dari peringkat terakhir menjadi peringkat sembilan. Selain itu, terdapat sebuah aplikasi

telemedicine yang disebutkan dalam persebaran kelas sentimen positif yaitu *halodoc*. Serta kata ‘dokter’ yang muncul dalam persebaran kata positif.

Selanjutnya berdasarkan Gambar 4.13 dan 4.15 dapat diketahui persebaran seluruh kata yang terdapat di dalam ulasan kelas sentimen positif. Pada gambar tersebut diketahui semakin banyak kata yang muncul dalam ulasan sentimen positif maka kata tersebut akan ditunjukkan dengan ukuran kata yang semakin besar. Hal tersebut dapat dilihat pada kata ‘sangat’ yang memiliki peringkat pertama pada 10 kata paling banyak muncul pada persebaran kata positif di Gambar 4.12, namun pada Gambar 4.13 kata ‘sangat’ tersebut kemudian hilang karena telah melalui *stopword removal* tahap II. Selain itu diketahui juga terdapat kata ‘respon’, dan ‘cepatnya’. Hal ini menunjukkan bahwa aplikasi ini memberikan respon yang cepat. Serta adanya kata ‘sopan’ hal ini menunjukkan pelayanan yang diberikan ramah / sopan.

5.3.2 Sentimen Negatif

Berdasarkan Gambar 4.16. dapat diketahui 10 kata paling banyak muncul pada ulasan negatif adalah ‘aplikasi’ (203 kata), ‘bayar’(143 kata), ‘dokter’(140 kata), ‘mau’ (127 kata), ‘telepon’ (125 kata), ‘data’ (117 kata), ‘chat’ (114 kata), ‘alodokter’ (108 kata), ‘nomor’ (103 kata), dan ‘apa’(91 kata). Pada Gambar 4.15. masih terdapat kata yang tidak memiliki arti atau makna dalam penelitian seperti ‘mau’, ‘kalau’, ‘sama’, ‘apa’, ‘nya’, dan ‘padahal’ sehingga perlu dilakukan *stopword removal* tahap II. Pada Gambar 4.18 menunjukkan 10 kata terbanyak muncul dalam ulasan negatif dengan tambahan kata ‘halodoc’ sejumlah 89, ‘rekening’ sejumlah 64. Kata-kata tersebut menggantikan kata sebelumnya yang sudah dihilangkan. Melalui informasi tersebut dapat diketahui, walaupun pada sentimen positif kata ‘dokter’ dan ‘halodoc’ muncul, akan tetapi pada ulasan negatif kata ‘dokter’ dan ‘halodoc’ juga tetap muncul dan termasuk dalam peringkat ketiga tertinggi dan peringkat kesembilan tertinggi. Hal ini menunjukkan terdapat dokter dan *halodoc* yang belum dapat memberikan kesan positif pada pengguna, sama halnya dengan kata ‘aplikasi’. Kemudian terdapat kata ‘bayar’, ‘chat’, ‘telepon’, dan ‘data’ yang selalu muncul dalam ulasan negatif. Selain itu terdapat kata ‘alodokter’ yang merupakan salah satu aplikasi yang menjadi objek penelitian paling banyak muncul dalam ulasan negatif.

Kemudian visualisasi *word cloud* pada Gambar 4.17 dapat diketahui persebaran seluruh kata pada ulasan negatif sebelum dilakukan proses *stopword removal* tahap II. Pada Gambar 4.19. menunjukkan persebaran seluruh kata setelah melalui proses *stopword removal* tahap II. Pada gambar-gambar tersebut diketahui terdapat kata ‘konsultasi’. Hal ini menunjukkan pada saat konsultasi belum dapat memberikan kesan positif pada pengguna. Dimana konsultasi tersebut merupakan salah satu fitur yang terdapat dalam aplikasi kesehatan (*telemedicine*). Kemudian terdapat kata ‘aplikasi’, dan ‘dokter’ dengan ukuran kata yang lebih besar dibandingkan dengan kata yang lain. Hal ini menunjukkan terdapat keluhan atau ulasan negatif yang berhubungan terhadap dua kata tersebut. Selain itu, terdapat kata ‘obat’ yang muncul, hal ini menunjukkan bahwa terdapat kejadian/kondisi yang berhubungan dengan obat namun masih memberikan kesan tidak positif pada pengguna.

Pada Tabel 4.10. dapat diketahui kata-kata yang memiliki kemungkinan untuk muncul secara bersamaan. Kata ‘telepon’ dan ‘data’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘aplikasi’ sebesar 56,7% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 2,8%. Kata ‘chat’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘dokter’ sebesar 62,3% dan didukung persentase kedua kata tersebut muncul dari keseluruhan ulasan sebesar 9,7%. Kata ‘telepon’ dan ‘data’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘alodokter’ sebesar 63,3% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 3,1%. Kata ‘rekening’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘alodokter’ sebesar 65,3% dan didukung persentase kedua kata tersebut muncul dari keseluruhan ulasan sebesar 5,3%. Kata ‘chat’ dan ‘bayar’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘dokter’ sebesar 68,2% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 4,6%. Kata ‘aplikasi’ dan ‘nomor’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘telepon’ sebesar 70,3% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 3,1%. Kata ‘chat’ dan ‘aplikasi’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘dokter’ sebesar 72% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 3%. Kata ‘telepon’ dan ‘rekening’ memiliki persentase untuk muncul secara bersamaan dengan

kata ‘alodokter’ sebesar 68% dan didukung persentase ketiga kata tersebut muncul dari keseluruhan ulasan sebesar 2,8%. Selanjutnya Kata ‘beli’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘obat’ sebesar 78,9% dan didukung persentase kedua kata tersebut muncul dari keseluruhan ulasan sebesar 2,5%. Kata ‘pelanggan’ memiliki persentase untuk muncul secara bersamaan dengan kata ‘pelayanan’ sebesar 90% dan didukung persentase kedua kata tersebut muncul dari keseluruhan ulasan sebesar 3%.

Pada Tabel 4.10. dapat diketahui nilai *confidence* tertinggi dengan nilai 90% dimiliki pada pasangan kata ‘pelanggan’ dan ‘pelayanan’. Hal tersebut menunjukkan bahwa setiap kata ‘pelanggan’ muncul akan memiliki kemungkinan untuk kata ‘pelayanan’ muncul secara bersamaan. Selain itu, pada nilai *support* tertinggi dengan nilai 9,7% dimiliki oleh pasangan kata ‘chat’ dan ‘dokter’. Hal ini menunjukkan bahwa pasangan kata tersebut memiliki kemungkinan muncul secara bersamaan dari keseluruhan ulasan negatif.

5.4 *Fishbone* Diagram

Fishbone diagram adalah suatu pendekatan terstruktur yang memungkinkan dilakukan suatu analisis lebih terperinci dalam menemukan penyebab-penyebab suatu masalah, ketidaksesuaian, dan kesenjangan yang ada (Gaspers, 2002). Pendekatan ini membantu dalam menganalisis penyebab yang mungkin terjadi terhadap dampak yang tidak diinginkan.

Berdasarkan diagram sebab-akibat (*fishbone*) yang ada pada Gambar 4.19. diketahui ulasan negatif yang muncul disebabkan oleh 5 faktor utama yaitu, *people*, *price*, *process*, *product*, dan *promotion*. Pada faktor *people*, di sebabkan oleh dokter dan *customer service*. Dokter yang saat menjalankan konsultasi kurang ramah dalam penyampaian atau membalas chat, dan dokter yang lama dalam membalas chat konsultasi atau tidak dapat tersambung untuk chat dengan dokter bersangkutan. Penyebab lain keluhan pengguna adalah *Customer Service* yang kurang responsif dalam menanggapi keluhan. Mulai dari tidak dibalasnya *email*, telepon yang tidak dapat dihubungi, dan pengguna yang tidak mengetahui cara atau media dalam menghubungi *Customer Service*.

Faktor *price*, disebabkan oleh harga obat yang tidak sesuai dengan mitra yang bekerjasama. Hal ini dimaksudkan pada harga obat atau keterangan tentang obat yang tersedia di aplikasi berbeda dengan mitra atau berubah-ubah. Hal tersebut menyebabkan pengguna diharuskan menyediakan biaya tambahan atau membatalkan pesanan. Selain itu, metode pembayaran kurang variatif, yang tersedia adalah gopay dan beberapa bank tertentu (mandiri atau BCA). Kemudian terdapat kegagalan dalam melakukan pembayaran, dikhususkan pada fitur beli obat. Pengguna mengalami kegagalan saat melakukan transaksi pembayaran atau tidak dapat melakukan order / pesan.

Faktor *product*, disebabkan oleh *update* / pembaruan terhadap ketersediaan obat pada aplikasi yang belum dilakukan. Hal ini menyebabkan kerugian bagi pengguna aplikasi atau driver yang melakukan proses pengambilan barang tersebut. Karena pengguna yang akan pesan obat kemudian dibatalkan sehingga kebutuhan akan obat tersebut tidak tercapai, selain itu kerugian pada driver menurunkan performa driver jika kemudian dibatalkan. Penyebab selanjutnya adalah pengiriman produk yang lama. Selanjutnya yang menjadi penyebab adalah pesanan yang dibatalkan sepihak (tidak dilakukan oleh konsumen atau driver) atau pesanan tidak masuk. Hal tersebut dapat terjadi dikarenakan pembatalan yang dilakukan oleh pihak aplikasi atau mitra aplikasi tersebut. Selain itu, terdapat pesanan yang tidak masuk yang dimaksudkan pada pesanan tersebut tidak masuk atau di terima oleh mitra aplikasi, namun di aplikasi pengguna sudah melakukan transaksi.

Faktor *process*, disebabkan oleh kegagalan masuk ke dalam aplikasi. Pengguna tidak dapat masuk ke dalam aplikasi, setiap membuka aplikasi tersebut, aplikasi *force close* / atau tutup paksa sendiri. Penyebab selanjutnya adalah, kode verifikasi yang tidak muncul. Hal ini terjadi saat pengguna melakukan proses pendaftaran yang diharuskan untuk melakukan verifikasi melalui nomor telepon, akan tetapi kode verifikasi tersebut belum atau tidak dikirim ke nomor tersebut. Penyebab selanjutnya adalah auto debit. Auto debit ini terjadi pada transaksi konsultasi, dimana pengguna yang menyimpan saldo didalam aplikasi tersebut kemudian melakukan proses konsultasi dan saldo tersebut otomatis terdebit. Namun, pengguna merasa belum melakukan konsultasi dikarenakan chat konsultasi yang dilakukan belum dibalaskan tetapi saldo yang tersimpan mengalami pengurangan. Selain itu, auto debit yang terjadi adalah pengguna yang tidak melakukan

konsultasi dan tidak menyimpan saldo di aplikasi mengalami auto debit pada rekening yang dimiliki. Penyebab selanjutnya adalah keamanan data. Pada hal ini pengguna merasa bahwa data yang mereka miliki atau tercatat dalam aplikasi tidak aman. Ini disebabkan banyaknya telepon yang mengatasnamakan suatu aplikasi yang selalu menghubungi. Penyebab terakhir pada faktor ini adalah lokasi pengambilan obat yang jauh. Pada fitur pembelian obat, hal ini menjadi sangat penting dikarenakan tidak hanya merugikan bagi pengguna namun juga driver yang melakukan pengambilan obat tersebut. lokasi pengambilan obat yang jauh menjadikan jarak tempuh dan waktu yang dibutuhkan lebih lama. Selain itu, pada proses pembelian obat lokasi pemilihan pengambilan obat tidak dilakukan oleh pengguna, namun dilakukan secara otomatis oleh aplikasi.

Faktor *promotion*, disebabkan oleh telepon yang mengatasnamakan suatu aplikasi dan dilakukan berulang kali. Tujuan penelepon adalah untuk menawarkan asuransi kesehatan dan atau menanyakan data pribadi atau melengkapi data tersebut. Hal ini menjadikan pengguna merasa terganggu dan memberikan prasangka data yang dimiliki tidak aman. Penyebab lain adalah, variasi promo yang diberikan. Pada hal ini promo yang dilakukan adalah chat gratis dengan dokter umum atau pengguna baru mendapat chat gratis dengan dokter umum.

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan dari hasil pengolahan data, serta hasil dan pembahasan pada penelitian ini, maka dapat diambil kesimpulan, sebagai berikut :

1. *Feedback* (umpan balik) yang diberikan pengguna pada periode waktu Maret – Juni 2020 adalah sangat suka berjumlah 84,3%, suka berjumlah 9%, netral berjumlah 2,51%, tidak suka berjumlah 0,92%, dan sangat tidak suka berjumlah 3,27%. Kemudian jumlah ulasan yang diberikan pada periode waktu Maret – Mei mengalami rata-rata penurunan sebesar 12,55% per bulan. Selanjutnya pada bulan Juni mengalami peningkatan yang signifikan sebesar 30% jumlah ulasan yang diberikan. Selain itu, dengan menurunnya jumlah ulasan yang diberikan pada periode Maret – April, Jumlah *feedback* pada kategori sangat tidak suka meningkat sebesar 23,6%.
2. Informasi yang didapatkan pada kelas sentimen positif untuk keseluruhan aplikasi kesehatan (*telemedicine*) adalah aplikasi sangat membantu pengguna, selain itu aplikasi memiliki respon yang cepat serta pelayanan yang sopan dan ramah. Pada kelas sentimen negatif, informasi yang didapatkan, pengguna memberikan keluhan terkait data pengguna, adanya promosi yang dilakukan dengan melakukan telepon secara berulang kali, kegagalan dalam masuk aplikasi atau dalam melakukan pendaftaran, adanya pelayanan yang diberikan tidak ramah, dan fitur beli obat.
3. Melalui hasil analisis yang dilakukan dengan menggunakan diagram sebab-akibat (*fishbone*) dan mengacu pada teori bauran pemasaran (*marketing mix*). Teridentifikasi masalah yang menjadi keluhan pengguna sejumlah 16 buah. Keluhan

tersebut kemudian di kelompokkan menjadi 5P yang terdiri dari *people*, *price*, *process*, *product*, dan *promotion*.

6.2 Saran

Berdasarkan penelitian ini, Saran yang dapat diberikan pada penelitian selanjutnya adalah:

1. Penggunaan data dalam pelabelan di usahakan dapat seimbang atau tidak berbeda terlalu jauh untuk jumlah setiap labelnya pada proses pembangunan model dan pengujian model klasifikasi.
2. Pada sistem pelabelan kelas sentimen untuk kategori netral menggunakan kamus *lexicon*, pelabelan yang dilakukan berdasarkan pada bobot kata yang terdapat dalam kamus, diharapkan pada penelitian selanjutnya sistem pelabelan kelas sentimen dapat melakukan deteksi sentimen pada tingkat frasa dan kalimat multipolar.
3. Peneliti selanjutnya, diharapkan mampu mengkombinasikan *tools* lainnya dalam peningkatan kualitas sehingga dapat menunjukkan hasil lebih representatif.

DAFTAR PUSTAKA

- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. *Conference on Management of Data* (pp. 207-216). Washington: ACM SIGMOID.
- Ahyari, A. (2000). *Manajemen Produksi*. Yogyakarta: BPFE UGM.
- Amilia, S. (2016). Pengaruh Bauran Pemasaran Terhadap Kepuasan Konsumen Pada Warung Kanasha di Kota Langsa. *Jurnal Manajemen dan Keuangan*, 459=468.
- APJII. (2017). *Penetrasi & Perilaku Pengguna Internet Indonesia*. Retrieved from APJII web site: <https://apjii.or.id/survei2017>
- APJII. (2018). *Penetrasi & Profil Perilaku Pengguna Internet Indonesia Survei 2018*. Retrieved from APJII web site: <https://apjii.or.id/survei>
- Armstrong, G., Adam, S., Denize, S., & Kotler, P. (2015). *Principle of Marketing 6th*. Melbourne: Pearson Australia.
- Baenanda, L. (2019, Mei 02). *Mengenal Lebih Jauh Revolusi Industri 4.0*. Retrieved from Binus University Knowledge Management & Innovation: <https://binus.ac.id/knowledge/2019/05/mengenal-lebih-jauh-revolusi-industri-4-0/>
- Berry, M. W., & Kogan, J. (2010). *Text Mining Application and Theory*. United Kingdom: Wiley.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter . *Integer Journal*, 32-41.
- Damanik, D., & Wibowo, F. S. (2018). Implementasi Strategi 4P Dalam Meningkatkan Frekuensi Public Evenet di Harris Hotel and Conventions Bekasi (HHCB). *Natioanl Conference of Creative Industry* (pp. 210-229). Jakarta: Universitas Bunda Mulia.
- Eka, R. (2019, Desember 4). *The Understanding of Wellness Market in Jakarta 2019*. Retrieved from Dailysocial.id Web site: <https://dailysocial.id/research>
- Eka, R. (2020, Maret 20). *Scaling Through Technology Democratization Startup Report 2019*. Retrieved from Daily Social id website: <https://dailysocial.id/research>

- Evanmartua34. (2020, May 31). *Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based*. Retrieved from Github Repository: <https://github.com/evanmartua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based>
- Fatmawati, & Affandes, M. (2017). Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk). *Jurnal CoreIT*, 24-30.
- Fauziah, A. (2018, November 5). *Capek Antri ke Dokter? Coba 5 Aplikasi Ini*. Retrieved from Digation.id web site: <https://www.digation.id/read/012196/capek-antri-ke-dokter-coba-5-aplikasi-ini>
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. USA: Cambridge University Press.
- Field, M. J. (1996). *Telemedicine: A Guide to Assessing Telecommunications for Health Care (1996)*. Washington DC: The National Academies Press.
- Gaspers, V. (2002). *Total Quality Management*. Jakarta: Gramedia Pustaka Utama.
- Gorunescu, F. (2011). *Data Mining Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Gummerus, J., Liljander, V., Pura, M., & van Riel, A. (2004). Customer Loyalty to Content-Based Web sites: The Case of an Online Health-Care Service. *Journal of Service Marketing*, 175-186.
- Gunawan, D., Riana, D., Ardiansyah, D., Akbar, F., & Alfariza, S. (2020). Komparasi Algoritma Support Vector Machine dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur. *Jurnal Teknik Komputer AMIK BSI*, 121-129.
- Gupta, V., & Lehal, G. S. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence*.
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 26-32.
- Han, J., & Kamber, M. (2006). *Data Mining : Concepts and Techniques Second Edition*. San Fransisco: Morgan Kauffman.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques*. USA: Elsevier.

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques Third Edition*. USA: Elsevier.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. Waltham: Morgan Kaufmann Publishers.
- Hidayat, A. N. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Elektronik Sistem Informasi dan Komputer (JESIK)*, 1-7.
- ILIE, G., & CIOCOIU, C. N. (2010). Application of Fishbone Diagram to Determine The Risk of An Event with Multiple Causes. *Management Research and Practice*, 1-20.
- Junaidi, A. (2015). Internet of Things, Sejarah, Teknologi dan Penerapannya : Review. *Jurnal Ilmiah Teknologi Informasi Terapan*, 62-66.
- Juran, J. M. (1999). *Juran's Quality Handbook (5th edition)*. McGraw-Hill.
- Karsandi, R. (2017). *Mengenal Bisnis Startup di Indonesia*. Retrieved from Grevia Networks: <http://www.grevia.com/article/45/mengenal-bisnis-startup-di-indonesia>
- Kim, S.-M., & Oh, J. (2011, November 4). *Health Information Acquisition Online and Its Influence on Intention to Visit a Medical Institution Offline*. Retrieved from Information Research: <http://informationr.net/ir/16-4/paper505#par05>
- Kodu, S. (2013). Harga, Kualitas Produk dan Kualitas Pelayanan Pengaruhnya Terhadap Keputusan Pembelian Mobil Toyota Avanza. *Jurnal EMBA*, 1251 - 1259.
- Kotler, P., & Amstrong, G. (2012). *Principle of Marketing*. New Jersey: Pearson Education Limited.
- Koto, F., & Rahamaningtyas, G. Y. (2017). InSet Lexicon: Evalutaion of a Word List for Indonesian Sentiment Analysis in Microblogs. *2017 International Conference on Asian Language Processing (IALP)*, (pp. 391-394). Singapore.
- Kristiyanti, D. A. (2015). Analisis Sentimen Review Produk Kosmetik Melalui Komparasi Feature Selection. *Konferensi Nasional Ilmu Pengetahuan dan Teknologi* (pp. 74 - 81). Bekasi: Konferensi Nasional Ilmu Pengetahuan dan Teknologi.
- Kristiyono, J. (2015). Budaya Internet: Perkembangan Teknologi Informasi Dan Komunikasi Dalam mendukung Penggunaan Media Di Masyarakat. *SCRIPTURA*, 23-30.

- Kurnianti, A. W. (2017). Komunikasi Pemasaran Transportasi Online NGuberJEK. *Jurnal Komunikasi dan Kajian Media*, 69-84.
- Kusumaningrum, A., Pujiastuti, A., & Zeny, M. (2017). Pemanfaatan Internet Of Things Pada Kendali Lampu. *Compiler*, 53-59.
- Larose, D. T. (2005). *Discovering Knowledge in Data An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data An Introduction to Data Mining Second Edition*. New Jersey: John Wiley & Sons.
- Lidwina, A., & Fitra, S. (2020, Mei 18). *Penggunaan Layanan Digital yang Sering Digunakan Selama Covid-19?* Retrieved from Databoks Katadata Web site: <https://databoks.katadata.co.id/datapublish/2020/05/18/apa-layanan-digital-yang-sering-digunakan-selama-covid-19#>
- Liliana, L. (2016). A New Model of Ishikawa Diagram for Quality Assessment. *IOP Conference Series: Material Science and Engineering* (p. Innovative Manufacturing Engineering and Energy Conference). IOP.
- Liputan6.com. (2019, Januari 10). *Pengertian Startup Adalah, Sejarah Singkat dan Karakteristiknya*. Retrieved from liputan6 web site: <https://www.liputan6.com/tekno/read/3867489/pengertian-startup-adalah-sejarah-singkat-dan-karakteristiknya>
- Liu, B. (2010). Sentiment Analysis and Subjectivity. In N. Indurkha, & F. J. Damerau, *Handbook of Natural Language Processing Second Edition* (pp. 627-661). USA: Chapman and Hall/CRC Press.
- Machfoedz. (2005). *Pengantar Pemasaran Modern Cetakan Pertama*. Yogyakarta: Unit Penerbit & Percetakan Akademi Manajemen Perusahaan YKPN.
- Maimon, O., & Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook*. London: Springer.
- Mantyala, M. V., Graziotin, D., & Kuuttila, M. (2017). The Evolution of Sentiment Analysis-A Review of Research Topics, Venues, and Top Cited Papers. *Elsevier*, 16-32.
- Marlius, D. (2016). Pengaruh Bauran Pemasaran Jasa Terhadap Minat Nasabah Dalam Menabung Pada Bank Nagari Cabang Muaralabuh. *Jurmak*, 12-22.

- Maulani, A. (2020). Pusat Pengembangan Startup (Startup Incubator dan Co-Working Space) di Kota Pontianak. *Jurnal Online Mahasiswa Arsitektur Universitas Tanjungpura*, 12-28.
- Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*. Heidelberg: Springer-Verlag.
- Mutakhirah, I., Saptono, F., Hasanah, N., & Wiryadinata, R. (2007). Pemanfaatan Metode Heuristik Dalam Pencarian Jalur Terpendek Dengan Algoritma Semut dan Algoritma Genetika. *Seminar Nasional Aplikasi Teknologi Informasi 2007 (SNATI 2007)*. Yogyakarta.
- Nugraha, I. (2008). Algoritma Genetik Untuk Optimasi Penjadwalan Kegiatan Belajar Mengajar. *Jurnal Teknik Informatika*.
- Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2016). Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes. *Prosiding Seminar Nasional Sains dan Teknologi* (pp. 156 -161). Semarang: Fakultas Teknik Universitas Wahid Hasyim Semarang.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1750-1757.
- Parasuraman, A., Zeithaml, V., & Berry, L. (1998). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, 12-40.
- Pearson, J., Pearson, A., & Green, D. (2007). Determining The Importance of Key Criteria in Web Usability. *Management Research*, 816-828.
- Petriella, Y. (2019, Agustus 19). *Penggunaan Aplikasi Kesehatan di Indonesia Baru 10% dari Total Penduduk*. Retrieved Oktober 2, 2020, from Ekonomi Bisnis: <https://ekonomi.bisnis.com/read/20190819/12/1138279/penggunaan-aplikasi-kesehatan-di-indonesia-baru-10-dari-total-penduduk>
- Praptiwi, D. Y. (2018). *Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine dan Maximum Entropy (Studi Kasus: Review Bukalapak pada Google Play)*. Yogyakarta: Program Studi Statistika Universitas Islam Indonesia.

- Prawiroharjo, P., Pratama, P., & Librianty, N. (2019). Layanan Telemedis di Indonesia: Keniscayaan, Resiko, dan Batasan Etika. *Jurnal Etika Kedokteran Indonesia*, 1-9.
- Pusparisa, Y., & Fitra, S. (2020, April 13). *Sektor Bisnis Startup yang Terkena Dampak Corona*. Retrieved from Databoks Katadata web site: <https://databoks.katadata.co.id/datapublish/2020/04/13/sektor-bisnis-startup-yang-terkena-dampak-corona>
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi dan Mulut Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 802-810.
- Rachman, F., & Purnami, S. W. (2012). Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal dan Support Vector Machine (SVM). *Jurnal Sains dan Seni ITS*, D130-135.
- Ramadhan, B. (2020, Februari 16). *Data Internet di Indonesia dan Perilakunya Tahun 2020*. Retrieved from Teknoia: <https://teknoia.com/data-internet-di-indonesia-dan-perilakunya-880c7bc7cd19>
- Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1725-1732.
- Sansoko, B. A. (2020, May 27). *Dasar Hukum Telemedicine Di Indonesia*. Retrieved Oktober 2, 2020, from Trust Medis: <https://trustmedis.com/dasar-hukum-telemedicine-di-indonesia/>
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santosa, R. G. (2004). *Statistik*. Yogyakarta: Andi Offset.
- Santoso, B. S., Rahmah, M., Setiasari, T., & Sularsih, P. (2015). Perkembangan dan Masa Depan Telemedika di Indonesia. *National Conference on Information Technology and Technical Engineering (CITEE)*. Yogyakarta: Researchgate.
- Santoso, J. (2003). E-Service Quality: A Model of Virtual Service Quality Dimension. *Managing Service Quality: An International Journal*, 233-246.
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 1-47.

- Setiawan, K. (2003). *Paradigma Sistem Cerdas (Jaringan Saraf Tiruan, Logika Fazi dan Algoritma Genetik)*. Malang: Banyumedia Publishing.
- Setyowati, D., & Annur, C. M. (2020, Mei 12). *Jalan Terang Startup Kesehatan Menuju Unicorn di Tengah Pandemi*. Retrieved from Katadata Web site: <https://katadata.co.id/telaah/2020/05/12/jalan-terang-startup-kesehatan-menuju-unicorn-di-tengah-pandemi/1>
- Sheikh, L. M., Tanveer, B., & Hamdani, S. A. (2004). Interesting Measures for Mining Association Rules. *INMIC* (pp. 641-6411). IEEE.
- Shwab, K. (2016). *The Fourth Industrial Revolution*. New York: Crown Business.
- Sugiyono. (2009). *Metode Penelitian Bisnis (Pendekatan Kuantitatif, Kualitatif dan R&D)*. Bandung: Alfabeta.
- Sunjaya, A. P. (2019). Potensi, Aplikasi dan Perkembangan Digital Health di Indonesia. *J Indon Med Assoc*, 167-169.
- Teknologi.id. (2018, Juni 20). *Jumlah Aplikasi yang Tersedia di Google Play Store*. Retrieved Juli 15, 2020, from teknologi.id: <https://teknologi.id/ulasan/jumlah-aplikasi-yang-tersedia-di-google-play-store/>
- Tjiptono, F. (2007). *Pemasaran Jasa*. Malang: Banyumedia Publishing.
- Valdia, A., Luzon, M. V., & Herrera, F. (2017). Sentiment Analysis in Tripadvisor. *IEEE Intelligent System*, 72-77.
- Vijayarani, S., Ilamthi, J., & Nithya. (2015). Preprocessing Techniques for Text Mining - an Overview. *Internatioanl Journal of Computer Science & Communication Networks*, 7-16.
- Wardhani, A. D., Slamet, I., & Sugiyanto. (2019). Analisis Sentimen Penggunaan Kotak Suara Kardus Menggunakan SVM Linearly Separable Data. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, (pp. 167-174).
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *Jurnal Evolusi*, 25-31.
- Wijanarko, Y. (2019, Maret 18). *Jumlah Startup di Indonesia Ratusan atau Ribuan?* Retrieved from Kominfo Website: https://kominfo.go.id/content/detail/17233/jumlah-startup-di-indonesia-ratusan-atau-ribuan/0/sorotan_media

- Wootton, R., Craig, J., & Patterson, V. (2017). *Introduction to Telemedicine*. The Royal Society of Medicine Press Limited.
- Yanto, R., & Kesuma, H. D. (2017). *Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule*. Jatisi.
- Yustiani, R., & Yunanto, R. (2017). Peran Marketplace Sebagai Alternatif Bisnis Di Era Teknologi Informasi. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 43-48.
- Zafikri, A. (2008). *Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) Pada Sistem Temu Kembali Informasi*. Program Studi S-1 Ilmu Komputer FMIPA USU.
- Zararsiz, G., Elmali, F., & Ozturk, A. (2012). Bagging Support Vector Machines for Leukemia Classification. *IJCSI International Journal of Computer Science Issues*, 355-358.
- Zeithaml, V., Parasuraman, A., & Malhotra, A. (2000). A Conceptual Framework for Understanding E-Service Quality: Implications Con Future Research and Managerial Practice. *MSI Working Paper Series*, 1-49.
- Zukhri, Z. (2014). *Algoritma Genetika Metode Komputasi Evolusioner Untuk Menyelesaikan Masalah Optimasi*. Yogyakarta: Andi.

LAMPIRAN

Source code Python

Preprocessing

```
# menghilangkan punctuation dan angka (Spell Normalization)
import string, re

def cleansing (data):
    data = data.lower()

    remove = string.punctuation
    translator = str.maketrans(remove, ' '*len(remove))
    data = data.translate(translator)
    angber = ''.join([x for x in data if not (x.isdigit()
                                         or x[0] == '-'
                                         and x[1:].isdigit())])

    return angber

# tokenizer
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize

def tokenize (txt):
    word_token = word_tokenize (txt)
    return word_token

# stop words sastrawi
!pip install sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
def stopw (txt_tokenized):
    stop = [word for word in txt_tokenized if word not in stopwords]
    return stop

# mengembalikan data
def gabung(txt):
```

```

data_gabung = ' '.join([c for c in txt if c not in string.punctuation])
return data_gabung

```

```

pd.options.display.max_colwidth = None
data['data_p'] = data['Review'].apply(lambda x: cleansing(x))
data['data_token'] = data['data_p'].apply(lambda x: tokenize(x))
data['data_stop'] = data['data_token'].apply(lambda x: stopw(x))
#data['data_stem'] = data['data_stop'].apply(lambda x: stemmer(x))
masih gagal
data['data_join'] = data['data_stop'].apply(lambda x: gabung(x))
data

```

Pelabelan

```

import pandas as pd

lexicon = pd.read_csv('full_lexicon.csv')
lexicon

negasi = ['bukan', 'tidak', 'ga', 'gk']

lexicon_word = lexicon['word'].to_list()
lexicon_num_words = lexicon['number_of_words']

# (Evanmartua34, 2020)
import numpy as np
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

sencol = []
senrow = np.array([])
nsen = 0
factory = StemmerFactory()
stemmer = factory.create_stemmer()
sentimen_list = []
# function to write the word's sentimen if it is founded
def found_word(ind, words, word, sen, sencol, sentimen, add):
    # if it is already included in the bag of words matrix, then just
    # increase the value
    if word in sencol:
        sen[sencol.index(word)] += 1
    else:
        #if not, than add new word
        sencol.append(word)
        sen.append(1)
        add += 1

```

```

    #if there is a negation word before it, the sentimen would be the
    #negation of it's sentimen
    if (words[ind-1] in negasi):
        sentimen += -lexicon['weight'][lexicon_word.index(word)]
    else:
        sentimen += lexicon['weight'][lexicon_word.index(word)]

    return sen,sencol,sentimen,add

# checking every words, if they are appear in the lexicon, and then
# calculate their sentimen if they do
for i in range(len(data['data_join'])):
    nsen = senrow.shape[0]
    words = word_tokenize(data['data_join'][i])
    sentimen = 0
    add = 0
    prev = [0 for ii in range(len(words))]
    n_words = len(words)
    if len(sencol)>0:
        sen =[0 for j in range(len(sencol))]
    else:
        sen =[]

    for word in words:
        ind = words.index(word)
        # check whether they are included in the lexicon
        if word in lexicon_word :
            sen,sencol,sentimen,add= found_word(ind,words,word,sen,
sencol,sentimen,add)
        else:
            # if not, then check the root word
            kata_dasar = stemmer.stem(word)
            if kata_dasar in lexicon_word:
                sen,sencol,sentimen,add= found_word(ind,words,kata_
dasar,sen,sencol,sentimen,add)
            # if still negative, try to match the combination of words
            # with the adjacent words
            elif(n_words>1):
                if ind-1>-1:
                    back_1 = words[ind-1]+' '+word
                    if (back_1 in lexicon_word):
                        sen,sencol,sentimen,add= found_word(ind,words,
back_1,sen,sencol,sentimen,add)
                elif(ind-2>-1):
                    back_2 = words[ind-2]+' '+back_1
                    if back_2 in lexicon_word:
                        sen,sencol,sentimen,add= found_word(ind,words,
back_2,sen,sencol,sentimen,add)

```

```

# if there is new word founded, then expand the matrix
if add>0:
    if i>0:
        if (nsen==0):
            senrow = np.zeros([i,add],dtype=int)
        elif(i!=nsen):
            padding_h = np.zeros([nsen,add],dtype=int)
            senrow = np.hstack((senrow,padding_h))
            padding_v = np.zeros([(i-
nsen),senrow.shape[1]],dtype=int)
            senrow = np.vstack((senrow,padding_v))
        else:
            padding =np.zeros([nsen,add],dtype=int)
            senrow = np.hstack((senrow,padding))
            senrow = np.vstack((senrow,sen))
    if i==0:
        senrow = np.array(sen).reshape(1,len(sen))
# if there isn't then just update the old matrix
elif(nsen>0):
    senrow = np.vstack((senrow,sen))

sentimen_list.append(sentimen)

sencol.append('sentimen')
sentimen_array = np.array(sentimen_list).reshape(senrow.shape[0],1)
sentimen_data = np.hstack((senrow,sentimen_array))
df_sen = pd.DataFrame(sentimen_data,columns = sencol)

sentimen_score = pd.DataFrame([])
sentimen_score['text'] = data['Review'].copy()
sentimen_score['sentimen'] = df_sen['sentimen'].copy()
sentimen_score['rating'] = data['nilai'].copy()
sentimen_score

```

Klasifikasi

```

import pandas as pd

pd.options.display.min_rows = 10 # syntax untuk mengatur jumlah display yang di lihat, sumber :https://www.youtube.com/watch?v=pLZZgizVrRM&list=PLIeJsyt\_FUfLeEbweFJdmZwb-SDNc\_BOd&index=2
pd.options.display.max_rows = 60 # syntax dengan .head() akan secara default memunculkan data sejumlah lima baris, selain itu jumlah data yg di display akan mempengaruhi data yang di preprocessing

data = pd.read_csv('full data label fix.csv')
data

```

```

# menghilangkan punctuation dan angka
import string, re

def cleansing (data):
    data = data.lower()

    remove = string.punctuation
    translator = str.maketrans(remove, ' '*len(remove))
    data = data.translate (translator)
    angber = ''.join([x for x in data if not (x.isdigit()
                                                or x[0] == '-'
                                                and x[1:].isdigit())])

    return angber

# tokenizer
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize

def tokenize (txt):
    word_token = word_tokenize (txt)
    return word_token

# stop words sastrawi
!pip install sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
def stopw (txt_tokenized):
    stop = [word for word in txt_tokenized if word not in stopwords]
    return stop

# mengembalikan data
def gabung(txt):
    data_gabung = ' '.join([c for c in txt if c not in string.punctuation])
    return data_gabung

data['data_p'] = data['text'].apply (lambda x: cleansing(x))
data['data_token'] = data['data_p'].apply (lambda x: tokenize (x))
data['data_stop'] = data['data_token'].apply (lambda x: stopw (x))

```

```

data['data_join'] = data['data_stop'].apply(lambda x: gabung (x))
data

data_fix = pd.DataFrame([])
data_fix ['text'] = data ['data_join'].copy()
data_fix ['sentimen'] = data ['label_rating'].copy()
data_fix

#Import libraries
import numpy as np
import pandas as pd
import nltk
from nltk.corpus import stopwords
import string
import pickle
from sklearn import svm

#Split the data into 80% training (X_train & y_train) and 20% testing (X_test & y_test) data sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_text, label, test_size = 0.20, random_state = 0)

# Get the shape of messages_bow
# messages_bow.shape
from sklearn.feature_extraction.text import TfidfVectorizer

# vectorize email text into tfidf matrix
# TfidfVectorizer converts collection of raw documents to a matrix of TF-IDF features.
# It's equivalent to CountVectorizer followed by TfidfTransformer.
vectorizer = TfidfVectorizer(
    input='content',      # input is actual text
    lowercase=True,      # convert to lower case before tokenizing
)
features_train_transformed = vectorizer.fit_transform(X_train)
features_test_transformed = vectorizer.transform(X_test)

# Classifier - Algorithm - SVM kernel linear
# fit the training dataset on the classifier
SVM_Classifier_kernel_linier = svm.SVC(C=1.0, kernel='linear', degree=8, gamma='auto')
SVM_Classifier_kernel_linier.fit(features_train_transformed,y_train)

# predict the labels on validation dataset

```



```

# Classifier - Algorithm - SVM kernel polynomial
# fit the training dataset on the classifier
SVM_Classifier_kernel_poly = svm.SVC(C=1.0, kernel='poly', degree=8,
    gamma='auto')
SVM_Classifier_kernel_poly.fit(features_train_transformed,y_train)

# predict the labels on validation dataset

# Classifier - Algorithm - SVM kernel sigmoid
# fit the training dataset on the classifier
SVM_Classifier_kernel_sigmoid = svm.SVC(C=1.0, kernel='sigmoid', deg
ree=8, gamma='auto')
SVM_Classifier_kernel_sigmoid.fit(features_train_transformed,y_train
)

# predict the labels on validation dataset

# Classifier - Algorithm - SVM linear
# fit the training dataset on the classifier
SVM_Classifier = svm.LinearSVC()
SVM_Classifier.fit(features_train_transformed,y_train)

# predict the labels on validation dataset

# Classifier - Algorithm - SVM rbf
# fit the training dataset on the classifier
SVM_Classifier_rbf = svm.SVC(C=1.0, kernel='rbf', degree=8, gamma='a
uto')
SVM_Classifier_rbf.fit(features_train_transformed,y_train)

# predict the labels on validation dataset

#Evaluate the model on the training data set SVM linier kernel
from sklearn.metrics import classification_report,confusion_matrix,
    accuracy_score
predictions_SVM = SVM_Classifier_kernel_linier.predict(features_test
_transformed)
print('kernel linier')
print(classification_report(y_test ,predictions_SVM ))
print('Confusion Matrix: \n',confusion_matrix(y_test,predictions_SV
M))
print()
print('Accuracy: ', accuracy_score(y_test,predictions_SVM))

```

```

#Evaluate the model on the training data set SVM kernel poly
from sklearn.metrics import classification_report,confusion_matrix,
    accuracy_score
predictions_SVM = SVM_Clasifier_kernel_poly.predict(features_test_t
ransformed)
print('polynomial')
print(classification_report(y_test ,predictions_SVM ))
print('Confusion Matrix: \n',confusion_matrix(y_test,predictions_SV
M))
print()
print('Accuracy: ', accuracy_score(y_test,predictions_SVM))

#Evaluate the model on the training data set SVM kernel sigmoid
from sklearn.metrics import classification_report,confusion_matrix,
    accuracy_score
predictions_SVM = SVM_Clasifier_kernel_sigmoid.predict(features_tes
t_transformed)
print('sigmoid')
print(classification_report(y_test ,predictions_SVM ))
print('Confusion Matrix: \n',confusion_matrix(y_test,predictions_SV
M))
print()
print('Accuracy: ', accuracy_score(y_test,predictions_SVM))

#Evaluate the model on the training data set SVM linier
from sklearn.metrics import classification_report,confusion_matrix,
    accuracy_score
predictions_SVM = SVM_Clasifier.predict(features_test_transformed)
print('linier')
print(classification_report(y_test ,predictions_SVM ))
print('Confusion Matrix: \n',confusion_matrix(y_test,predictions_SV
M))
print()
print('Accuracy: ', accuracy_score(y_test,predictions_SVM))

#Evaluate the model on the training data set SVM rbf
from sklearn.metrics import classification_report,confusion_matrix,
    accuracy_score
predictions_SVM = SVM_Clasifier_rbf.predict(features_test_transform
ed)
print('RBF')
print(classification_report(y_test ,predictions_SVM ))
print('Confusion Matrix: \n',confusion_matrix(y_test,predictions_SV
M))
print()
print('Accuracy: ', accuracy_score(y_test,predictions_SVM))

```

Visualisasi Diagram Batang

Visualisasi seluruh kelas

```
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
df = pd.read_csv('Hasil_SVM.csv')

#keseluruhan kata
visual = pd.DataFrame([])
visual['teks'] = df['Labeled_Data'].values.copy()
visual

#keseluruhan
word_dict = {}
for i in range(0, len(visual['teks'])):
    sentence = visual['teks'][i]
    word_token = word_tokenize(str(sentence))
    for j in word_token:
        if j not in word_dict:
            word_dict[j] = 1
        else:
            word_dict[j] += 1

df_top = pd.DataFrame((word_dict.items()))
df_top = df_top.rename(columns={0:'word', 1:'counts'})
df_top = df_top.sort_values(by='counts', ascending=False)
df_top.head(10)

import matplotlib.pyplot as plt

df_top.head(10).sort_values(by='counts', ascending=True).plot.barh(x
='word', y='counts', rot=0, figsize=(20,10), fontsize=30)
plt.savefig('grafik .png', bbox_inches='tight')
```

Visualisasi Kelas Positif

```
#visualisasi positif
pos_dict = {}
for i in range(0, len(positif['Labeled_Data'])):
    sentence = positif['Labeled_Data'][i]
    token = word_tokenize(str(sentence))
    for b in token:
        if b not in pos_dict:
            pos_dict[b] = 1
```

```

else:
    pos_dict[b] += 1

positif_top = pd.DataFrame((pos_dict.items()))
positif_top = positif_top.rename(columns={0:'word', 1:'counts'})
positif_top = positif_top.sort_values(by='counts', ascending=False)
positif_top.head(10)

positif_top.head(10).sort_values(by='counts', ascending=True).plot.barh(x='word', y='counts', rot=0, figsize=(20,10), fontsize=30)
plt.savefig('grafik positif .png', bbox_inches='tight')

```

Visualisasi Kelas Negatif

```

#visualisasi negatif
neg_dict = {}
for i in range(0, len(negatif['Labeled_Data'])):
    sentence = negatif['Labeled_Data'][i]
    tokenisasi = word_tokenize(str(sentence))
    for c in tokenisasi:
        if c not in neg_dict:
            neg_dict[c] = 1
        else:
            neg_dict[c] += 1

negatif_top = pd.DataFrame((neg_dict.items()))
negatif_top = negatif_top.rename(columns={0:'word', 1:'counts'})
negatif_top = negatif_top.sort_values(by='counts', ascending=False)
negatif_top.head(10)

negatif_top.head(10).sort_values(by='counts', ascending=True).plot.barh(x='word', y='counts', rot=0, figsize=(20,10), fontsize=30)
plt.savefig('grafik negatif .png', bbox_inches='tight')

```

Word Cloud

```

from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
import pandas as pd
import os
from PIL import Image
import numpy as np

currdir = os.path.dirname('cloud (1).png')
mask = np.array(Image.open(os.path.join(currdir, 'cloud (1).png')))

```

```

# Start with one review:
# text = df.data_label
pos_word = str(pos.Labeled_Data)
neg_word = str(neg.Labeled_Data)
#net_word = str(net.Labeled_Data)
# Create and generate a word cloud image:
print ("Positif :")
wordcloud_pos = WordCloud(background_color='white',mask=mask).generate(pos_word)
plt.imshow(wordcloud_pos, interpolation='bilinear')
plt.axis("off")
plt.show()
wordcloud_pos.to_file("SVM_pos.png")
print ("Negatif :")
wordcloud_neg = WordCloud(background_color='white', mask=mask).generate(neg_word)
plt.imshow(wordcloud_neg, interpolation='bilinear')
plt.axis("off")
plt.show()
wordcloud_neg.to_file("SVM_neg.png")

```

Asosiasi

```

import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')

unique_items_list = []

# for each index it will iter row by row
for index, row in data.iterrows():

    # splitting items with , and creating a new list for row & it will
    # all going add it again
    # ...item_series list for each iteration..so item_series will be
    # list of lists..
    items_series = list(row.str.split(','))

    # again reading each list elements from item_Series which is big
    # list as mentioned above code
    for each_row_list in items_series:

```

```

# iterating each item from each_row_lists
for item in each_row_list:

    # for first iteration..unique_items_list is empty so first
    # item directly append to it.
    #...from next onwards..it will start to check condition
    'not in'
    #....& if item not found in unique_items_list list then
    # it will append to it.
    #.....finally we will get one unique item list..
    if item not in unique_items_list:
        unique_items_list.append(item)

df_apriori = pd.DataFrame(columns=unique_items_list)
df_apriori

dataset1 =df_apriori.copy()

dataset1.head()

dataset1.shape

dataset1.info()

# Sincd efault datatype saved as 'object'. Converting in 'integer'
# datatype

dataset1 = dataset1.astype('uint8')
dataset1.info()

freq_items

freq_items.to_csv('asosiasi kata negatif.csv')

# for this we need support value dataframe..that is fre_items from
# measure1.

confidence_association = association_rules(freq_items, metric='confidence',
min_threshold=0.5)

# min_threshold is nothing but setting min % crieteria. In this case
# i have choosen 20%
#...confidence should be minimum 20%.

```

```
# checking combination in first 10 rows from dataset  
confidence_association.sort_values(by = ['support'])
```

