

SKRIPSI

**IMPLEMENTASI ALGORITMA *BIDIRECTIONAL LONG SHORT TERM MEMORY*
(*BILSTM*) DENGAN ANALISIS SENTIMEN *EMOJI EMBEDDED* TERHADAP
TINGKAT KECEMASAN COVID-19 DAN SOSIAL EKONOMI MASYARAKAT**

Oleh :

JENIE MARCELINA

(065118138)



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
UNIVERSITAS PAKUAN
BOGOR
2024**

SKRIPSI

**IMPLEMENTASI ALGORITMA *BIDIRECTIONAL LONG SHORT TERM MEMORY*
(*BILSTM*) DENGAN ANALISIS SENTIMEN *EMOJI EMBEDDED* TERHADAP
TINGKAT KECEMASAN COVID-19 DAN SOSIAL EKONOMI MASYARAKAT**

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam

Oleh :

JENIE MARCELINA

(065118138)



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
UNIVERSITAS PAKUAN
BOGOR
2024**

HALAMAN PERSEMBAHAN

“Alhamdulillah Rabbil Aalamin, puji syukur kepada Allah SWT. Terimakasih atas karunia-Mu yang telah memberikan kemudahan dan kelancaran sehingga skripsi Ini dapat terselesaikan dengan baik. Saya ucapkan terimakasih yang sebesar-besarnya untuk diri saya sendiri yang telah berjuang dan tidak berhenti berusaha serta tetap kuat dan tegar selama ini dalam perjalanan yang tidak selalu sesuai dengan apa yang diharapkan”.

“Skripsi ini saya persembahkan kepada Orang tua tersayang Papa dan Mama, yang telah memberikan dukungan moril maupun materi serta doa yang tiada henti untuk kesuksesan saya. Karena tiada kata seindah lantunan doa dan tiada doa yang paling khusuk selain doa yang terucap dari kedua orang tua. Ucapan terimakasih saja takkan pernah cukup untuk membalas kebaikan orang tua saya, karena itu terimalah persembahan bakti dan cinta ku untuk kalian papaku dan mamaku”.

“Skripsi ini saya persembahkan kepada Kakak ku tersayang Furry Agnestya Sari. Terimakasih atas doa, dukungan moril dalam setiap tugas yang penulis kerjakan, serta selalu menjadi pengingat dikala penulis melakukan kesalahan dalam setiap pengambilan keputusan. Maka dari itu, terimalah persembahan bukti dan cinta ku untukmu, dan aku dapat menunjukkan padamu bahwa, Finally, I did it kak!”.

“Skripsi ini saya persembahkan untuk teman-teman terbaik di penghujung perjalanan ku yang senantiasa memberikan arahan, dukungan, serta motivasi. Dan teman seperjuangan saya (Maria, Sela, Ismi, Diva, Nurmin, Khusnul, Verdi, Elih) tanpa semangat, dukungan dan bantuan kalian semua takkan mungkin aku sampai disini, terimakasih untuk kenangan manis yang telah mengukir selama ini. Dengan perjuangan dan kebersamaan kita pasti bisa! Semangat!!”

“Sesungguhnya sesudah kesulitan itu ada kemudahan, maka apabila kamu telah selesai (dari semua segala urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain. Dan hanya kepada Tuhanmulah hendaknya kamu berharap.” (QS. Al Insyirah, 6-8)

“ Where there is love there is hate, where there is light there is darkness. Then you don't have to think too much about what people say. Because when there are people who hate you, then rest assured that there will be more people who also love you even more than what you imagine. keep up the spirit, and be yourself. “ – RM BT

HALAMAN PENGESAHAN

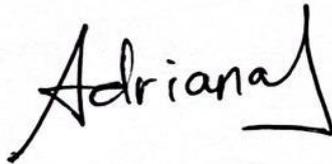
Judul : Implementasi Algoritma *Bidirectional Long Short Term Memory* (BiLSTM)
Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan
Covid-19 dan Sosial Ekonomi Masyarakat.

Nama : Jenie Marcelina

NPM : 065118138

Mengesahkan,

Pembimbing Pendamping
Program Studi Ilmu Komputer
FMIPA – UNPAK



Adriana Sari Aryani, S.Kom., M.Cs.

Pembimbing Utama
Program Studi Ilmu Komputer
FMIPA – UNPAK



Dr. Eneng Tita Tosida, S.Tp., M.Si., M.Kom.

Mengetahui,

Ketua Program Studi Ilmu Komputer
FMIPA – UNPAK



Arie Qur'ania, S.Kom., M.Kom.

Dekan
FMIPA – UNPAK



Asep Denih, S.Kom., M.Sc., Ph.D.

PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI

Dengan ini saya menyatakan bahwa:

Karya tulis ini bukan merupakan karya tulis yang pernah dipublikasikan atau sudah pernah dipakai untuk mendapatkan gelar sarjana di Universitas lain, kecuali pada bagian-bagian di mana sumber informasinya dicantumkan dengan cara referensi yang semestinya.

Demikian pernyataan ini saya buat dengan sebenar-benarnya. Apabila kelak dikemudian hari terdapat gugatan, Penulis bersedia dikenakan sanksi sesuai dengan peraturan yang berlaku.



PERNYATAAN PELIMPAHAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Saya yang bertanda tangan di bawah ini:

Nama : Jenie Marcelina
NPM : 065118138
Judul Skripsi : Implementasi Algoritma *Bidirectional Long Short Term Memory* (BiLSTM) Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat.

Dengan ini saya menyatakan bahwa Paten dan Hak Cipta dari produk Skripsi dan Tugas Akhir di atas adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun.

Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan Paten, hak cipta dari karya tulis saya kepada Universitas Pakuan.

Bogor, Januari 2024



Jenie Marcelina

RIWAYAT HIDUP



JENIE MARCELINA, lahir pada 03 Maret 2000 di Bekasi, Jawa Barat. Penulis merupakan anak kedua dari pasangan Bapak Agus Hernawan dan Ibu Santi Dhiana. Penulis memulai pendidikan formal pada tahun 2005 di TK Al-Ikhsan dan lulus pada 2006, kemudian melanjutkan pendidikan sekolah dasar di SD Abdi Negara 2 dan lulus pada tahun 2012. Penulis melanjutkan sekolah menengah pertama di SMP Negeri 3 Cibarusah pada tahun 2012 sampai dengan 2015 dan menyelesaikan pendidikan sekolah menengah atas di SMAN 1 Cibarusah kemudian lulus pada tahun 2018. Lalu pada tahun 2018 penulis melanjutkan pendidikan tingkat sarjana (S1) dengan Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam di Universitas Pakuan. Selama masa perkuliahan penulis aktif menjadi anggota Badan Eksekutif Mahasiswa (BEM) dan aktif sebagai anggota Himpunan Mahasiswa Ilmu Komputer (HIMAKOM) Universitas Pakuan. Setelah melaksanakan panjang serangkaian perkuliahan serta melaksanakan praktek lapang, pada bulan Januari 2023 penulis melakukan penelitian tugas akhir dengan mengambil data secara online di media Sosial Twitter dengan mengambil judul “Implementasi Algoritma *Bidirectional Long Short Term Memory (BiLSTM)* Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat”.

RINGKASAN

Jenie Marcelina (2024). Implementasi Algoritma *Bidirectional Long Short Term Memory* (BiLSTM) Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat. Dibawah bimbingan Ibu Eneng Tita Tosida dan Ibu Adriana Sari Aryani.

Pandemi COVID-19 telah menghadirkan tantangan multidimensi di Indonesia, mempengaruhi sosial, ekonomi, dan kesehatan masyarakat secara signifikan pada tingkat kecemasan. Kecemasan masyarakat terkait pandemi dapat tercermin dalam media online, terutama Twitter, yang menjadi saluran utama untuk berbagi informasi dan ekspresi emosi. Penelitian ini bertujuan untuk memahami tingkat kecemasan masyarakat sehubungan pasca pandemi COVID-19 dengan menggunakan suatu metode klasifikasi. Klasifikasi dilakukan menggunakan metode *Knowledge Discovery in Database* (KDD) dengan algoritma *Bidirectional LSTM* dan analisis sentimen penyematan emoji, serta dilakukan juga pengujian *K-Fold Cross Validation* dengan macam optimizer. Hasil akhir tingkat akurasi terbaik yang didapatkan sebesar 98,08%. Hal tersebut menunjukkan bahwa model klasifikasi yang dibuat sudah baik.

Kata Kunci: Kecemasan masyarakat, Covid-19, *Emoji embedded*, *Bidirectional LSTM*, *Sentiment analysis*.

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan nikmat serta hidayah-Nya terutama nikmat kesempatan dan kesehatan serta izin-Nya penulis dapat menyelesaikan penulisan laporan skripsi dengan judul “**Implementasi Algoritma *Bidirectional Long Short Term Memory* Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 Dan Sosial Ekonomi Masyarakat**” ini.

Dalam penulisan tugas akhir ini, penulis dengan senang hati ingin mengucapkan terima kasih yang sebesar-besarnya kepada berbagai pihak terutama kepada:

1. Dr. Eneng Tita T., S.Tp., M.Si., M.Kom. sebagai dosen sekaligus pembimbing utama, yang berkenan meluangkan waktu untuk memberikan bimbingan dan saran dalam menyelesaikan laporan penelitian ini,
2. Adriana Sari Aryani, S.Kom., M.Cs. sebagai dosen sekaligus pembimbing kedua yang juga berkenan meluangkan waktu untuk memberikan arahan serta saran dalam penyusunan laporan penelitian ini.
3. Kedua Orang Tua, Mama Santi dan Papa Agus tercinta atas segala dukungan, motivasi, do'a, dan selalu menjadi penyemangat bagi penyusun.
4. Kaka kandung tercinta Furry Agnestya Sari yang selalu menjadi penyemangat, selalu memberi dukungan dan menjadi motivasi bagi penyusun.
5. Rekan seperjuangan persekripsian khususnya Maria Qibtia, Sela Juniasti Sofwah, Ismiatun Hasanah, Diva Randika, Nurmin Kuma, Elih Febriani, teman seperjuangan BERKARAKTER, dan rekan HIMAKOM 2018 yang telah memberikan dukungan, semangat, motivasi kepada penulis selama ini serta semua pihak yang telah memberi semangat baik secara langsung maupun tidak langsung yang tidak dapat disebutkan satu persatu.

Penyusun menyadari memiliki keterbatasan waktu dan kemampuan dalam penulisan laporan ini, sehingga masih jauh dari kata sempurna. Oleh karena itu, segala kritik dan saran yang membangun akan diterima dengan senang hati. Mudah-mudahan Allah SWT akan membalas semua kebaikan kepada semua pihak yang membantu. Akhir kata, penyusun mengucapkan terimakasih. Semoga laporan ini dapat bermanfaat bagi penyusun pada khususnya dan para pembaca pada umumnya.

Bogor, Januari 2024

Jenie Marcelina
065118138

DAFTAR ISI

HALAMAN PERSEMBAHAN.....	i
HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI	Error! Bookmark not defined.
PERNYATAAN PELIMPAHAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA	Error! Bookmark not defined.
RIWAYAT HIDUP	v
RINGKASAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	2
1.3 Ruang Lingkup Penelitian	3
1.4 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Landasan Teori.....	4
2.1.1 Tingkat Kecemasan	4
2.1.2 Text Mining.....	4
2.1.3 Analisis Sentimen.....	4
2.1.4 Bidirectional Long Short-Term Memory (BiLSTM).....	5
2.1.5 Emoji Embedding	7
2.1.6 Confusion Matrix.....	8
2.1.7 K-Fold Cross Validation	8
2.2 Penelitian Terdahulu	10
2.2 Tabel Perbandingan Penelitian	11
BAB III METODE PENELITIAN.....	12
3.1 Metode Penelitian.....	12
3.1.1 Tahap Pengumpulan Data	12
3.1.2 Pembersihan Data dan Integrasi Data	12
3.1.3 Seleksi Data dan Transformasi Data	13
3.1.4 Proses Data Mining	13

3.1.5	Evaluasi Pola (<i>Pattern Evaluation</i>)	13
3.2	Flowchart Penelitian	13
3.3	Waktu dan Tempat Pelaksanaan	14
3.4	Alat dan Bahan	14
3.4.1	Alat	14
3.4.2	Bahan	14
BAB IV	PERANCANGAN DAN IMPLEMENTASI	15
4.1	Perancangan Sistem	15
4.1.1	Tahap Pengumpulan Data (<i>Data Selection</i>)	15
4.1.2	Pembersihan Data dan Integrasi Data	16
4.1.3	Selecting Data dan Transformasi Data (<i>Selecting and Transformation</i>)	22
4.1.4	Data Mining	23
4.1.5	Evaluasi Pola (<i>Pattern Evaluation</i>)	27
4.1.6	Presentasi Pengetahuan (<i>Knowledge Evaluation</i>)	28
4.2	Implementasi Sistem	28
BAB V	HASIL DAN PEMBAHASAN	29
5.1	Hasil	29
5.1.1	Tahap Pengumpulan Data (<i>Data Selection</i>)	29
5.1.2	Proses Pembersihan data dan Integrasi Data	29
5.1.3	Proses Pembobotan Kata	29
5.1.4	Proses Model BiLSTM	30
5.1.5	Hasil Implementasi dan Pengujian Model	30
5.2	Pembahasan	31
5.2.1	Hasil Klasifikasi Sentimen Aspek Kesehatan	31
5.2.2	Hasil Klasifikasi Sentimen Aspek Sosial	32
5.2.3	Hasil Klasifikasi Sentimen Aspek Ekonomi	33
5.3	Hasil Evaluasi Pola (<i>Pattern Evaluation</i>)	34
5.4	Wordcloud	36
5.5	Pengujian <i>K-Fold Cross Validation</i>	37
BAB VI	KESIMPULAN DAN SARAN	38
6.1	Kesimpulan	38
6.2	Saran	38
DAFTAR PUSTAKA	39
LAMPIRAN	41

DAFTAR GAMBAR

Gambar 1. Ilustrasi LSTM (Rizki et al, 2020)	5
Gambar 2. Ilustrasi BiLSTM.....	7
Gambar 3. Tahapan KDD (Han et al., 2022).....	12
Gambar 4. Diagram Alur Penelitian.....	13
Gambar 5. Proses Crawling Data	15
Gambar 6. Tahap Preprocessing.....	17
Gambar 7. Implementasi menggunakan Jupyter Notebook (Anaconda3)	28
Gambar 8. Proses Model Bidirectional Long Short-Term Memory	30
Gambar 9. Hasil Confussion Matrix.....	31
Gambar 10. Hasil Klasifikasi Sentimen Aspek Kesehatan dengan Emoji.....	31
Gambar 11. Hasil Klasifikasi Sentimen Aspek Sosial dengan Emoji.....	32
Gambar 12. Hasil Klasifikasi Sentimen Aspek Ekonomi dengan Emoji.....	33
Gambar 13. Hasil Training dan Validation Accuracy	34
Gambar 14. Hasil Training dan Validation Loss.....	34
Gambar 15. Hasil Confussion Matrix Aspek Kesehatan dengan Emoji	35
Gambar 16. Hasil Confussion Matrix Aspek Sosial dengan Emoji	35
Gambar 17. Hasil Confussion Matrix Aspek Sosial dengan Emoji	35
Gambar 18. Wordcloud Aspek Kesehatan dengan Emoji.....	36
Gambar 19. Wordcloud Aspek Sosial dengan Emoji.....	36
Gambar 20. Wordcloud Aspek Ekonomi dengan Emoji.....	36

DAFTAR TABEL

Tabel 1. Contoh Emoji dan Deskripsi Nama Labelnya.....	7
Tabel 2. Tabel Perbandingan Penelitian Terdahulu	11
Tabel 3. Keterangan Jumlah Data Berdasarkan Setiap Aspek Dalam Penelitian	16
Tabel 4. Jenis Emoji Yang Ditemukan Dalam Aspek Kesehatan	17
Tabel 5. Coverting Emoji	18
Tabel 6. Remove Username, hashtag, URL, & Mention	19
Tabel 7. Remove Punctuation	19
Tabel 8. Case Folding.....	19
Tabel 9. Tokenization.....	20
Tabel 10. Remove Stopwords	20
Tabel 11. Stemming	20
Tabel 12. Word Translation from Bahasa to English.....	21
Tabel 13. Sampel Data Sebelum Preprocessing dan Hasil Preprocessing	21
Tabel 14. Tabel Perhitungan Skor Sentiment.....	21
Tabel 15. Sampel data One Hot Encoding	22
Tabel 16. Hasil Prediksi	27
Tabel 17. Hasil Confusion Matrix.....	27
Tabel 18. Hasil Rata-rata Akurasi Pengujian	35
Tabel 19. Hasil Akurasi Data Dengan Aspek Emoji Pengujian K-Fold Cross Validation	37
Tabel 20. Hasil Akurasi Data Dengan Aspek Tanpa Emoji Pengujian K-Fold Cross Validation.....	37

DAFTAR LAMPIRAN

Lampiran 1. SK Penelitian	42
Lampiran 2. Kartu Bimbingan Mahasiswa	44
Lampiran 3. Gambar Proses Scrapping Tweet Kecemasan Dalam 3 Aspek	46
Lampiran 4. Hasil Wawancara Dengan Pakar Bahasa Indonesia	47
Lampiran 5. Biodata Pakar Bahasa Indonesia.....	48
Lampiran 6. Deskripsi Label Nama Dan Volume Emoji Dalam Dataset Setiap Aspek	49
Lampiran 7. Perhitungan skor sentimen contoh data dan perhitungan manualnya.....	55
Lampiran 8. Proses Cleaning Data.....	57
Lampiran 9. Hasil Klasifikasi Aspek Tanpa Penyematan Emoji.....	59
Lampiran 10. Hasil Training Dan Validation Seluruh Aspek Dengan Atau Tanpa Emoji....	60
Lampiran 11. Hasil Confussion Matrix Seluruh Aspek Tanpa Penyematan Emoji.....	62
Lampiran 12. Hasil Wordcloud Seluruh Aspek Tanpa Penyematan Emoji.....	63

BAB I PENDAHULUAN

1.1 Latar Belakang

Sejak paruh awal tahun 2020, pandemi covid-19 menciptakan tantangan baru pada sosial ekonomi serta krisis kesehatan yang belum pernah terjadi sebelumnya di Indonesia. Kebijakan yang dilakukan seperti pembatasan mobilitas telah membantu menekan penyebaran virus, tetapi juga menciptakan krisis kesehatan, sosial, dan ekonomi masyarakat. Survey penilaian yang mencakup lebih dari 12.000 rumah tangga telah dilakukan (Unicef *et al.*, 2021) oleh bimbingan Fiskal Badan Kebijakan dan Badan Pusat Statistik (BPS), hasil yang di tunjukkan bahwa dampak covid-19 membuat kondisi sosial ekonomi sangat parah, setidaknya satu anggota dari dua rumah tangga kehilangan pekerjaan. Pandemi juga menempatkan anak-anak dan remaja pada posisi yang rentan karena harus menghadapi tantangan kesulitan belajar (pembelajaran jarak jauh), dan keterbatasan sosial interaksi. Pandemi covid-19 memang menimbulkan berbagai masalah sosial, ekonomi, dan kesehatan di tengah-tengah masyarakat, sehingga keadaan yang signifikan ini dapat menjadi pemicu munculnya perasaan kecemasan, takut, tekanan mental akibat *social distancing* dan *physical distancing*, isolasi, serta ketidakpastian (Indonesia, 2020).

Kondisi ini menciptakan beban psikologis, dengan munculnya perasaan kecemasan, ketakutan, dan tekanan mental. Meskipun kecemasan adalah reaksi yang normal, tingkat kecemasan yang tinggi dapat mengganggu keseimbangan dan kualitas hidup individu (Suwandi *et al.*, 2020). Survey kesehatan mental menunjukkan bahwa sejak dimulainya pandemic hingga tahun 2022, 71,7% dari responden mengalami masalah kecemasan (PDSKJI, 2020). Kesadaran rendah akan isu kesehatan mental dan stigma negatif dapat membuat individu enggan mencari bantuan, memilih menggunakan media online sebagai outlet untuk berbagi cerita dan mencari dukungan.

Media Online, terutama Twitter, memainkan peran penting selama pandemi, menjadi saluran utama untuk mendapatkan informasi, berinteraksi, dan mengekspresikan perasaan. Postingan online, termasuk opini masyarakat menciptakan polarisasi dan dapat mencerminkan tingkat kecemasan. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini masyarakat di media online, khususnya Twitter, untuk memahami tingkat kecemasan akibat pandemic Covid-19. Analisis sentimen akan dilakukan dengan mengklasifikasikan dokumen berdasarkan kata-kata kunci yang mencerminkan sentimen positif, negatif, dan netral, dengan harapan dapat memberikan wawasan tambahan tentang dinamika kecemasan masyarakat selama pandemic Covid-19.

Orlando Troisi, Giuseppe Fenza, Mara Grimaldi, Francesca Loia (Troisi, 2022) mengangkat topik penelitian terkait kecemasan masyarakat terhadap peran teknologi di masa sebelum dan sesudah pandemi covid-19 dengan memanfaatkan metode Fuzzy Formal Concept Analysis. Hasil yang didapatkan menunjukkan bahwa probabilitas kejadian topik di dalam tweet atau postingan twitter masyarakat pada periode pasca covid meningkat 70% dibandingkan pra-covid, kata yang dianggap sebagai indikator valid untuk sentimen pasca-covid adalah kecemasan akan kemampuan, dan kesulitan budaya digital.

Pada penelitian oleh Syafrial Fachri Pane, Jenly Ramdan (Pane & Ramdan, 2022) telah menggunakan metode deep learning yaitu LSTM untuk pengujian 3000 tweet atau postingan dari Twitter terhadap kebijakan PPKM dari pemerintah untuk masyarakat pada masa pandemi covid-19, penelitian ini berhasil diklasifikasi dan mendapatkan nilai akurasi sebesar 0,917, serta peneliti menyarankan untuk penelitian selanjutnya menambahkan suatu algoritma untuk meningkatkan nilai akurasi model dengan melakukan klasifikasi sentimen emoji.

Liang-Chu Chen, Chia-Meng Lee, Mu-Yen Chen (Chen, 2020) melakukan penelitian analisis sentimen pada sosial media dengan menggunakan kedua metode yang dibahas sebelumnya, yaitu LSTM dan BiLSTM. Hasil akhir untuk akurasi dan ukuran F1 dari model LSTM mendapat nilai 84,08% dan 82,41%, sedangkan akurasi dan ukuran F1 dari metode BiLSTM mendapat nilai 92,68% dan 88,41%.

Dalam penelitian lain ada pula yang memanfaatkan metode deep learning lainnya seperti yang dilakukan oleh Rachdian Habi Yahya, Warih Maharani, Rifki Wijaya (Yahya *et al.*, 2023) untuk Analisis Sentimen Manajemen Bencana Indonesia oleh pengguna twitter menggunakan BiLSTM. Dengan Menggunakan algoritma BiLSTM hasil performansi klasifikasi di dapatkan nilai akurasi mencapai 84%.

Chuchu Liu, Fan Fang, Xu Lin, Tie Cai, Xu Tan, Jianguo Liu, Xin Lu (Liu et al., 2021) dalam penelitian nya Improving Sentimen Analysis Accuracy with Emoji Embedding mendapatkan akurasi tertinggi sekitar 0,95 atau 95% dalam menganalisis teks online berbahasa cina dengan memanfaatkan metode CEmo-LSTM yang berbasis BiLSTM dan melakukan penyematan emoji.

Penelitian oleh Eneng Eneng Tita Tosida, Erniyati, dan Krisna (Tosida *et al.*, 2021) mengevaluasi opini masyarakat terhadap kebijakan pemerintah, seperti physical distancing dan LSSR, menggunakan metode Support Vector Machine (SVM) karena memiliki tingkat akurasi tertinggi dalam klasifikasi teks. Hasil analisis dibagi menjadi empat topik, termasuk “patuh PSBB belajar dirumah”, “patuh PSBB WFH”, “patuh PSBB transportasi”, dan “patuh physical distancing”. Akurasi pakar berkisar antara 56-78%, sedangkan akurasi sistem mencapai 88-98% untuk berbagai topik kepatuhan tergantung pada jenis kebijakan.

Berdasarkan dari penelitian terdahulu, metode SVM dan BiLSTM telah terbukti efektif dalam analisis sentimen. SVM, sebagai algoritma machine learning, memiliki tingkat akurasi yang tinggi dalam klasifikasi teks. Sementara itu, BiLSTM, sebagai algoritma deep learning, memiliki keunggulan mendeteksi informasi dan memahami konteks arti kata atau kalimat, khususnya dalam menganalisis sentimen tingkat kecemasan masyarakat terkait pandemic Covid-19. Perbedaan utama antara keduanya adalah SVM merupakan algoritma machine learning dan BiLSTM dan algoritma lainnya pada penelitian terdahulu termasuk dalam algoritma deep learning. Penggunaan BiLSTM dalam menganalisis sentimen melibatkan akuisisi data dari twitter, focus pada tingkat kecemasan terkait aspek sosial, ekonomi, dan kesehatan masyarakat selama pandemi. Hasil klasifikasi mencakup sentimen positif, negatif, dan netral, yang kemudian dibedakan menjadi sentimen dengan tingkat kecemasan positif, sentimen dengan tingkat kecemasan negatif, dan sentimen dengan tingkat kecemasan netral.

Pada penelitian ini pula dibutuhkan untuk pengujian penyematan emoji pada sentimen. Hal ini dilakukan karena yang telah dibahas pada penelitian terdahulu, penyematan emoji dapat melengkapi informasi tekstual. Emoji yang secara visual signifikan dapat berguna dalam pemahaman atas emosi dari teks terkait. Dalam model analisis sentimen, penyematan emoji ini akan efektif dalam hasil yang lebih baik jika di kombinasikan dengan metode BiLSTM dalam pengujian nya.

Berdasarkan penelitian dan masalah yang telah dibahas, selanjutnya dilakukan penelitian yang akan di fokuskan guna mengetahui Implementasi Algoritma BiLSTM dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah Mengimplementasikan Algoritma BiLSTM Dengan Analisis Sentimen Menggunakan *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat.

1.3 Ruang Lingkup Penelitian

Penelitian ini memiliki ruang lingkup yang dibatasi sebagai berikut:

1. Proses pengambilan data melalui *scrapping* tweets dari Twitter dengan kata kunci terkait aspek sosial, ekonomi, dan kesehatan.
2. Pengambilan data dilakukan dengan fokus pada aspek sosial menggunakan kata kunci seperti sosial, PSBB, interaksi, stigma, dan daring; pada aspek ekonomi menggunakan kata kunci seperti ekonomi, PPKM, resesi, inflasi; dan pada aspek kesehatan menggunakan kata kunci seperti kesehatan mental, cemas, covid-19, vaksin, dan pelayanan kesehatan, dan pengambilan data dilakukan selama periode 1 hingga 31 desember 2022.
3. Data tweets yang diambil dijumlahkan berdasarkan aspeknya, dengan jumlah total mencapai 3.707 data.
4. Dilakukan penyematan emoji pada data tweets, dengan jumlah jenis emoji yang ditemukan mencapai 223 jenis dari total 3.570 jenis emoji pada kamus Unicode. Jenis serta *Volume* emoji yang terdapat pada dataset setiap aspek bisa dilihat pada Lampiran 6.
5. Dilakukan analisis terhadap data yang sama tanpa penyematan emoji untuk perbandingan hasil akhir evaluasi.
6. Penggunaan metode Bidirectional Long Short-Term Memory (BiLSTM) dalam penelitian ini.
7. Penambahan pengujian K-Fold Cross Validation pada klasifikasi BiLSTM untuk membandingkan akurasi dengan hasil klasifikasi BiLSTM tanpa pengujian K-Fold Cross Validation.
8. Perancangan aplikasi menggunakan bahasa pemrograman Python dan menggunakan aplikasi Jupyter Notebook (Anaconda3).
9. Output aplikasi berupa visualisasi grafik jumlah sentiment yang telah diklasifikasi, visualisasi wordcloud, dan grafik hasil akurasi validasi setiap aspek.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini sebagai berikut:

1. Mendapatkan informasi mengenai kecemasan akan kehidupan sosial-ekonomi dan pandemi Covid-19 di Twitter untuk menklasifikasikannya menjadi sentimen positif, negatif, atau netral.
2. Memprediksi tingkat kecemasan terhadap Covid-19 dan kondisi sosial-ekonomi masyarakat pada Desember 2022 serta menemukan metode analisis sentimen yang akurat di Twitter.
3. Menjadi referensi untuk penelitian selanjutnya, khususnya dalam text mining, Bidirectional Long Short-Term Memory (BiLSTM), dan penggunaan emoji dalam analisis sentimen.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Tingkat Kecemasan

Menurut Kaplan, Sadock, dan Grebb (Grebb, 2010), kecemasan merupakan respon pada suatu situasi tertentu yang mengancam, dan hal tersebut normal terjadi mengiringi perkembangan, perubahan, pengalaman baru terhadap hal yang belum pernah dilakukan, serta dalam proses menemukan identitas diri dan arti hidup. Kecemasan merupakan reaksi yang dapat dialami siapapun. Namun cemas yang berlebih, apalagi yang telah menjadi gangguan dapat menghambat seseorang dalam keseimbangan hidupnya.

WHO sebagai badan kesehatan dunia mendefinisikan kesehatan individu secara menyeluruh yaitu sehat secara fisik, mental, sampai dengan sosial. Dalam langkah ideal upaya penanganan covid-19 seharusnya tidak hanya melakukan penanganan pada kesehatan fisik saja, tetapi perlu penanganan pula untuk kesehatan mental dan sosial, sehingga perlu untuk diketahui seberapa besar kecemasan yang disebabkan karena pandemi covid-19 pada masyarakat (Muyasaroh, 2020)

Kecemasan adalah salah satu gangguan kesehatan mental dengan jumlah kasus yang tinggi bersama dengan gangguan kesehatan mental lainnya yaitu depresi. Pada Mei 2020 dengan melibatkan 393 responden dengan rentan umur 16-24 tahun sebanyak 95% dinyatakan bahwa individu tersebut mengalami gejala kecemasan, 88% pernah mengalami gejala depresi dalam mengalami permasalahan selama hidupnya di rentan usia tersebut, dan lebih dari 90% dari rasa kecemasan dan depresi yang dialami oleh individu disebabkan oleh kesulitan keuangan, akademik, hingga perasaan kesepian (Kaligis et al., 2021).

2.1.2 Text Mining

Text Mining merupakan suatu algoritma untuk mencari sebuah informasi dari sebuah teks yang terstruktur. *Text Mining* dapat membantu pengguna untuk mengetahui isi dari sebuah bacaan tanpa harus membaca seluruh isi dari bacaan tertentu. Pengelolaan pada sumber data pada *Text Mining* dilakukan dalam beberapa tahapan. Tahapan ini dikenal dengan istilah *Text Processing*. *Text Processing* terdiri dari beberapa tahapan, yaitu *Transform Cases*, *Filter Stopword*, dan *Filter Tokenize*, dan *Stemming* (Flores et al, 2020).

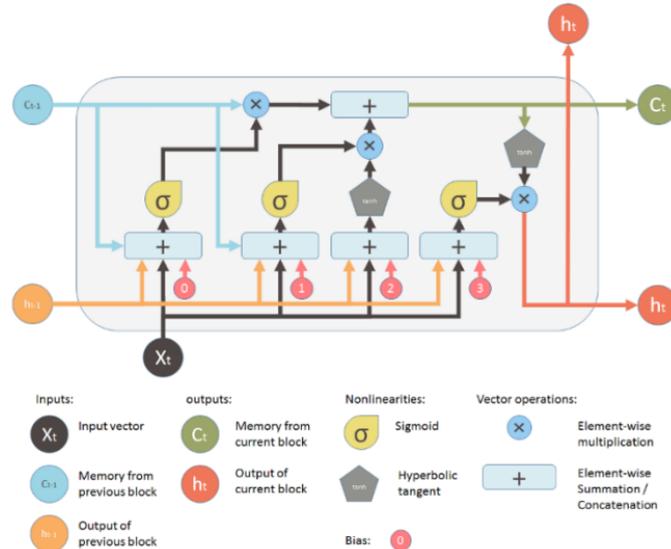
2.1.3 Analisis Sentimen

Analisis sentimen adalah bidang studi yang menganalisis suatu pendapat, sentimen, evaluasi, sikap, dan emosi individu terhadap suatu produk, peristiwa, topik, atau isu yang diungkapkan dengan bahasa tulisan (Liu et al., 2021). Analisis sentimen merupakan salah satu teknik yang digunakan untuk mengetahui opini publik dengan mengekstrak informasi berupa sikap seseorang terhadap suatu isu atau kejadian dengan mengelompokkan polaritas dari sebuah teks. Pengelompokan yang dilakukan adalah untuk melihat apakah teks tersebut menyiratkan sentimen positif, negatif, atau netral (Aditama et al., 2020).

Analisis sentimen yang menggunakan media sosial sebagai tempat mengambil data akan memberi wawasan mengenai konsepsi, sikap, dan perilaku untuk pengambilan keputusan penting bagi para pelaku bisnis, politik, serta perwakilan masyarakat. Pandemi covid-19 memberi dampak buruk bagi manusia dan negara. Kebutuhan medis dan perawatan kesehatan diperlukan, serta sangat penting untuk melindungi individu dan masyarakat dari tekanan psikologis (kecemasan, ketakutan, penyakit mental, dsb (Hung et al., 2020).

2.1.4 Bidirectional Long Short-Term Memory (BiLSTM)

Long Short-Term Memory (LSTM) adalah pengembangan arsitektur *Recurrent Neural Network* (RNN) untuk menangani masalah *Vanishing Gradient*, dimana kemiringan fungsi kerugian menurun secara eksponensial pada saat memproses data sekuensial yang panjang (Rumelhart *et al*, 1985) Hal ini menyebabkan RNN gagal menangkap long term dependencies sehingga dapat mengurangi performa prediksi. LSTM mengganti lapisan RNN dengan *blok memory cell* menggunakan mekanisme gerbang yang terdiri dari *forget gate*, *input gate*, dan *output gate* (Saxena, 2018). Sama halnya dengan RNN, LSTM tersusun atas neuron yang diproses secara berulang. Struktur neuron tunggal pada LSTM dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi LSTM (Rizki *et al*, 2020)

Dari Gambar 1 diatas, LSTM memiliki dua bagian cell diantaranya adalah *Cell State* dan *hidden state*. *Cell State* merupakan jaringan utama aliran data yang memungkinkan data mengalir pada dasarnya tidak berubah. Namun beberapa transformasi linear dapat terjadi. Data dapat ditambahkan atau dihapus dari *Cell State* melalui sigmoid gates. Gates tersebut mirip dengan lapisan atau serangkaian operasi matriks, yang berisi bobot individu yang berbeda. LSTM dirancang untuk menghindari masalah long-term dependency (Athallah, 2021).

LSTM memiliki rumus untuk perhitungan dalam setiap layernya, berikut rumus-rumus pada LSTM:

a) *Forget Gate*

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (1)$$

b) *Input Gate*

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (2)$$

c) *Memory Update*

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \quad (3)$$

$$C_t = f_t \cdot c_{t-1} + i_t \cdot C_t \quad (4)$$

d) *Output Gate*

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (5)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (6)$$

Rumus – rumus layer pada LSTM menggunakan fungsi aktivasi *sigmoid* dimana fungsi aktivasi merupakan fungsi yang menggambarkan hubungan antara tingkat aktivitas internal yang berbentuk linear maupun non-linear. Dengan tujuan untuk menentukan *neuron* tersebut diaktifkan atau tidak. Fungsi aktivasi yang digunakan secara umum yaitu *sigmoid*, *tanh*, dan *softmax*. Penjelasan mengenai fungsi aktivasi tersebut adalah sebagai berikut:

1. Sigmoid

Fungsi aktivasi *sigmoid* merupakan sebuah fungsi non-linear. Masukkan untuk fungsi aktivasi ini adalah bilangan real dan *output* dari fungsi tersebut memiliki range antara 0 sampai 1, dengan perhitungan fungsi aktivasi *sigmoid* sebagai berikut:

$$g(x) = \frac{1}{1+e^{-x}} \quad (7)$$

Fungsi ini mentransformasikan range nilai dari *input* x menjadi anatara 0 dan 1. Jika masukannya negatif, maka keluaran yang akan didapatkan adalah 0. Sedangkan jika masukannya positif, maka keluaran yang akan didapatkan adalah 1.

2. Tanh

Fungsi aktivasi tanh merupakan fungsi non-linier. Masukkan untuk fungsi aktivasi ini berupa bilangan real dan *output* dari fungsi ini memiliki range antara -1 sampai 1, dengan perhitungan fungsi aktivasi tanh sebagai berikut:

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

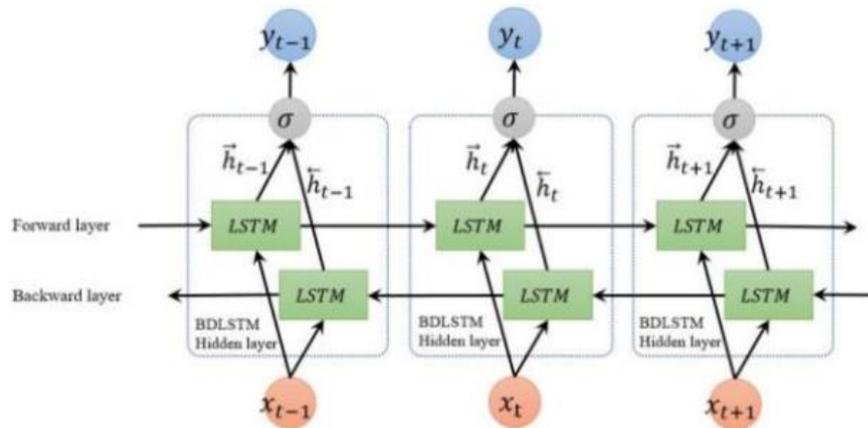
Dalam pengaplikasiannya fungsi *tanh* lebih menjadi pilihan jika dibandingkan dengan fungsi sigmoid dikarenakan fungsi ini memiliki *output* yang merupakan *zero-centered*.

3. Softmax

Fungsi aktivasi *softmax* adalah rentang probabilitas *output* dengan nilai 0 hingga 1 dan jumlah semua probabilitas akan sama dengan satu. Dengan perhitungan fungsi aktivasi *softmax* berikut:

$$g(x) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (9)$$

Salah satu kelemahan LSTM adalah tidak cukup memperhitungkan informasi dari kata terakhir karena hanya membaca kalimat dari satu arah saja, awal ke akhir. Oleh karena nya, terdapat perkembangan dari model LSTM yaitu BiLSTM. BiLSTM memiliki dua lapisan yang prosesnya saling berkebalikan arah, model ini sangat baik untuk mengenali pola dalam kalimat karena setiap kata dalam dokumen diproses secara sekuensial, karena data dapat dipahami bila pembelajaran secara berurut setiap kata. Lapisan dibawahnya bergerak maju (*forward*), yaitu memahami dan memproses dari kata pertama menuju kata terakhir sedangkan lapisan diatasnya bergerak mundur (*backward*), yaitu memahami dan memproses dari kata terakhir menuju kata pertama. Dengan adanya lapisan dua arah yang saling berlawanan ini maka model dapat memahami dan mengambil perspektif dari kata terdahulu dan kata terdapan, sehingga proses pembelajaran semakin dalam yang berdampak pada model akan lebih memahami konteks pada data tersebut. Secara Teknis, BiLSTM menerapkan dua LSTM terpisah, satu untuk arah depan dan satu untuk arah mundur (Fadli & Hidayatullah, 2021). Ilustrasi struktur Algoritma BiLSTM dapat dilihat pada Gambar 2.



Gambar 2. Ilustrasi BiLSTM

Setiap hidden unit keluaran unit pada lapisan bawah dan atas digabungkan membentuk nilai fitur kata tersebut dengan ukuran lebih panjang daripada menggunakan LSTM biasa. Karena lebih panjang nilai fitur, maka informasi yang akan di proses pada tahap selanjutnya yaitu feed *forward* akan mengklasifikasikan dengan lebih akurat.

2.1.5 Emoji Embedding

Media sosial seperti twitter telah di eksplorasi karena pengguna yang masif menghasilkan konten dengan sentimen pengguna, di mana emoji juga digunakan secara luas. Emoji merupakan karakter gambar yang biasa digunakan dalam pesan elektronik maupun postingan seseorang di dalam media sosial (Chen et al., 2018). Model ini dapat digunakan dalam analisis sentimen karena setiap emoji dapat mengekspresikan satu atau lebih emosi. Meluasnya penggunaan media sosial membuat kebutuhan akan analisis emosi dan analisis psikologis pengguna media sosial merupakan hal yang perlu dan sangat penting bagi pemerintah dan organisasi, yang meningkatkan pentingnya prediksi emoji untuk analisis emosi. Bila model ini berprinsip meningkatkan tingkat akurasi analisis, metode penyematan kata berbasis emosi yang disebut penyematan emoji atau *emoji embedding* dapat melakukannya dengan cara mengekstrak keduanya, semantik dan emosi kata-kata (Tavan et al., 2020). Jumlah emoji pada kamus Unicode terdapat 3.570 jenis, namun hanya 223 jenis emoji yang dipakai dalam penelitian. Berikut pada Tabel 1 merupakan beberapa emoji yang dipakai dalam penelitian ini.

Tabel 1. Contoh Emoji dan Deskripsi Nama Labelnya

No.	Emoji	Deskripsi
1.	😭	loudly crying face
2.	😊	smiling face
3.	😄	face with tears of joy
4.	😡	angry face
5.	☹️	frowning face
⋮	⋮	⋮
100.	🙏	folded hands
⋮	⋮	⋮
223.	❤️	red heart

2.1.6 Confusion Matrix

Confusion matrix juga sering disebut *error matrix*. Pada dasarnya *confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui. *Confusion matrix* terdiri dari tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi (Setyo Nugroho *et al.*, 2021).

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <small>Type I Error</small>
	0 (Negative)	FN (False Negative) <small>Type II Error</small>	TN (True Negative)

Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).

1. *True Positive* (TP) adalah data positif yang diprediksi positif oleh sistem.
2. *True Negative* (TN) adalah data negatif yang diprediksi negatif oleh sistem.
3. *False Positive* (FP) adalah data negatif yang diprediksi positif oleh sistem.
4. *False Negative* (FN) adalah data positif yang diprediksi negatif oleh sistem.

Adapun beberapa parameter yang biasanya digunakan dalam mengukur performansi suatu metode antara lain adalah *Precision*, *Recall*, dan *Accuracy*.

1. *Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$Precision = \frac{TP}{FP+TP} \quad (10)$$

2. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Recall = \frac{TP}{FN+TP} \quad (11)$$

3. *Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (12)$$

4. *F - score* atau f-measure digunakan untuk melakukan perbandingan rata – rata presisi dan recall. Dapat dihitung dengan persamaan berikut.

$$F - score = 2 \cdot \frac{Precision * Recall}{Precision+Recall} \quad (13)$$

2.1.7 K-Fold Cross Validation

K-Fold Cross Validation merupakan suatu teknik dalam data mining yang bertujuan untuk mencapai tingkat akurasi optimal dengan membagi data menjadi dua subset, yaitu data latih dan data uji. Salah satu metode Cross Validation yang umum digunakan adalah

K-Fold Cross Validation. Metode ini berfungsi untuk mengevaluasi kinerja suatu algoritma dengan cara membagi dataset secara acak menjadi sejumlah K partisi, kemudian mengelompokkan data tersebut. dalam pendekatan K-Fold Cross Validation, dataset dibagi menjadi beberapa partisi secara acak, dan proses eksperimen dilakukan sebanyak K kali. Setiap eksperimen menggunakan satu partisi sebagai data uji dan menggunakan sisa partisi lainnya sebagai data latih.

2.2 Penelitian Terdahulu

Berikut merupakan penelitian terdahulu yang dipakai untuk mendukung penelitian penulis.

1.	Judul	: Covid-19 sentiments in smart cities: The role of technology anxiety before and during the pandemic.
	Peneliti	: Orlando Troisi, Giuseppe Fenza, Mara Grimaldi, Francesca Loia. (2021).
	Isi	: penelitian ini menggunakan metode Fuzzy Formal Concept Analysis (FFCA). Hasilnya menunjukkan peningkatan probabilitas kejadian topik dalam tweet selama periode pasca Covid sebesar 70%. Indikator valid sentimen pasca Covid melibatkan akan kemampuan dan kesulitan budaya digital.
2.	Judul	: Exploration of social media for sentiment analysis using deep learning.
	Peneliti	: Liang-Chu Chen, Chia-Meng Lee, Mu-Yen Chen (2019).
	Isi	: Analisis sentimen menggunakan metode LSTM dan BiLSTM pada media sosial menunjukkan hasil akhir. Model LSTM memiliki akurasi 84,08% dan F1 82,41%, sementara BiLSTM mencapai akurasi 92,68% dan F1 88,41%.
3.	Judul	: Improving Sentimen Analysis Accuracy with Emoji Embedding.
	Peneliti	: Chuchu Liu, Fan Fang, Xu Lin, Tie Cai, Xu Tan, Jianguo Liu, Xin Lu (2021).
	Isi	: pengujian menyatakan bahwa penerapan emoji meningkatkan akurasi analisis sentimen. Metode CEmo-LSTM mencapai akurasi tertinggi sekitar 95% dalam menganalisis teks online berbahasa Cina.
4.	Judul	: Disaster Management Sentiment Analysis Using the BiLSTM Method
	Peneliti	: Rachdian Habi Yahya, Warih Maharani, Rifki Wijaya (2023).
	Isi	: penelitian ini menggunakan metode deep learning BiLSTM untuk menguji 2.686 tweet terkait manajemen bencana lingkungan di Indonesia. Hasilnya, dengan pembagian data sebesar 80:20, penelitian berhasil mengklasifikasi dengan akurasi 84%.
5.	Judul	: Sentimen Analysis Using the Support Vektor Machine For Community Compliance Representation in The Covid-19 Pandemic Period.
	Peneliti	: Eneng Tita Tosida, Erniyati, & Krisna (2021).
	Isi	: akurasi pakar dan sistem untuk berbagai query adalah sebagai berikut: "Patuh Belajar Dirumah" pakar 56%, sistem 98%; "Patuh WFH" pakar 72%, sistem 93%; "Patuh PSBB Transportasi" pakar 58%, sistem 88%; dan "Patuh Physical Distancing" pakar 78%, sistem 90%.

2.2 Tabel Perbandingan Penelitian

Tabel perbandingan penelitian terdahulu dapat dilihat pada Tabel 2.

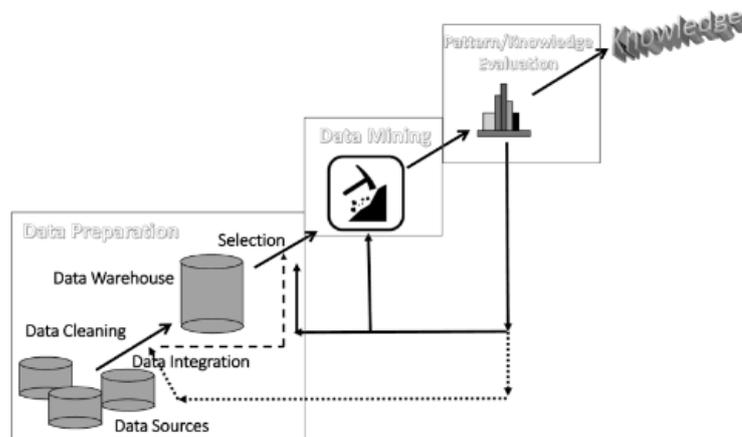
Tabel 2. Tabel Perbandingan Penelitian Terdahulu

No.	Penelitian	Output		Metode Penelitian							Topik Analisis Sentimen	Jumlah data	Aspek	Penyematan emoji	Tingkat Akurasi
		Model	WEB	SVM	LSTM	BiLSTM	FFCA	Emoji Embedding	CNN	CEmo-LSTM					
1.	Orlando, dkk (2021).	✓					✓				kecemasan teknologi melalui analisis sentimen publik.	32.334	Teknologi	-	70%
2.	Liang-Chu Chen, dkk (2019).	✓			✓	✓					Analisis sentimen pada media sosial Ptt.cc (PTT) milik Taiwan.	17.819	Sosial	-	82% - 93%
3.	Chuchu Liu, dkk (2021).	✓						✓		✓	Analisis sentimen individu yang pasif (mis. Kesedihan) akibat covid-19.	38.183.194	Kesehatan	15.609 .843 emoji	95%
4.	Rachdian Habi Yahya, Warih Maharani, Rifki Wijaya (2023).	✓				✓					Analisis sentimen manajemen bencana di Indonesia.	2.686	Lingkungan	-	84%
5.	Eneng Tita Tosida, Erniyati, & Krisna (2021).	✓		✓							Analisis opini masyarakat terhadap sistem yang diterapkan oleh pemerintah yaitu <i>physical distancing</i> dan PSBB.	3.797	Sosial	-	88% - 98%
6.	Jenie Marcelina (2022).	✓				✓		✓			Analisis sentimen tingkat kecemasan terhadap covid-19, dan sosial ekonomi masyarakat	3.707	Sosial, Ekonomi, & Kesehatan	4.600 emoji	98,08%

BAB III METODE PENELITIAN

3.1 Metode Penelitian

Metode yang diterapkan untuk penelitian ini yaitu menggunakan data mining atau dapat disebut juga *Knowledge Discovery and Data Mining* atau KDD. Metode KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan suatu pola atau hubungan dalam set data berukuran besar (Han *et al.*, 2022). Sebagai sebuah suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap dan metode KDD dapat dilihat pada Gambar 3.



Gambar 3. Tahapan KDD (Han *et al.*, 2022)

KDD adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data. Pola yang ditemukan bersifat baru, dapat bermanfaat, dan dapat dimengerti. Dalam penelitian ini dilakukan langkah – langkah atau alur penelitian yang meliputi pengumpulan data Twitter lalu data tersebut dilakukan *Preprocessing*, setelah itu dilakukan *word embedding* dengan *one hot encode* dan dilakukan klasifikasi dengan BiLSTM, kemudian sistem tersebut diuji dengan menggunakan *confusion matrix*.

3.1.1 Tahap Pengumpulan Data

Dalam tahapan pengumpulan data, data dapat dikumpulkan dari berbagai sumber sesuai dengan kebutuhan analisis, yang kemudian akan diseleksi kembali. Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

3.1.2 Pembersihan Data dan Integrasi Data

Pembersihan data atau *cleaning data* merupakan proses menghilangkan noise, data yang tidak relevan dan konsisten. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi focus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3.1.3 Seleksi Data dan Transformasi Data

Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada yang ingin dicapai. Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

3.1.3.1 Word Embedding

Word embedding merupakan Setelah didapatkan hasil *Preprocessing*, tahap selanjutnya adalah pembobotan kata yaitu transformasikan kata menjadi numerik atau vector.

3.1.4 Proses Data Mining

Proses data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

3.1.4.1 Metode *Bidirectional Long Short-Term Memory*

Merupakan tahapan untuk menentukan analisis sentimen dengan *emoji embedded* terhadap tingkat kecemasan covid-19 dan sosial ekonomi masyarakat, dari komentar pengguna akan di klasifikasi masuk kedalam sentimen dengan tingkat kecemasan positif, sentimen dengan tingkat kecemasan negatif, atau sentimen dengan tingkat kecemasan netral.

3.1.5 Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola menarik yang akan ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

3.2 Flowchart Penelitian



Gambar 4. Diagram Alur Penelitian

3.3 Waktu dan Tempat Pelaksanaan

Penelitian dilaksanakan mulai dari bulan November 2022 s.d bulan Januari 2022 (3 bulan). Tempat pelaksanaan penelitian di lokasi Laboratorium Ilmu Komputer Program Studi Ilmu Komputer FMIPA Universitas Pakuan Bogor yang beralamat di Gedung MIPA 2. Jl. Pakuan, RT.02/RW.06, Tegallega, Kecamatan Bogor Tengah, Kota Bogor, Jawa Barat 16129.

3.4 Alat dan Bahan

3.4.1 Alat

1. *Hardware*
 - a. Personal Computer
 - b. AMD A9-9420 RADEON R5, 5 COMPUTE CORES 2C + 3G 3.0GHZ
 - c. RAM 8,00 GB
2. *Software*
 - a. Google Chrome
 - b. *Jupyter Notebook* (Anaconda Navigator)
 - c. Ms. Office 2016
 - d. Ms. Excel 2016
 - e. Ms. Visio 2016

3.4.2 Bahan

Bahan – bahan yang digunakan dalam penelitian ini yaitu,

1. Data twitter pada bulan Desember 2022 dan data portal berita mengenai tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat.
2. Jurnal, skripsi, dan buku sebagai referensi dalam pembuatan laporan penelitian.

BAB IV PERANCANGAN DAN IMPLEMENTASI

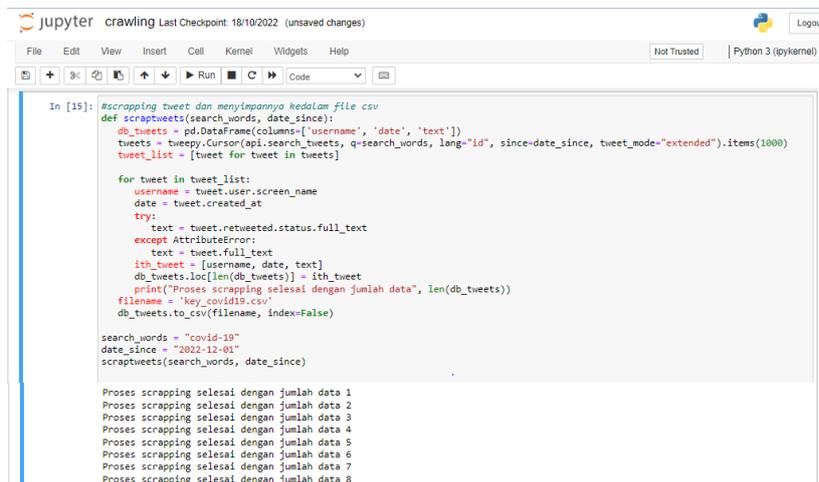
4.1 Perancangan Sistem

Dalam tahapan ini bertujuan untuk memberikan gambaran mengenai perancangan sistem atau rancangan desain yang akan dikembangkan. Sistem menggunakan kaidah *Knowledge Discovery in Databases* (KDD). Gambaran umum sistem yang akan dibangun menggunakan beberapa *package library* yang disediakan oleh Python. Diagram alur atau flowchart merupakan bagan yang menunjukkan alur kerja yang sedang dikerjakan di dalam sistem secara keseluruhan dan menjelaskan urutan dari prosedur-prosedur yang ada dalam sistem. Gambar 4 adalah diagram alur atau flowchart program implementasi algoritma BiLSTM dengan analisis sentimen menggunakan *emoji embedded* terhadap tingkat kecemasan covid-19 dan sosial ekonomi masyarakat.

4.1.1 Tahap Pengumpulan Data (*Data Selection*)

Pengumpulan data dilakukan dengan cara melakukan pengambilan data dari media sosial Twitter dengan menggunakan *crawling* data yang dicari berdasarkan *keyword* yang dibutuhkan dalam tiga aspek yaitu aspek sosial, ekonomi, dan kesehatan. Penentuan aspek berdasarkan topik yang sering dibicarakan di media sosial Twitter. Kemudian dilakukan wawancara dan validasi pada pakar guru bahasa Indonesia. Adapun profil biodata pakar dapat dilihat pada lampiran 5 dan hasil wawancara dapat dilihat pada lampiran 4.

Berikut adalah hasil *crawling* data twitter menggunakan *Tweepy* Anaconda Notebook disimpan dalam bentuk file csv. Data dicari berdasarkan *keyword* yang diinputkan yaitu berdasarkan aspek-aspek dan kata kunci yang telah ditentukan. Untuk proses *crawling* dapat dilihat pada Gambar 5 dan proses *crawling* untuk aspek lainnya dapat dilihat pada lampiran 3.



```
In [15]: #scrapping tweet dan menyimpannya kedalam file csv
def scraptweets(search_words, date_since):
    db_tweets = pd.DataFrame(columns=['username', 'date', 'text'])
    tweets = tweepy.Cursor(api.search_tweets, q=search_words, lang="id", since=date_since, tweet_mode="extended").items(1000)
    tweet_list = [tweet for tweet in tweets]

    for tweet in tweet_list:
        username = tweet.user.screen_name
        date = tweet.created_at
        try:
            text = tweet.retweeted.status.full_text
        except AttributeError:
            text = tweet.full_text
        ith_tweet = [username, date, text]
        db_tweets.loc[len(db_tweets)] = ith_tweet
        print("Proses scrapping selesai dengan jumlah data", len(db_tweets))
    filename = 'key_covid19.csv'
    db_tweets.to_csv(filename, index=False)

search_words = "covid-19"
date_since = "2022-12-01"
scraptweets(search_words, date_since)

Proses scrapping selesai dengan jumlah data 1
Proses scrapping selesai dengan jumlah data 2
Proses scrapping selesai dengan jumlah data 3
Proses scrapping selesai dengan jumlah data 4
Proses scrapping selesai dengan jumlah data 5
Proses scrapping selesai dengan jumlah data 6
Proses scrapping selesai dengan jumlah data 7
Proses scrapping selesai dengan jumlah data 8
```

Gambar 5. Proses Crawling Data

Sebagai contoh kita ingin *crawling* data twitter dengan kata kunci “covid-19” dengan tweet yang dimulai dari tanggal 1 Desember 2022 dengan maksimal jumlah tweet adalah 1000, lalu hasil *crawling* akan disimpan dengan format nama *key_covid19.csv*. Kemudian dicari kata kunci lain sesuai kebutuhan. Data hasil *scrapping* setiap kata kunci yang didapatkan lalu dikelompokkan dalam aspek terkait dan proses pengecekan dilakukan secara manual. Setelah proses seleksi, kata kunci kemudian divalidasi oleh pakar Bahasa Indonesia. Data berdasarkan kata kunci terkait kemudian disatukan dalam satu aspek yang

disimpan dalam format .csv, sehingga didapatkan hasil akhir jumlah data masing-masing kata kunci dapat dilihat dalam Tabel 3.

4.1.2 Pembersihan Data dan Integrasi Data

Pembersihan data atau *cleaning data* merupakan proses menghilangkan noise, data yang tidak relevan dan konsisten. Serta integrasi data merupakan penggabungan data dari beberapa sumber. Data yang digunakan dalam penelitian ini merupakan hasil *scrapping* dari media sosial twitter serta dari portal berita. Kemudian kita masukkan data sesuai hasil pencarian *keyword* tersebut yang telah dicari. Kemudian data tersebut dikumpulkan untuk dilakukan penelitian.

Pengumpulan data dilakukan dengan cara melakukan *scrapping* dari media sosial twitter dan portal berita dengan menggunakan *tool scrapping* melalui anaconda Navigator Notebook Developer Twitter. Data yang dicari berdasarkan *keyword* yang di *input*-kan yaitu berdasarkan aspek sosial, ekonomi, dan kesehatan pada tahap pengumpulan data (*data selection*) sebelumnya sehingga didapatkan hasil akhir jumlah data berikut.

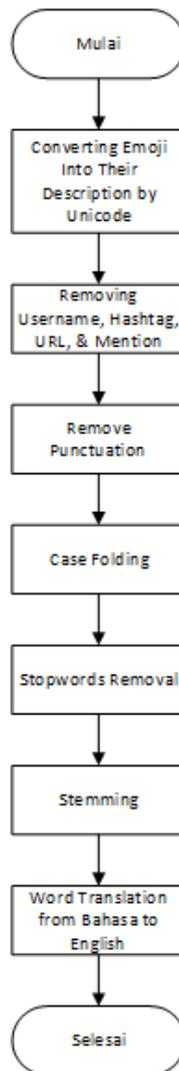
Tabel 3. Keterangan Jumlah Data Berdasarkan Setiap Aspek Dalam Penelitian

Aspek	Kata Kunci	Jumlah
Sosial	sosial, PSBB, interaksi, stigma, dan daring.	1.115
Ekonomi	ekonomi, PPKM, resesi, inflasi.	1.288
Kesehatan	kesehatan mental, cemas, covid-19, vaksin dan pelayanan kesehatan.	1.304
Total Data		3.707

Data yang dijadikan sampel dalam penelitian ini berjumlah 5 diambil dari twitter, Dimana 1 data *tweet* yang akan dijadikan sampel *preprocessing* dan perhitungan manualisasi nya. Contoh data yang akan dilakukan *preprocessing* dan perhitungan manualisasi nya dapat dilihat dibawah ini:

arelxy_,2022-12-13 15:11:53+00:00,Para pemimpin dan kabinet banyak yang mundur serta hura hura di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan 😞

Pada umumnya data yang diperoleh memiliki isi yang tidak sempurna seperti atribut yang tidak relevan. Data yang diperoleh belum sepenuhnya siap digunakan untuk proses klasifikasi secara langsung karena data masih tidak terstruktur dengan baik dan masih terdapat banyak noise. Data masih memuat angka, tanda baca, serta kata-kata yang tidak bermakna, maka itu perlu dilakukan *preprocessing* yang bertujuan untuk menyeragamkan bentuk kata, menghilangkan karakter selain huruf, dan mengurangi volume kosakata sehingga data lebih terstruktur. Dalam *preprocessing*, terdapat tahapan proses yang dimulai dari *converting emoji*, *removing username*, *hashtag*, *URL*, & *mention*, *remove punctuation*, *case folding*, *stopwords removal*, *stemming*, dan kemudian data perlu di translate dari bahasa Indonesia kedalam bahasa Inggris. Tahapan *Preprocessing* yang dilakukan dapat dilihat pada Gambar 6.



Gambar 6. Tahap Preprocessing

a) *Converting Emoji*

Proses penghilangan noise pada teks antara lain meliputi *username*, *hashtag*, *URL*, *mention*, tanda baca, dan karakter non-alfanumerik lainnya. Karena penelitian ini juga berfokus pada emoji, maka keberadaannya perlu dipertimbangkan. Sebelum dilakukan untuk menghilangkan noise dan karakter non-alfanumerik pada data, pilihan terbaik adalah mengubah emoji menjadi nama label atau deskripsi nama spesifiknya berdasarkan penamaan di kamus Unicode karena emoji adalah karakter non-alfanumerik. Konversi ini dilakukan dalam tahap awal sebelum menghapus URL, nama pengguna, dan juga sebelum menghapus karakter non-alfanumerik. Terdapat 3.570 jenis emoji di dalam kamus Unicode. Berikut contoh hasil emoji yang ditemukan dalam data aspek kesehatan berdasarkan dari banyaknya frekuensi emoji tersebut, dapat dilihat pada Tabel 4 dan variasi serta jumlah emoji yang ditemukan lainnya didalam data tweet setiap aspek dapat dilihat pada lampiran 6.

Tabel 4. Jenis Emoji Yang Ditemukan Dalam Aspek Kesehatan

No.	Emoji	Label Nama / Deskripsi Emoji	Jumlah Emoji
1	😭	loudly crying face	138

2	😭	face with tears of joy	97
3	🤪	rolling on the floor laughing	89
⋮	⋮	⋮	⋮
10	😄	grinning face	48
⋮	⋮	⋮	⋮
100	😞	dissapointed face	1
⋮	⋮	⋮	⋮
138	🙇	person bowing	1

Untuk menemukan teks berisi emoji, hal ini dilakukan secara manual yang kemudian emoji akan di ekstrasi secara manual pula dengan mengubah emoji menjadi nama label atau deskripsi nya berdasarkan penamaan di kamus Unicode. Mengubah emoji menjadi nama label atau deskripsi nya merupakan cara terbaik yang digunakan karna mempertimbangkan bahwa beberapa emoji mungkin tidak dapat dikelompokkan dalam kategori sentimen tertentu, dan juga agar data teks serta emoji dapat terintegrasi dengan baik (Arifiyanti & Wahyuni, 2020). Bentuk emoji yang sudah menjadi label nama atau deskripsi nya berbentuk sebuah teks, maka karena itu emoji tidak akan terhapus pada pembersihan karakter non-alfa numerik di proses *cleaning* data selanjutnya. Hasil ekstrasi emoji menjadi nama label atau deskripsi dalam contoh data dapat dilihat pada Tabel 5.

Tabel 5. Covertng Emoji

Converting Emoji	
arelxy_,2022-12-13 15:11:53+00:00,Para pemimpin dan kabinet banyak yang mundur serta hura hura di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan 😞	arelxy_,2022-12-13 15:11:53+00:00,Para pemimpin dan kabinet banyak yang mundur serta hura hura di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan worried face

b) *Remove username, hashtag, URL, & mention*

URL yang terdapat pada data hasil *scrapping* Twitter membuat data tidak efektif dan tidak memiliki arti. Untuk itu perlu adanya penghapusan URL tersebut. kemunculan alamat web atau URL ini disebabkan karena banyaknya pengguna mempromosikan sebuah produk pada situs mereka sehingga pengguna lain bisa langsung masuk ke halaman web yang dituju. Simbol “@” atau symbol *mention* digunakan untuk menunjuk atau mengajak teman berkomunikasi langsung. Pada suatu analisis sentimen, nama pengguna tidak diperhatikan sehingga perlu dihapus. Hashtag merupakan suatu penunjuk sebuah kata yang dibicarakan oleh sesame pengguna twitter yang memiliki symbol “#”. Biasanya akan digunakan sebagai judul topik pembicaraan dan juga berfungsi sebagai pengelompokkan terhadap percakapan yang berhubungan dengan kata yang diberi simbol hashtag. Hasil *Remove username, hashtag, URL, & Proses remove username, hashtag, url, & mention* dapat dilihat pada Tabel 6.

Tabel 6. Remove Username, hashtag, URL, & Mention

Remove username, hashtag, URL, & mention	
<p>arelxy_,2022-12-13 15:11:53+00:00,Para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan worried face</p>	<p>Para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan worried face</p>

c) *Remove Punctuation*

Menghapus semua karakter non-alphabet misalnya symbol, spasi, dan lainnya. Proses *remove punctuation* dapat dilihat pada Tabel 7.

Tabel 7. Remove Punctuation

Remove Punctuation	
<p>Para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan worried face</p>	<p>Para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi apalagi inflasi termasuk lima besar se asia tenggara dan masuk besar negara yang mengalami krisis pangan worried face</p>

d) *Case Folding*

Case Folding adalah proses penyeragaman bentuk huruf dimana dalam proses ini hanya menerima huruf latin antara “a” sampai “z”. karakter lain selain huruf dianggap sebagai delimitter sehingga karakter tersebut akan dihapus dari dokumen. Kemudian penyeragaman dilakukan dengan mengubah isi dokumen menjadi huruf kecil secara keseluruhan (dari “a” sampai dengan “z”). hal ini bertujuan agar kata yang ditulis dengan huruf awal capital dan huruf non-kapital tidak terdeteksi memiliki arti yang berbeda. Proses *case folding* dapat dilihat pada Tabel 8.

Tabel 8. Case Folding

Case Folding	
<p>Para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi apalagi inflasi termasuk lima besar se asia tenggara dan masuk besar negara yang mengalami krisis pangan worried face</p>	<p>para pemimpin dan kabinet banyak yang mundur serta hura hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi apalagi inflasi termasuk lima besar se asia tenggara dan masuk besar negara yang mengalami krisis pangan worried face</p>

e) Tokenization

Pada tahap tokenisasi bertujuan untuk membuat kalimat dibagi menjadi beberapa bagian menjadi dalam bentuk kata atau frasa dalam sebuah dokumen. Proses tokenisasi dapat dilihat pada Tabel 9.

Tabel 9. Tokenization

Tokenization	
para pemimpin dan kabinet banyak yang mundur serta hura hura di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi apalagi inflasi termasuk lima besar se asia tenggara dan masuk besar negara yang mengalami krisis pangan worried face	'para', 'pemimpin', 'dan', 'kabinet', 'banyak', 'yang', 'mundur', 'serta', 'hura', 'hura', 'di', 'setiap', 'negara', 'dan', 'kami', 'sudah', 'katakan', 'tidak', 'ada', 'yang', 'bisa', 'keluar', 'dari', 'tekanan', 'tatanan', 'dunia', 'baru', 'atas', 'kehancuran', 'ekonomi', 'apalagi', 'inflasi', 'termasuk', 'lima', 'besar', 'se', 'asia', 'tenggara', 'dan', 'masuk', 'besar', 'negara', 'yang', 'mengalami', 'krisis', 'pangan', 'worried', 'face'

f) *Remove Stopwords*

Stopwords diproses pada sebuah kalimat jika mengandung kata- kata yang sering keluar dan dianggap tidak penting seperti waktu, penghubung dan lain sebagainya sehingga perlu dilakukan penghapusan. Pada proses ini *library remove stopwords* yang digunakan adalah *library nltk*. Proses *remove stopwords* dapat dilihat pada Tabel 10.

Tabel 10. Remove Stopwords

Remove Stopwords	
'para', 'pemimpin', 'dan', 'kabinet', 'banyak', 'yang', 'mundur', 'serta', 'hura', 'hura', 'di', 'setiap', 'negara', 'dan', 'kami', 'sudah', 'katakan', 'tidak', 'ada', 'yang', 'bisa', 'keluar', 'dari', 'tekanan', 'tatanan', 'dunia', 'baru', 'atas', 'kehancuran', 'ekonomi', 'apalagi', 'inflasi', 'termasuk', 'lima', 'besar', 'se', 'asia', 'tenggara', 'dan', 'masuk', 'besar', 'negara', 'yang', 'mengalami', 'krisis', 'pangan', 'worried', 'face'	'pemimpin', 'kabinet', 'mundur', 'negara', 'tekanan', 'tatanan', 'dunia', 'kehancuran', 'ekonomi', 'inflasi', 'asia', 'tenggara', 'masuk', 'mengalami', 'krisis', 'pangan', 'worried', 'face'

g) *Stemming*

Stemming merupakan proses untuk mencari stem (kata dasar) dari kata hasil *stopwords removal* (filtering). Hal ini bertujuan untuk menghilangkan imbuhan yang terdapat didalam kata tersebut. Proses *stemming* dapat dilihat pada Tabel 11.

Tabel 11. Stemming

Stemming	
'pemimpin', 'kabinet', 'mundur', 'negara', 'tekanan', 'tatanan', 'dunia', 'kehancuran', 'ekonomi', 'inflasi',	'pimpin', 'kabinet', 'mundur', 'negara', 'tekan', 'tatanan', 'dunia', 'hancur', 'ekonomi', 'inflasi', 'asia', 'tenggara',

'asia', 'tenggara', 'masuk', 'mengalami', 'krisis', 'pangan', 'worried', 'face'	'masuk', 'alami', 'krisis', 'pangan', 'worried', 'face'
---	---

h) *Word Translation from Bahasa to English*

Selain menggunakan *library Regex* atau *Regular Expression* untuk mengekstrak emoji, kita akan gunakan pula *library googletrans* dan *mtranslate* untuk menerjemahkan semua data dari Bahasa Indonesia ke Bahasa Inggris. hal ini dilakukan untuk menyesuaikan dokumen teks dengan hasil ekstraksi emoji menjadi deskripsi nama labelnya yang masih berbahasa Inggris. Proses *word translation from bahasa to English* dapat dilihat pada Tabel 12.

Tabel 12. Word Translation from Bahasa to English

Word Translation from Bahasa to English	
'pimpin', 'kabinet', 'mundur', 'negara', 'tekan', 'tatanan', 'dunia', 'hancur', 'ekonomi', 'inflasi', 'asia', 'tenggara', 'masuk', 'alami', 'krisis', 'pangan', 'worried', 'face'	'leader', 'cabinet', 'resign', 'country', 'press', 'order', 'world', 'destroy', 'economy', 'inflation', 'asia', 'southeast', 'entry', 'natural', 'crisis', 'food', 'worried', 'face'

Sampel dari hasil *Preprocessing* data yang didapatkan dapat dilihat pada Tabel 13, sebagai berikut:

Tabel 13. Sampel Data Sebelum *Preprocessing* dan Hasil *Preprocessing*

Data	Sebelum Preprocessing	Hasil Preprocessing
1	arelx_2022-12-13 15:11:53+00:00, Para pemimpin dan kabinet banyak yang mundur serta hura hura di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan 😞	'leader', 'cabinet', 'resign', 'country', 'press', 'order', 'world', 'destroy', 'economy', 'inflation', 'asia', 'southeast', 'entry', 'natural', 'crisis', 'food', 'worried', 'face'

Dari data diatas akan dilakukan pelabelan data. Emoji yang sebelumnya merupakan karakter non-alfanumerik telah diekstraksi menjadi text agar bisa lanjut dilakukan pelabelan dan pembobotan kata. Pelabelan data dapat dilakukan dengan perhitungan berdasarkan rumus perhitungan skor sentimen, yaitu Skor = Jumlah kata positif – Jumlah kata negatif (Nurrahman, 2021). Salah satu contoh kata positif dan negatif dapat dilihat pada Tabel 14.

Tabel 14. Tabel Perhitungan Skor Sentiment

Komentar	Positif	Negatif	Skor	Pelabelan
['leader', 'cabinet', 'resign', 'riot', 'harass', 'state', 'press', 'order', 'world', 'destroy', 'economy', 'inflation', 'asia', 'southeast', 'entry', 'country', 'natural', 'crisis', 'food', 'worried', 'face']	Country Economy Natural	Resign Destroy Inflation Crisis Worried	3 – 5 : -2	
Jumlah	3	5	-2 < 0	Negative

Sehingga dengan demikian diperoleh perhitungan sebagai berikut : Skor = (Jumlah kata positif) – (Jumlah kata negatif) Skor = 5 – 6 = -1 Skor Akhir yang diperoleh dari simulasi perhitungan bernilai < 0. Sehingga hasil dari klasifikasi ulasan tersebut adalah Negatif. Berikut adalah 5 dokumen teks yang diambil dari data tweet Kecemasan covid-19 dan sosial ekonomi masyarakat:

- Dokumen 1 : country, economy, natural, resign, destroy, inflation, crisis, worried
- Dokumen 2 : society, happy, life, economy, money, stigma, trample, sad, struggle, frowning
- Dokumen 3 : conscious, society, attitude, follow, stigma, stereotype, dictation, fearful
- Dokumen 4 : mental, health, culture, strong, issue, stigma, bias, weak, crying
- Dokumen 5 : phenomenon, strong, society, rich, smiling, stigma, poor, slump, lost, suffer, bipolar

Lalu dihitung skor hasil positif negatif dengan rumus seperti berikut, kemudian untuk perhitungan skor komentar 2 – 5 dapat dilihat di lampiran 7.

4.1.3 Selecting Data dan Transformasi Data (*Selecting and Transformation*)

Seleksi data yang ada pada database sering tidak semuanya akan dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Transformasi data adalah data yang ditransformasikan atau digabungkan ke dalam format dimana kemudian akan diproses dalam data mining. Pada proses ini dilakukan transformasi data dengan *word embedding* untuk merubah huruf pada kalimat menjadi vektor, dikarenakan LSTM tidak mengenali huruf, kemudian *output* yang didapatkan adalah nilai yang nantinya dapat digunakan untuk proses klasifikasi.

4.1.3.1 Word Embedding

Word embedding merupakan Setelah didapatkan hasil *Preprocessing*, tahap selanjutnya adalah pembobotan kata yaitu transformasikan kata menjadi numerik atau vektor, data yang diambil sebagai sampel adalah data ke-2. Sampel data di transformasikan menjadi vektor yang diolah dengan *word embedding* yaitu menggunakan cara *One Hot Encoding*. Pada *One Hot Encoding*, angka yang dipresentasikan akan menjadi nilai vektor yang semua elemen bernilai 0 kecuali hanya satu elemen yang bernilai 1 yaitu elemen yang sama dengan kata yang terdapat pada kamus dengan *one hot encoding*. Karena BiLSTM tidak mengenali huruf, maka pembobotan kata menjadi vektor perlu dilakukan agar data tersebut dapat diproses dengan metode BiLSTM. Berikut adalah contoh dari penerapan BiLSTM, dan sampel data di transformasikan menjadi vektor yang diolah dengan *word embedding* yaitu menggunakan cara *One Hot Encoding* dapat dilihat pada Tabel 15.

Tabel 15. Sampel data *One Hot Encoding*

Data	<i>One Hot Encoding</i>							
	state	economy	natural	resign	destroy	inflation	crisis	worried
state	1	0	0	0	0	0	0	0
economy	0	1	0	0	0	0	0	0
natural	0	0	1	0	0	0	0	0
resign	0	0	0	1	0	0	0	0
destroy	0	0	0	0	1	0	0	0

inflation	0	0	0	0	0	1	0	0
crisis	0	0	0	0	0	0	1	0
worried	0	0	0	0	0	0	0	1

Setelah dilakukan *one hot encoding*, selanjutnya melakukan inisialisasi bobot awal secara *random*, ketika melakukan perhitungan diawal akan mendapatkan nilai *error* yang besar dan akan membuat hasil akhir jauh dari prediksi sebenarnya. Nilai *weight* akan diinisialisasikan secara *random* dengan jangkauan dari:

$$W = \left(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}} \right) \quad W = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \quad W = (-0,707,0,707)$$

Keterangan:

- w = range untuk nilai bobot
- d = nilai angka acak untuk menentukan nilai jangkauan bobot

Dari perhitungan di atas, nilai bobot akan diinisialisasikan dengan jangkauan dari (-0.707,0.707). Setelah melakukan proses transformasi menggunakan *one hot encoding* pada Tabel 3, data tersebut kemudian akan dimasukkan ke dalam metode LSTM. Pada data di Tabel 3, kita akan jadikan sebagai data latih untuk melakukan proses pembelajaran

$$\begin{matrix} X_1 = (1\ 0\ 0\ 0\ 0\ 0\ 0\ 0) \\ X_2 = (0\ 1\ 0\ 0\ 0\ 0\ 0\ 0) \\ X_3 = (0\ 0\ 1\ 0\ 0\ 0\ 0\ 0) \\ X_4 = (0\ 0\ 0\ 1\ 0\ 0\ 0\ 0) \\ X_5 = (0\ 0\ 0\ 0\ 1\ 0\ 0\ 0) \\ X_6 = (0\ 0\ 0\ 0\ 0\ 1\ 0\ 0) \\ X_7 = (0\ 0\ 0\ 0\ 0\ 0\ 1\ 0) \\ X_8 = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 1) \end{matrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Kemudian tentukan nilai *weight* dan *bias* secara acak dari hasil perhitungan jangkauan yang sudah ditentukan sebelumnya.

$$W_f = \begin{pmatrix} 0,6 \\ 0,2 \\ 0,3 \\ 0,5 \\ 0,4 \\ 0,1 \\ 0,2 \\ 0,1 \end{pmatrix} \quad W_c = \begin{pmatrix} 0,4 \\ 0,3 \\ 0,3 \\ 0,1 \\ 0,2 \\ 0,5 \\ 0,6 \\ 0,1 \end{pmatrix} \quad W_i = \begin{pmatrix} 0,1 \\ 0,2 \\ 0,2 \\ 0,7 \\ 0,3 \\ 0,4 \\ 0,5 \\ 0,6 \end{pmatrix} \quad W_o = \begin{pmatrix} 0,2 \\ 0,2 \\ 0,1 \\ 0,5 \\ 0,7 \\ 0,6 \\ 0,3 \\ 0,1 \end{pmatrix} \quad B_i = 0,2 \quad B_o = 0,1 \quad B_f = 0,3 \quad B_c = 0,5$$

4.1.4 Data Mining

Proses data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

4.1.4.1 Klasifikasi Bidirectional Long Short Term Memory (BiLSTM)

Berikut ini merupakan contoh dalam penerapan metode BiLSTM untuk pemrosesan kata yang didapat dari *tweets* hasil *Preprocessing* sebelumnya.

Langkah pertama yang dilakukan adalah perhitungan X_t . Hal ini merupakan langkah pertama yang dilakukan di LSTM untuk memutuskan informasi apa yang dihapus dari *cell state*. *forget gate* akan memproses h_{t-1} dan X_t sebagai *input* dan menghasilkan *output* berupa angka 0 sampai 1.

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)$$

$$f_t = \sigma \left(\begin{pmatrix} 0,6 \\ 0,2 \\ 0,3 \\ 0,5 \\ 0,4 \\ 0,1 \\ 0,2 \\ 0,1 \end{pmatrix} \cdot \begin{bmatrix} 0, \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{bmatrix} + 0,3 \right) = \sigma \begin{pmatrix} 0,9 \\ 0,3 \\ 0,3 \\ 0,3 \\ 0,3 \\ 0,3 \\ 0,3 \\ 0,3 \end{pmatrix} \approx f_t = \begin{pmatrix} 0,7109 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \end{pmatrix}$$

Langkah selanjutnya adalah memutuskan informasi yang nantinya akan disimpan pada *cell state*, langkah ini terbagi menjadi 2 bagian. Bagian pertama bernama *input gate* dimana bagian ini memutuskan nilai yang akan diperbarui. Selanjutnya \tilde{c}_t akan ditambah ke *cell state*.

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$

$$i_t = \sigma \left(\begin{pmatrix} 0,1 \\ 0,2 \\ 0,2 \\ 0,7 \\ 0,3 \\ 0,4 \\ 0,5 \\ 0,6 \end{pmatrix} \cdot \begin{bmatrix} 0, \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{bmatrix} + 0,2 \right) = \sigma \begin{pmatrix} 0,3 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \end{pmatrix} \approx i_t = \begin{pmatrix} 0,5744 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \end{pmatrix}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c)$$

$$\tilde{c}_t = \tanh \left(\begin{pmatrix} 0,4 \\ 0,3 \\ 0,3 \\ 0,1 \\ 0,2 \\ 0,5 \\ 0,6 \\ 0,1 \end{pmatrix} \cdot \begin{bmatrix} 0, \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{bmatrix} + 0,5 \right) = \tanh \begin{pmatrix} 0,9 \\ 0,5 \\ 0,5 \\ 0,5 \\ 0,5 \\ 0,5 \\ 0,5 \\ 0,5 \end{pmatrix} \approx \tilde{c}_t = \begin{pmatrix} 0,7163 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \end{pmatrix}$$

Langkah ketiga adalah memperbarui *cell state* yang lama (c_{t-1}) menjadi nilai yang baru dengan mengkalikan (c_{t-1}) dengan f_t , untuk menghapus informasi sebelumnya, dan menambahkan dengan *input gate* (i_t) yang dikalikan dengan \tilde{c}_t yang merupakan nilai baru dan digunakan untuk memperbarui *state*.

$$C_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t$$

$$C_t = \begin{pmatrix} 0,7109 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \end{pmatrix} \cdot 0 + \begin{pmatrix} 0,5744 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \end{pmatrix} \cdot \begin{pmatrix} 0,7163 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \end{pmatrix} = \begin{pmatrix} 0,4114 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \end{pmatrix}$$

Langkah selanjutnya adalah untuk memutuskan hasil *output*, apakah hasil tersebut sesuai atau tidaknya dengan apa yang sudah ditentukan sebelumnya.

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

$$O_t = \sigma \left(\begin{pmatrix} 0,2 \\ 0,2 \\ 0,1 \\ 0,5 \\ 0,7 \\ 0,6 \\ 0,3 \\ 0,1 \end{pmatrix} \cdot 0, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 0,1 \right) = \sigma \begin{pmatrix} 0,3 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \end{pmatrix} \approx o_t = \begin{pmatrix} 0,5744 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \end{pmatrix}$$

$$h_t = O_t \cdot \tanh(C_t)$$

$$h_t = \begin{pmatrix} 0,5744 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \end{pmatrix} \cdot \tanh \begin{pmatrix} 0,4114 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \end{pmatrix} = \begin{pmatrix} 0,2238 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \end{pmatrix}$$

Kemudian perhitungan dilanjutkan sampai X_8 dan diperoleh hasil sebagai berikut:

	f_t	i_t	\tilde{c}_t	c_t	o_t	h_t
X_1	$\begin{pmatrix} 0,7109 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \\ 0,5744 \end{pmatrix}$	$\begin{pmatrix} 0,5744 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \\ 0,5498 \end{pmatrix}$	$\begin{pmatrix} 0,7163 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \\ 0,4621 \end{pmatrix}$	$\begin{pmatrix} 0,4114 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \\ 0,2540 \end{pmatrix}$	$\begin{pmatrix} 0,5744 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \\ 0,2549 \end{pmatrix}$	$\begin{pmatrix} 0,2238 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \\ 0,1305 \end{pmatrix}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

X_8	$\begin{pmatrix} 1,2773 \\ 1,8176 \\ 1,5874 \\ 1,3613 \\ 1,4346 \\ 1,7387 \\ 1,7706 \\ 1,6096 \end{pmatrix}$	$\begin{pmatrix} 1,6950 \\ 1,9036 \\ 1,7014 \\ 1,2996 \\ 1,5488 \\ 1,8094 \\ 1,9035 \\ 1,2542 \end{pmatrix}$	$\begin{pmatrix} 0,8194 \\ 0,3381 \\ 0,6242 \\ 0,5675 \\ 0,6449 \\ 0,4732 \\ 0,3641 \\ 0,6011 \end{pmatrix}$	$\begin{pmatrix} 24,0862 \\ 52,3979 \\ 47,5005 \\ 18,2493 \\ 26,7904 \\ 62,9450 \\ 55,4838 \\ 48,0575 \end{pmatrix}$	$\begin{pmatrix} 1,6521 \\ 1,9986 \\ 1,8375 \\ 1,4413 \\ 1,3558 \\ 1,8894 \\ 1,9600 \\ 1,8229 \end{pmatrix}$	$\begin{pmatrix} 1,6121 \\ 1,0586 \\ 0,5310 \\ 1,4006 \\ 1,3286 \\ -0,6788 \\ -1,6544 \\ 1,2639 \end{pmatrix}$
-------	--	--	--	--	--	--

Secara teknis, BiLSTM menerapkan dua LSTM terpisah, satu untuk arah depan (*forward layer*) dan satu untuk arah mundur (*backward layer*). Berikut merupakan perhitungan LSTM untuk arah mundur (*backward layer*) masih dengan cara yang sama seperti perhitungan untuk arah maju (*forward layer*). perhitungan dilakukan dari X_8 sampai dengan X_1 , kemudian *output* keduanya akan digabungkan sebelum dilakukannya evaluasi model.

	f_t	i_t	\tilde{c}_t	c_t	o_t	h_t
X_8	$\begin{pmatrix} 1,2816 \\ 1,5994 \\ 1,6317 \\ 1,3677 \\ 1,4354 \\ 1,7928 \\ 2,0312 \\ 1,5753 \end{pmatrix}$	$\begin{pmatrix} 1,6968 \\ 1,6625 \\ 1,7362 \\ 1,3071 \\ 1,5496 \\ 2,0741 \\ 2,8723 \\ 1,2104 \end{pmatrix}$	$\begin{pmatrix} 0,8160 \\ 0,6737 \\ 0,5778 \\ 0,5649 \\ 0,6444 \\ 0,1592 \\ -0,4562 \\ 0,6208 \end{pmatrix}$	$\begin{pmatrix} 32,2533 \\ 84,9252 \\ 78,5096 \\ 25,6978 \\ 39,4534 \\ 113,1778 \\ 111,3883 \\ 76,4563 \end{pmatrix}$	$\begin{pmatrix} 1,6554 \\ 1,7322 \\ 1,8580 \\ 1,4491 \\ 1,3570 \\ 2,3596 \\ 2,4863 \\ 1,7215 \end{pmatrix}$	$\begin{pmatrix} -1,3448 \\ 1,2243 \\ -1,3570 \\ 1,4350 \\ -0,9318 \\ -1,5670 \\ -0,6946 \\ -1,0186 \end{pmatrix}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_1	$\begin{pmatrix} 1,1609 \\ 1,5609 \\ 1,6407 \\ 2,0462 \\ 1,7001 \\ 1,6657 \\ 1,5849 \\ 1,6246 \end{pmatrix}$	$\begin{pmatrix} 1,6348 \\ 1,6199 \\ 1,7432 \\ 2,3275 \\ 1,7848 \\ 1,5338 \\ 1,4535 \\ 1,2941 \end{pmatrix}$	$\begin{pmatrix} 0,8390 \\ 0,7245 \\ 0,5683 \\ 0,4060 \\ 0,4840 \\ 0,7758 \\ 0,8363 \\ 0,5853 \end{pmatrix}$	$\begin{pmatrix} 201,7940 \\ 32,2280 \\ 35,3809 \\ 138,1735 \\ 97,7687 \\ 54,5779 \\ 35,5645 \\ 37,6891 \end{pmatrix}$	$\begin{pmatrix} 1,5439 \\ 1,6851 \\ 1,8621 \\ 2,2778 \\ 1,8197 \\ 1,4763 \\ 1,6348 \\ 1,7629 \end{pmatrix}$	$\begin{pmatrix} 0,4747 \\ 1,3740 \\ -0,2024 \\ -1,8495 \\ 0,1654 \\ 0,6435 \\ -0,2041 \\ -0,5191 \end{pmatrix}$

Setelah perhitungan selesai, kedua nilai pada kolom h_t *forward layer* dan *backward layer* dihitung menggunakan persamaan *output* BiLSTM. Lalu kemudian nilai terbaru dihitung menggunakan fungsi aktivasi *softmax*.

$$y_t = W_{hy}h_{\bar{t}} + W_{hy}h_{\bar{t}} + b_y$$

$$y_t = \begin{pmatrix} 1,7520 \\ 1,7163 \\ 0,6314 \\ 0,3653 \\ 0,5165 \\ 0,4929 \\ 1,9868 \\ -0,1703 \end{pmatrix} \Rightarrow y_1 = \text{softmax} \begin{pmatrix} 1,7520 \\ 1,7163 \\ 0,6314 \\ 0,3653 \\ 0,5165 \\ 0,4929 \\ 1,9868 \\ -0,1703 \end{pmatrix}$$

Hasil dari proses perhitungan akan menghasilkan berupa *vektor output*. Untuk mengetahui hasil dari prediksi algoritma BiLSTM, dilihat dari nilai probabilitas terbesar pada *vektor output*, maka nilai “0,2792” diambil sebagai prediksi yang dapat dilihat di Tabel 16.

Tabel 16. Hasil Prediksi

Output	Label
0,2208	Negatif
0,2130	Positif
0,0720	Negatif
0,0551	Positif
0,0641	Negatif
0,0626	Positif
0,2792	Negatif
0,0323	Negatif

Dari proses pengujian data diatas, dapat disimpulkan bahwa data tersebut masuk kedalam kelas negative.

4.1.5 Evaluasi Pola (*Pattern Evaluation*)

Tahap ini untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Setelah menerapkan metode BiLSTM, kemudian perlu untuk menghitung akurasi dari data yang didapat untuk melihat apakah sudah sesuai atau belum. Pada penelitian ini dilakukan penilaian dari keberhasilan metode berdasarkan keakuratan memprediksi menggunakan *confusion matrix*. Parameter yang akan digunakan dalam penelitian ini yaitu parameter akurasi. Dari hasil perhitungan metode diatas akan dicocokkan dengan 4 data aktual menggunakan *confusion matrix* seperti pada Tabel 17.

Tabel 17. Hasil *Confusion Matrix*

Prediksi	Aktual		
	Negatif	Positif	Neutral
Negatif	2	1	0
Positif	0	1	0
Netral	0	0	1

Dari Tabel 17 diatas didapatkan hasil :

- True Positive* = 1 dokumen pada kelas positif, 2 dokumen pada kelas negatif, dan 1 dokumen pada kelas neutral yang berarti jumlah dokumen positif diprediksi positif.
- True Negative* = 0 dokumen, yang berarti jumlah dokumen negatif diprediksi negatif
- False Positive* = 0 dokumen, yang berarti jumlah dokumen negatif diprediksi positif
- False Negative*= 1 dokumen, yang berarti jumlah dokumen positif diprediksi negatif.

Dengan demikian hasil *accuracy* sebagai berikut :

$$accuracy = \frac{TN + TP}{TN + TP + FN + FP} = \frac{0 + (2 + 1 + 1)}{5} = 0,80 \text{ atau } 80\%$$

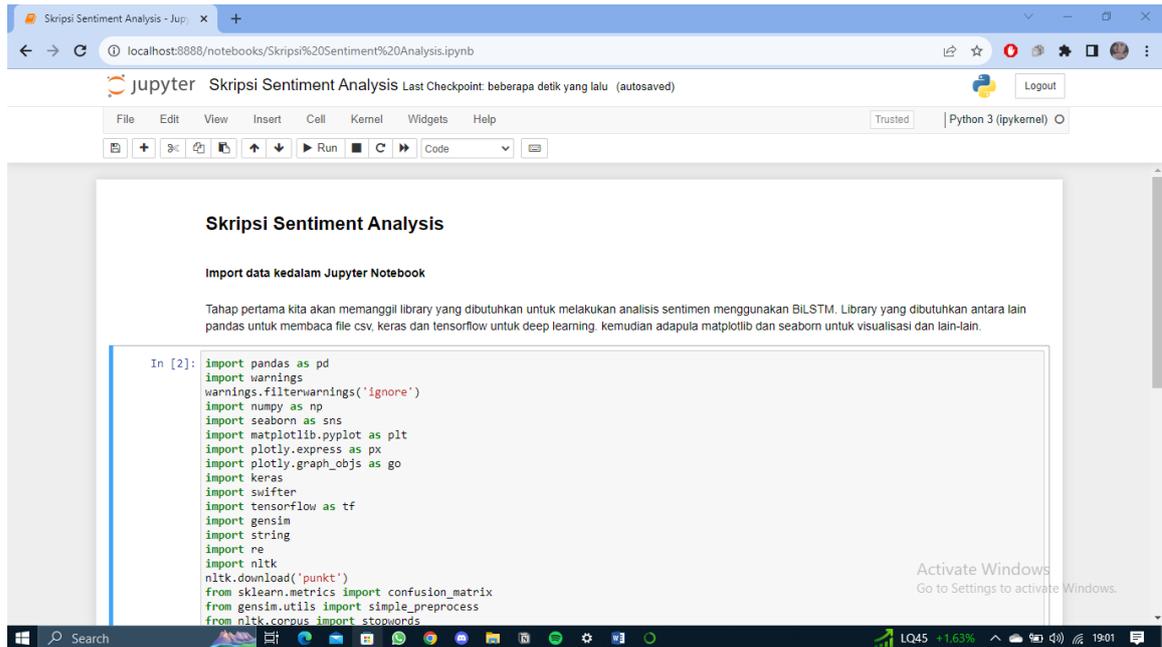
Dari hasil perhitungan pada Tabel 6, didapatkan hasil akurasi sebesar 80%.

4.1.6 Presentasi Pengetahuan (*Knowledge Evaluation*)

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah di mengerti oleh pihak yang berkepentingan.

4.2 Implementasi Sistem

Tahap implementasi merupakan tahap pembangunan sistem yang sudah dirancang, dalam tahapan ini pembangunan sistem dilakukan dengan menggunakan *software* Jupyter Notebook (Anaconda3) dengan bahasa pemrograman Python. Implementasi menggunakan Jupyter Notebook (Anaconda3) dapat dilihat pada Gambar 7.



Gambar 7. Implementasi menggunakan Jupyter Notebook (Anaconda3)

BAB V HASIL DAN PEMBAHASAN

5.1 Hasil

Pada bab sebelumnya telah diuraikan perancangan program yang di bangun. Berikut ini adalah hasil dan pembahasan program yang sudah dibuat beserta pembahasan program.

5.1.1 Tahap Pengumpulan Data (*Data Selection*)

Pengambilan data dari sosial media twitter yang terdiri dari topik aspek kesehatan, sosial, dan ekonomi. Pengambilan data dilakukan secara manual dengan cara menginputkan kata kunci atau *keyword* twitter yang sudah ditetapkan pada masing–masing aspek sehingga data yang di dapatkan tidak real time. Data yang digunakan berasal dari *tweet* pengguna media sosial twitter pada bulan Desember 2022. Data aspek yang dibahas divalidasi berdasarkan artikel survey kesehatan mental oleh Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI, 2022) dimana artikel tersebut menunjukkan bahwa dalam kurun waktu 2020 sejak pandemic covid-19 dimulai hingga tahun 2022 masih menunjukkan tingginya masalah kecemasan dan disetujui pakar Bahasa Indonesia.

Data hasil *scrapping* berdasarkan kata kunci yang telah ditentukan kemudian disimpan dan digabungkan menjadi satu dalam bentuk file excel dan csv sesuai dengan aspek yang bersangkutan. Jumlah data yang telah didapatkan berdasarkan setiap aspeknya dapat dilihat pada Tabel 3, dan untuk proses pengumpulan kata kunci menjadi satu file aspek dapat dilihat di lampiran 3.

5.1.2 Proses Pembersihan data dan Integrasi Data

Pembersihan data (*cleaning data*) adalah proses menghilangkan noise pada data yang tidak relevan dan konsisten. Namun, karena penelitian ini juga mempertimbangkan nilai emoji yang dapat mempengaruhi isi dari data tweet, maka sebelum dilakukannya proses pembersihan data dimana hal tersebut dapat menghilangkan emoji yang merupakan karakter non-alfanumerik, emoji yang tersemat di setiap dokumen tweet dalam dataset dikonversikan menjadi nama label atau deskripsi nama spesifiknya berdasarkan penamaan di kamus Unicode. Kemudian tahapan proses pembersihan data dimulai dari *preprocessing* dimana didalam proses *preprocessing* itu sendiri terdapat beberapa proses di dalamnya yaitu *cleaning data*, *case folding*, *tokenizing*, *stopwords*, *stemming*, dan *translation*.

Proses *case folding* dilakukan otomatis melalui syntax *case_folding*, kemudian *stopwords* diproses melalui *library* NLTK. *Library* NLTK dan Sastrawi digunakan untuk melakukan analisis terhadap dokumen yang dimiliki. Dalam prosesnya kita akan mengenal aspek dalam NLP (*Natural Language Processing*) yaitu *tokenization* dan *stopwords*.

Proses *translate* data kedalam bahasa inggris memanfaatkan *library googletrans* dan *mtranslate* agar penyematan emoji pada data dapat memberikan pengaruh nilai dalam hasil akhir dengan signifikan. *Stemmer* yang dipakai menggunakan Sastrawi untuk *stemming* bahasa Indonesia. Untuk hasil proses *cleaning* data dapat dilihat pada lampiran 8.

5.1.3 Proses Pembobotan Kata

Data tweet berbentuk kalimat yang perlu diubah terdahulu ke dalam format yang sesuai dalam bentuk numerik atau vector. Karna BiLSTM tidak mengenali huruf dalam proses nya, maka tujuan pembobotan kata menjadi vector perlu dilakukan agar data

tersebut bisa diproses. Angka yang dipresentasikan akan menjadi nilai vector yang semua elemen bernilai 0 kecuali hanya satu elemen yang bernilai 1.

5.1.4 Proses Model BiLSTM

Pada proses ini, hasil embedding yang telah dilakukan pada data di tiap aspek kemudian dilanjutkan kedalam pemrosesan model BiLSTM. Untuk kodifikasi program BiLSTM dapat dilihat pada Gambar 8 sebagai berikut:

```
# Model BiLSTM dengan word embedding untuk 3 kelas
model = Sequential()
model.add(Embedding(max_words, 64, input_length=max_len))
model.add(Bidirectional(LSTM(64, return_sequences=True)))
model.add(Dropout(0.5))
model.add(Bidirectional(LSTM(128)))
model.add(Dense(256, activation='relu')) # Menambah Layer Dense dengan Lebih banyak neuron
model.add(Dropout(0.5)) # Menambah Layer dropout
model.add(Dense(3, activation='softmax'))
model.summary()

import tensorflow as tf
from tensorflow.keras.callbacks import EarlyStopping

# Kompilasi model
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

# Convert labels to TensorFlow tensors
train_labels_tf = tf.convert_to_tensor(train_labels, dtype=tf.int32)
test_labels_tf = tf.convert_to_tensor(test_labels, dtype=tf.int32)

# Callback EarlyStopping
early_stopping = EarlyStopping(monitor='val_accuracy', patience=5, restore_best_weights=True)

# Melatih model
history = model.fit(x_train, train_labels_one_hot, epochs=25, batch_size=32, validation_data=(x_test, test_labels_one_hot),
                    verbose=1, shuffle=True)
```

Gambar 8. Proses Model *Bidirectional Long Short-Term Memory*

Pada kodingan LSTM ini, didefinisikan ada 64 jumlah unit (neuron) dalam hidden layers, termasuk lapisan embedding dan BiLSTM. Jumlah unit dalam hidden layers menentukan kapasitas model mempelajari representasi yang kompleks dari data pelatihan. Jumlah unit yang besar dapat membantu model untuk mengatasi pola yang lebih rumit dalam data, tetapi juga dapat meningkatkan resiko overfitting. Namun, penerapan dropout dengan nilai 0,5 pada lapisan BiLSTM adalah untuk mencegah overfitting. Kemudian optimizer adam merupakan algoritma pengoptimalan jaringan saraf yang baik untuk masalah deep learning. Dengan menggunakan hal-hal tersebut jaringan BiLSTM ini tidak bergantung dengan semua syaraf-syaraf nya melainkan memaksa jaringan tersebut untuk mencari *pattern*-nya untuk mencari kata-kata yang memiliki arti dengan tujuan meningkatkan tingkat akurasi permasalahan yang kita cari.

5.1.5 Hasil Implementasi dan Pengujian Model

Pada proses ini dilakukan implementasi dan pengujian model untuk melihat hasil implementasinya dengan menggunakan confusion matrix dengan kodingan program pada Gambar 9 sebagai berikut:

```
from sklearn.metrics import confusion_matrix, classification_report

# print(y_test.argmax(axis=1))
# print(predictions.argmax(axis=1))
# print(y_test)

matrix = confusion_matrix(test_labels_one_hot.argmax(axis=1), predictions.argmax(axis=1))
print(matrix)
```

```

import seaborn as sns

conf_matrix = pd.DataFrame(matrix, index = ['negative', 'neutral', 'positive'], columns = ['negative', 'neutral', 'positive'])
#Normalizing

plt.figure(figsize = (15,15))
sns.heatmap(conf_matrix, annot=True, annot_kws={"size": 25}, cmap='Reds',fmt='g')
plt.xlabel("Predicted")
plt.ylabel("Actual")

```

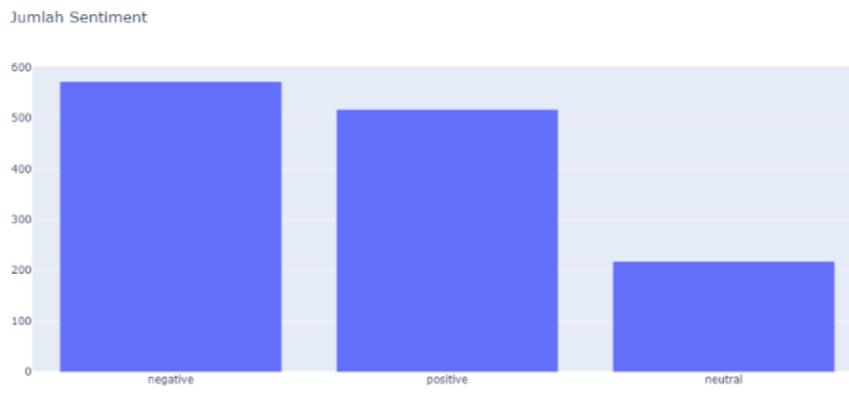
Gambar 9. Hasil Confussion Matrix

5.2 Pembahasan

Berdasarkan program yang telah diimplementasikan terdapat beberapa langkah yang harus dilakukan. Data ulasan twitter dibagi menjadi aspek-aspek terkait isu kecemasan yang ditimbulkan oleh adanya pandemic covid-19. Aspek dibagi menjadi 3 topik atau isu yaitu aspek kesehatan, aspek sosial, dan aspek ekonomi.

5.2.1 Hasil Klasifikasi Sentimen Aspek Kesehatan

Hasil analisis sentimen aspek kesehatan dengan penyematan emoji sebanyak 1.304 data. Berdasarkan hasil analisis sentimen pada variabel yaitu sentimen dengan tingkat kecemasan negatif sebanyak 571, sentimen dengan tingkat kecemasan positif sebanyak 516, dan sentimen dengan tingkat kecemasan netral sebanyak 217 maka diperoleh hasil sebagai berikut, dapat dilihat pada Gambar 10 serta hasil klasifikasi sentimen aspek kesehatan tanpa penyematan emoji dapat dilihat pada lampiran 9.



Gambar 10. Hasil Klasifikasi Sentimen Aspek Kesehatan dengan Emoji

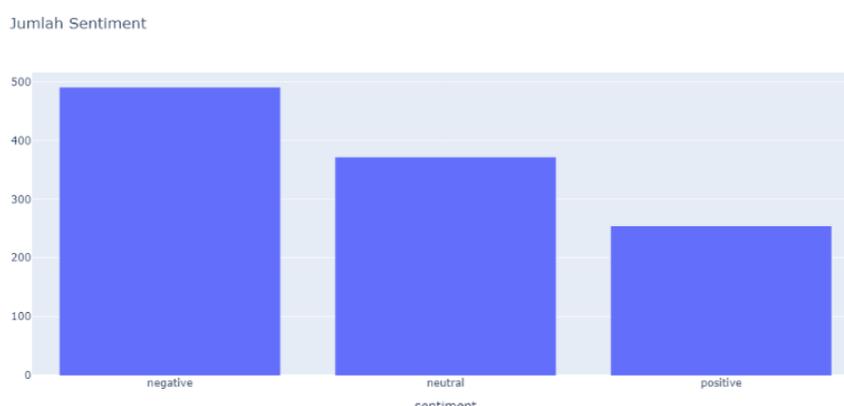
Dilihat dari tingkat kecemasan masyarakat, tingkat kecemasan negatif pada aspek kesehatan memiliki data yang paling banyak dibahas oleh masyarakat tentang kecemasan yang terjadi akibat adanya pandemic covid-19. Berdasarkan Gambar 10 diketahui bahwa rata-rata masyarakat Indonesia dalam kurun waktu bulan desember 2022 mengekspresikan kecamasan nya terhadap pandemic covid-19 dari segi kesehatan diperoleh sentimen dengan tingkat kecemasan negatif sebanyak 571, lebih banyak dibandingkan dengan sentimen dengan tingkat kecemasan positif dan sentimen dengan tingkat kecemasan netral. Untuk sentimen dengan tingkat kecemasan negatif sebanyak 199, sentimen dengan tingkat kecemasan positif sebanyak 158, dan sentimen dengan tingkat kecemasan netral sebanyak 74 dari data uji sebanyak 431 data. Atau sentimen dengan tingkat kecemasan negatif sebesar 46%, sentimen dengan tingkat kecemasan positif sebesar 37% dan sentimen dengan tingkat kecemasan netral sebesar 17%. Dalam hal ini, sentimen masyarakat terhadap pandemi covid-19 yang sudah terjadi dalam kurun waktu 2 tahun berlalu sejak desember 2022 dalam segi aspek kesehatan cenderung dominan kearah tingkat kecemasan yang negatif dibandingkan tingkat kecemasan yang positif dan netral.

Melihat tingkat kecemasan masyarakat yang cenderung dominan kearah tingkat kecemasan negatif didukung oleh argument kementerian kesehatan RI. Menurut juru bicara kementerian kesehatan, dr. Muhammad Syahril mengungkapkan Indonesia masih berada di level 1 transmisi komunitas, kendati demikian vaksinasi tetap harus digalakan. Meskipun terjadi nya penurunan kasus dalam pandemic covid-19, angka kematian akibat covid-19 mengalami fluktuasi. Dari rata-rata mingguan sebanyak 46 orang. Terjadi kenaikan sebanyak 54 orang pada 1 Desember 2022, yang juga diiringi dengan peningkatan *positivity rate*. Maka dari itu, upaya vaksinasi menjadi bagian upaya atau strategi kementerian kesehatan dalam mencapai atau menuju berakhirnya pandemic covid-19 (Rokom, 2022).

Psikiater lembaga kesehatan PBNU, dr. Citra Fitri Agustina menyampaikan bahwa pandemic covid-19 yang telah terjadi sejak 2020 lalu juga memberi dampak yang cukup signifikan dalam hal kesehatan mental masyarakat sejak 12 bulan terakhir dengan mencapai sebanyak 2,45 juta orang didiagnosis mengalami gangguan kejiwaan. Perubahan aktivitas selama pandemic membuat lebih tertekan, lebih cemas, lebih kesepian, atau lebih sulit berkonsentrasi dibandingkan sebelum pandemic covid-19 melanda. Isolasi sosial dan perasaan kesepian meningkatkan risiko depresi dan kecemasan, dan durasi kesepian selama pandemic covid-19 ini memiliki dampak yang paling besar terhadap kesehatan mental masyarakat (UGM, 2022).

5.2.2 Hasil Klasifikasi Sentimen Aspek Sosial

Hasil analisis sentimen aspek sosial sebanyak 1.115 data. Berdasarkan hasil analisis sentimen pada variabel yaitu sentimen dengan tingkat kecemasan negatif sebanyak 490, sentimen dengan tingkat kecemasan netral sebanyak 371, dan sentimen dengan tingkat kecemasan positif sebanyak 254 maka diperoleh hasil sebagai berikut, dapat dilihat pada Gambar 11 serta hasil klasifikasi aspek sosial tanpa penyematan emoji dapat dilihat pada lampiran 9.



Gambar 11. Hasil Klasifikasi Sentimen Aspek Sosial dengan Emoji

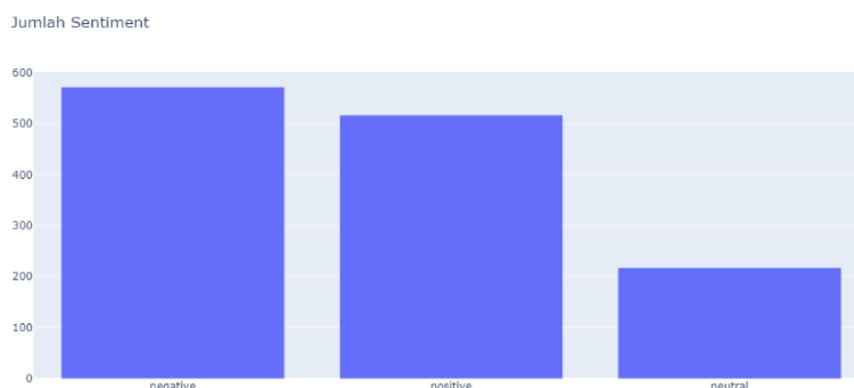
Dilihat dari tingkat kecemasan masyarakat, tingkat kecemasan negatif pada aspek sosial memiliki data yang paling banyak dibahas oleh masyarakat tentang kecemasan yang terjadi akibat adanya pandemic covid-19. Berdasarkan Gambar 13 diketahui bahwa rata-rata masyarakat Indonesia dalam kurun waktu bulan desember 2022 mengekspresikan kecemasan nya terhadap pandemic covid-19 dari segi sosial diperoleh sentimen dengan tingkat kecemasan negatif dan netral lebih banyak dibandingkan dengan sentimen dengan tingkat kecemasan positif. Untuk sentimen dengan tingkat kecemasan negatif sebanyak 128, sentimen dengan tingkat kecemasan netral sebanyak 118, dan sentimen dengan tingkat

kecemasan positif sebanyak 81 dari data uji sebanyak 391 data. Atau sentimen dengan tingkat kecemasan negatif sebesar 33%, sentimen dengan tingkat kecemasan netral sebesar 31% dan sentimen dengan tingkat kecemasan positif sebesar 21%. Dalam hal ini, sentimen masyarakat terhadap pandemi covid-19 yang sudah terjadi dalam kurun waktu 2 tahun berlalu sejak desember 2022 dalam segi aspek sosial cenderung dominan kearah tingkat kecemasan yang negatif dibandingkan tingkat kecemasan yang netral dan positif.

Melihat tingkat kecemasan masyarakat yang cenderung dominan kearah tingkat kecemasan negatif, selama pandemic covid-19 terdapat aturan baru yang telah diatur oleh pemerintah dan wajib dipatuhi oleh masyarakat yang menjadi suatu keterbatasan. Interaksi menjadi suatu pembahasan yang menarik dalam ruang lingkup sosial sejak pandemic covid-19 dimulai, dimana kita mengetahui bahwa kita perlu beradaptasi dengan peraturan-peraturan baru yang menyebabkan terjadinya perubahan baik dari rutinitas maupun kurangnya interaksi sosial (Yatimah, 2022).

5.2.3 Hasil Klasifikasi Sentimen Aspek Ekonomi

Hasil analisis sentimen aspek ekonomi sebanyak 1.288 data. Berdasarkan hasil analisis sentimen pada variabel yaitu sentimen dengan tingkat kecemasan negatif sebanyak 487 data, sentimen dengan tingkat kecemasan positif sebanyak 401 data, dan sentimen dengan tingkat kecemasan netral sebanyak 400 data maka diperoleh hasil sebagai berikut, dapat dilihat pada Gambar 12 serta hasil klasifikasi aspek ekonomi tanpa penyematan emoji dapat dilihat pada lampiran 9.



Gambar 12. Hasil Klasifikasi Sentimen Aspek Ekonomi dengan Emoji

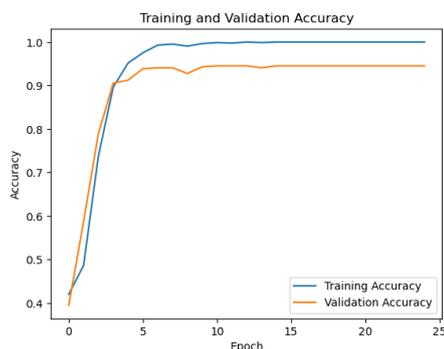
Dilihat dari tingkat kecemasan masyarakat, tingkat kecemasan negatif pada aspek ekonomi memiliki data yang paling banyak dibahas oleh masyarakat tentang kecemasan yang terjadi akibat adanya pandemic covid-19. Berdasarkan Gambar 14 diketahui bahwa rata-rata masyarakat Indonesia dalam kurun waktu bulan desember 2022 mengekspresikan kecemasa nya terhadap pandemic covid-19 dari segi ekonomi diperoleh sentimen dengan tingkat kecemasan negatif sebanyak 487, lebih banyak dibandingkan dengan sentimen dengan tingkat kecemasan positif dan sentimen dengan tingkat kecemasan netral. Untuk sentimen dengan tingkat kecemasan negatif sebanyak 136, sentimen dengan tingkat kecemasan positif sebanyak 114, dan sentimen dengan tingkat kecemasan netral sebanyak 118 dari data uji sebanyak 368 data. Atau sentimen dengan tingkat kecemasan negatif sebesar 37%, sentimen dengan tingkat kecemasan positif sebesar 31% dan sentimen dengan tingkat kecemasan netral sebesar 32%. Dalam hal ini, sentimen masyarakat terhadap pandemi covid-19 yang sudah terjadi dalam kurun waktu 2 tahun berlalu sejak desember

2022 dalam segi aspek ekonomi cenderung dominan kearah tingkat kecemasan yang negatif dibandingkan tingkat kecemasan yang positif dan netral.

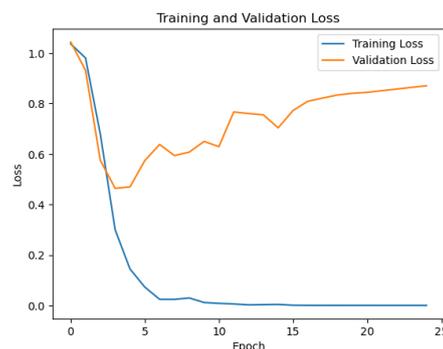
Melihat tingkat kecemasan masyarakat yang cenderung dominan kearah tingkat kecemasan negatif tidak selaras dengan berita yang dilansir oleh biro komunikasi dan pelayanan publik kementerian kesehatan RI, dimana Presiden Joko Widodo resmi mencabut pemberlakuan pembatasan kegiatan masyarakat (PPKM) di seluruh wilayah di Indonesia pada jumat (30/12). Presiden Joko Widodo beranggapan bahwa Indonesia termasuk negara yang berhasil mengendalikan pandemic covid-19 dengan baik dan sekaligus bisa menjaga stabilitas ekonominya karena dalam beberapa bulan terakhir pandemic covid-19 semakin terkendali di Indonesia (Kemenkes RI, 2022). Namun jika menyimak diferensiasi perekonomian global pasca covid-19 yang dilansir oleh infobanknews.com, Rezkiana Nisaputra berbicara bahwa masalah ketidakpastian geopolitik menjadi kondisi yang membawa ancaman pada perekonomian global serta kondisi ekonomi dan keuangan yang lebih ketat dan inflasi yang tinggi akan terus menekan pertumbuhan global pada 2022 dan 2023. Organisasi untuk kerjasama ekonomi dan pembangunan (OECD) setuju dengan perkiraan pesimistis tersebut. Menurut OECD, saat ini negara-negara di dunia sedang menghadapi prospek ekonomi yang sangat sulit. Bukan tentang resesi global, melainkan perlambatan pertumbuhan yang signifikan untuk ekonomi dunia pada 2023, serta inflasi yang masih tinggi, meskipun melandai di banyak negara (infobanknews.com, 2022).

5.3 Hasil Evaluasi Pola (Pattern Evaluation)

Pada penelitian ini model yang dibangun pada tiap aspek menggunakan jaringan Bidirectional Long Short-Term Memory, dimana model ini akan melakukan pemaksimalan sensitivitas dari RNN, yaitu memproses input sekuensial pada arah yang berbeda kemudian dipersatukan. Menggunakan BiLSTM dapat menangkap pattern yang lebih kompleks dibandingkan dengan jaringan LSTM *single layer*, yang juga dapat bekerja dengan lebih baik dengan adanya penyematan emoji dalam algoritmanya. Berikut contoh dari akurasi sebesar 94,52% dan loss dari aspek kesehatan dengan penyematan emoji yang dibuatkan grafik dan dapat dilihat pada Gambar 13 dan 14 serta hasil akurasi dan loss dari aspek lainnya dengan atau tanpa emoji dapat dilihat pada lampiran 10.



Gambar 13. Hasil Training dan Validation Accuracy

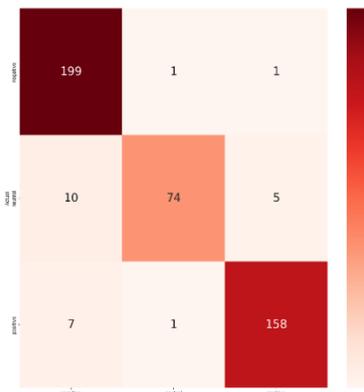


Gambar 14. Hasil Training dan Validation Loss

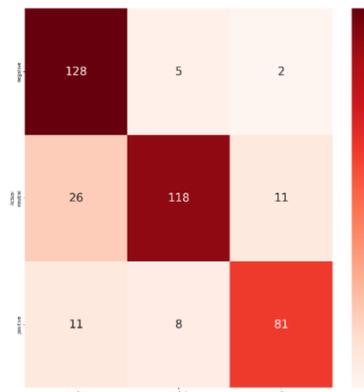
Karena pada model yang dibuat ini menentukan 3 kelas sentimen yaitu positif, negatif, dan netral maka yang menjadi permasalahan untuk dihadapi yaitu merupakan *multiclass, multilabel classification* oleh karena itu kita menggunakan fungsi aktivasi softmax pada layer terakhir dan *categorical crossentropy* untuk *loss function* nya. Pada model yang

dibuat ini juga menggunakan optimizer adam. Optimizer adam sering dipilih sebagai optimizer default yang sering disarankan dalam *deep learning* untuk tugas-tugas klasifikasi *multiclass*.

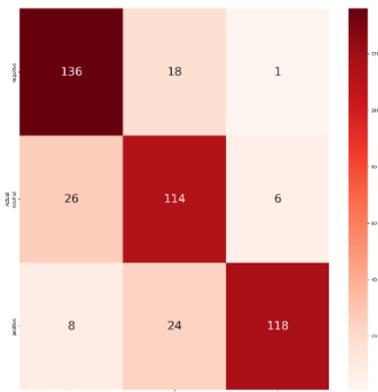
Kemudian dilakukan pengujian *confussion matrix* untuk melihat hasil prediksi dengan hasil uji nya. Berikut hasil validasi klasifikasi BiLSTM keseluruhan yang didapatkan dari hasil *confussion matrix* pada tiap aspek dengan penyematan emoji dan didalam penelitian ini, peneliti juga mengolah data yang sama namun dengan tanpa penyematan emoji. Penelitian ini mengklasifikasikan komentar yang menunjukkan hasil positif, negatif, dan netral. Hasil *confussion matrix* dapat dilihat pada Gambar 15, 16 dan 17 untuk pengujian tiap aspek dengan penyematan emoji. Hasil *confussion matrix* tanpa penyematan emoji dapat dilihat pada lampiran 11.



Gambar 15. Hasil Confussion Matrix Aspek Kesehatan dengan Emoji



Gambar 16. Hasil Confussion Matrix Aspek Sosial dengan Emoji



Gambar 17. Hasil Confussion Matrix Aspek Sosial dengan Emoji

Berdasarkan hasil dari confusion matriks pada Gambar 15, Gambar 16, dan Gambar 17 menunjukkan bahwa sentimen negatif mendominasi seluruh aspek dan dengan hasil persentasi akurasi yang lebih tinggi dari aspek lainnya. Hasil klasifikasi dari tiap-tiap aspek dengan penyematan emoji didapatkan akurasi sebesar 94,52% untuk Aspek Kecemasan Kesehatan, 83,85% untuk Aspek Kecemasan Sosial, dan 81,60% untuk Aspek Kecemasan Ekonomi, dan hasil klasifikasi dari tiap-tiap aspek tanpa penyematan emoji didapatkan akurasi sebesar 76% untuk Aspek Kecemasan Kesehatan, 58% untuk Aspek Kecemasan Sosial, dan 65% untuk Aspek Kecemasan Ekonomi. Dengan demikian dapat kita lihat hasil rata-rata akurasi dari kedua pengujian data secara menyeluruh terdapat pada Tabel 18.

Tabel 18. Hasil Rata-rata Akurasi Pengujian

No	Aspek dengan Emoji	Akurasi	Aspek tanpa Emoji	Akurasi
1	Aspek Kesehatan	94,52%	Aspek Kesehatan	76%
2	Aspek Sosial	84,85%	Aspek Sosial	58%
3	Aspek Ekonomi	81,60%	Aspek Ekonomi	65%
	Rata-rata	86,99%	Rata-rata	66,33%

Didapatkan hasil jumlah ulasan tingkat kecemasan positif, tingkat kecemasan negatif, dan tingkat kesemasan netral pada masing-masing aspek dilihat dari *true positive*, *true negative*, dan *true neutral* pada hasil *confussion matrix* dengan rata-rata nilai pada semua aspek dengan penyematan emoji dengan akurasi 86,99%, dan rata-rata nilai pada semua aspek tanpa penyematan emoji dengan akurasi 66,33%. Dapat disimpulkan bahwa metode

algoritma BiLSTM dengan penyematan emoji dapat mengklasifikasikan sentimen cukup baik, dengan model klasifikasi akurasi tinggi dan error yang cenderung sedikit.

5.4 Wordcloud

Evaluasi pola performa dari hasil positif, negatif, dan netral setiap aspek. Berdasarkan dataset aspek yang ditarik. Wordcloud akan menampilkan kata yang paling sering muncul dengan menyerupai bentuk awan, semakin sering kata tersebut muncul maka akan semakin besar ukurannya. Wordcloud didapatkan dari kata dasar proses stemming. Kata yang telah dicocokkan sesuai corpus positif, negatif, dan netral dihitung setiap kemunculan kata kemudian data teks stemming ditampilkan dalam bentuk visual. Berikut wordcloud berdasarkan aspek dengan penyematan emoji dapat dilihat pada Gambar 18, 19, dan 20. Hasil wordcloud tanpa penyematan emoji dapat dilihat pada lampiran 12.



Gambar 18. Wordcloud Aspek Kesehatan dengan Emoji



Gambar 19. Wordcloud Aspek Sosial dengan Emoji



Gambar 20. Wordcloud Aspek Ekonomi dengan Emoji

Hasil dari wordcloud pada aspek kesehatan dengan emoji menunjukkan bahwa kecemasan menjadi fokus utama dalam data klasifikasi ulasan pengguna Twitter. Terdapat kata-kata seperti 'health', 'vaccine', dan 'impact' yang menunjukkan kesadaran individu terhadap isu kesehatan selama pandemic. Sementara itu, kata-kata seperti 'anxiety', 'anxious', 'lockdown', dan 'crying' mendominasi, menunjukkan intensitas tinggi individu dalam mengekspresikan perasaan mereka mengenai isu kesehatan melalui percakapan online. Diskusi ini juga mencerminkan perhatian perhatian terhadap keadaan Indonesia dalam menghadapi isu kesehatan, dan penggunaan kata 'face' menunjukkan tingginya penggunaan emoji untuk memperkuat pesan.

Wordcloud pada aspek sosial dengan emoji mengungkapkan bahwa masih banyak individu yang membahas interaksi dan stigma dalam isu sosial. Hal ini terlihat dari kata-kata seperti 'stigma', 'interaction', dan 'social'. Perbincangan ini mendorong individu untuk mengekspresikan pikiran dan perasaan mereka melalui emoji yang disematkan dalam pesan, seperti terlihat dari kata-kata 'face', 'loudly', 'frowning', dan 'smiling'.

Analisis wordcloud pada aspek ekonomi dengan emoji menunjukkan bahwa kata-kata yang paling sering muncul dan dibahas adalah 'recession', 'inflation', dan 'economic'. Kata 'pandemic' juga banyak dibicarakan, menunjukkan bahwa resesi dan inflasi banyak dibahas karena dampak signifikan pandemic terhadap perekonomian Indonesia pada tahun 2023. Penggunaan kata 'face' kembali menunjukkan intensitas penggunaan emoji oleh individu untuk memperjelas pesan yang disampaikan.

5.5 Pengujian *K-Fold Cross Validation*

Pada pengujian keseluruhan aspek sebelumnya didapatkan hasil yang cukup baik menggunakan optimizer Adam tanpa mengatur *tuning parameter* nya. Walaupun telah didapatkan hasil yang cukup baik, data setiap aspek dengan atau tanpa penyematan emoji sebelumnya akan diuji coba kembali dengan pengujian *k-fold cross validation* dan pengaturan learning rate secara manual menggunakan optimizer Adam dengan nilai learning rate 0,01 serta pengaturan learning rate secara manual menggunakan optimizer SGD (Stochastic Gradient Descent) dengan learning rate 0,01 dan momentum 0,9. Berdasarkan hasil akurasi menggunakan BiLSTM dengan pengujian *k-fold cross validation* dapat dilihat pada Tabel 19.

Tabel 19. Hasil Akurasi Data Dengan Aspek Emoji Pengujian *K-Fold Cross Validation*

No.	Aspek dengan emoji	Akurasi dengan Adam	Akurasi dengan K-Fold Cross Validation + Adam: Learning Rate 0,01	Akurasi dengan K-Fold Cross Validation + SGD: Learning Rate 0,01
1.	Kesehatan	94,52%	99,29%	63,12%
2.	Sosial	84,85%	97,10%	45,64%
3.	Ekonomi	81,60%	97,85%	65,59%
Rata-rata		86,99%	98,08%	58,11%

Sementara dataset tanpa penyematan emoji didapatkan hasil sebagai berikut dengan pengujian *k-fold cross validation*, dapat dilihat pada Tabel 20.

Tabel 20. Hasil Akurasi Data Dengan Aspek Tanpa Emoji Pengujian *K-Fold Cross Validation*

No.	Aspek tanpa emoji	Akurasi dengan Adam	Akurasi dengan K-Fold Cross Validation + Adam: Learning Rate 0,01	Akurasi dengan K-Fold Cross Validation + SGD: Learning Rate 0,01
1.	Kesehatan	76%	95,39%	45,23%
2.	Sosial	58%	96,30%	46%
3.	Ekonomi	65%	96,40%	46%
Rata-rata		66,33%	96,03%	45,74%

Didapatkan hasil perbandingan hasil akurasi klasifikasi dengan BiLSTM sebelum dan sesudah dilakukan pengujian dengan *k-fold cross validation* serta mengatur tuning parameter nya. Dari Tabel 19 kita dapat mengetahui bahwa mengatur tuning parameter dari optimizer Adam dapat meningkatkan performa atau tingkat akurasi model menjadi lebih baik, dimana nilai rata-rata pada akurasi awal dari seluruh aspek dengan penyematan emoji hanya mencapai 86,99% dan seluruh aspek tanpa penyematan emoji mencapai 66,33%, setelah dilakukan pengujian ini untuk aspek dengan penyematan emoji mencapai akurasi 98,08% dan seluruh aspek tanpa penyematan emoji mencapai 96,03%. Sebaliknya dengan mengatur tuning parameter dari optimizer SDG pada seluruh aspek dengan penyematan emoji atau tanpa penyematan emoji mendapatkan hasil yang tidak lebih baik dari hasil akurasi awal tanpa pengujian *k-fold cross validation* dan penentuan *learning rate*. Hal ini dipengaruhi banyak faktor termasuk ukuran dataset, karakteristik masalah, dan lainnya. cukup menentukan optimizer Adam dengan penentuan tuning parameter 0,01 lebih cocok dalam menghadapi masalah pengklasifikasian sentimen untuk tingkat kecemasan.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Penelitian ini membandingkan akurasi klasifikasi menggunakan BiLSTM sebelum dan setelah menerapkan pengujian k-fold cross validation serta penyesuaian tuning parameter, menghasilkan hasil terbaik. Penyesuaian parameter pada optimizer Adam dengan menyetel nilai learning rate menjadi 0,01 mampu meningkatkan performa model secara signifikan, terutama dalam hal tingkat akurasi. Rata-rata akurasi untuk aspek yang melibatkan emoji mencapai 98,08%, dengan aspek kesehatan mencapai akurasi 99,29%, aspek sosial mencapai 97,10%, dan aspek ekonomi mencapai 97,85%. Temuan ini mengindikasikan bahwa penyetelan parameter Adam secara efektif meningkatkan kinerja model, khususnya dalam menangani sentimen dengan tingkat kecemasan.

Hasil pengujian menunjukkan bahwa klasifikasi sentimen menggunakan algoritma BiLSTM dengan k-fold cross validation dan penyesuaian parameter pada dataset pelatihan yang diintegrasikan dengan emoji memberikan akurasi yang lebih baik daripada klasifikasi sentimen tanpa emoji. Meskipun emoji tidak secara langsung menentukan sentimen, penggunaannya membantu meningkatkan optimasi prediksi sentimen pada teks.

Kesimpulan dari penelitian menunjukkan bahwa metode Bidirectional Long Short-Term Memory (BiLSTM) berhasil diterapkan untuk analisis sentimen terkait tingkat kecemasan Covid-19 dan sosial-ekonomi masyarakat dengan emoji. Meskipun data tweet menunjukkan proporsi lebih banyak sentimen negatif, menandakan tingkat kecemasan yang masih tinggi terkait isu kesehatan, sosial, dan ekonomi, meskipun sudah dua tahun sejak pandemi Covid-19. Dampak pandemi tetap signifikan setelah berakhirnya, dan tingkat kecemasan yang tinggi menunjukkan bahwa Covid-19 bukanlah satu-satunya faktor yang mempengaruhi kecemasan masyarakat.

6.2 Saran

Pada penelitian ini hal yang perlu dilakukan pembaharuan adalah:

1. Dataset yang beragam, karena hasil *scrapping* masih mengandung banyak duplikasi tweets yang dapat mengganggu kemampuan model BiLSTM dalam mengenali sentimen dengan tepat.
2. Penggunaan emoji dengan memasukkan emoji berdasarkan polaritas sentimennya bertujuan untuk mengurangi variasi fitur emoji yang dapat mempengaruhi stabilitas kinerja model.
3. Dalam penggunaan K-Fold Cross Validation, ada kemungkinan untuk mencoba menggunakan parameter lain untuk meningkatkan keakuratan model.
4. Pihak pemerintah yang berwenang dalam bidang kesehatan dapat menggunakan hasil klasifikasi ini sebagai evaluasi, menunjukkan bahwa kesehatan mental masyarakat masih memerlukan perhatian yang serius.

DAFTAR PUSTAKA

- Aditama, M. I., Irfan Pratama, R., Hafizzana, K., Wiwaha, U., & Rakhmawati, N. A. (2020). Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19. <https://t.co/hlfyHfiZQj>
- Arifiyanti, A. A., & Wahyuni, E. D. (2020). Emoji And Emoticon In Tweet Sentiment Classification. *Proceeding - 6th Information Technology International Seminar, ITIS 2020, 145–150*. <https://doi.org/10.1109/ITIS50118.2020.9320988>
- Athallah, H. D. (2021). Analisis Terhadap Industri Otomotif.
- Chen, Y., Yuan, J., You, Q., & Luo, J. (2018). Twitter Sentiment Analysis via Bi-sense Emoji Embedding and. *26th ACM International Conference on Multimedia, 117–125*.
- Dewi Durrotun Yatimah. (2022). Interaksi Sosial Remaja Pada Masa Pandemi Covid-19 Di Dusun Jetak, Desa Dagen, Kecamatan Jaten, Kabupaten Karanganyar *skripsi. 8.5.2017, 1–120*. https://eprints.iain-surakarta.ac.id/3920/1/full_teks_skripsi_dewi_171221200.pdf
- Fadli, H. F., & Hidayatullah, A. F. (2021). Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM.
- Flores Veronica Ambassador, Lie Jasa, L. L. (2020). Analisis Sentimen Untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram. *Majalah Ilmiah Teknologi Elektro, 49(19.1)*.
- Grebb, J. A. (2010). Sinopsis Psikiatri: Ilmu Pengetahuan Perilaku Psikiatri Klinis. Edisi 2.
- Hung, M., Lauren, E., Hon, E. S., Birmingham, W. C., Xu, J., Su, S., Hon, S. D., Park, J., Dang, P., & Lipsky, M. S. (2020). Social network analysis of COVID-19 Sentiments: Application of artificial intelligence. *Journal of Medical Internet Research, 22(8)*. <https://doi.org/10.2196/22590>
- infobanknews.com. (2022). Menyimak Diferensiasi Perekonomian Global Pasca Covid-19 *Infobanknews*. <https://infobanknews.com/menyimak-diferensiasi-perekonomian-global-pasca-covid-19/>
- Jiawei Han, Jian Pei, H. T. (2022). Data Mining Concept and Techniques. *Fourth Edi*. [https://books.google.co.id/books?hl=id&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=KDD+method+by+han+et+al+2022&ots=_N1GUJykt2&sig=x9VbgYu4j9D90cwOMEa0-N4vtM&redir_esc=y#v=onepage&q=KDD method by han et al 2022&f=false](https://books.google.co.id/books?hl=id&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=KDD+method+by+han+et+al+2022&ots=_N1GUJykt2&sig=x9VbgYu4j9D90cwOMEa0-N4vtM&redir_esc=y#v=onepage&q=KDD+method+by+han+et+al+2022&f=false)
- Kaligis, F., Ismail, R. I., Wiguna, T., Prasetyo, S., Indriatmi, W., Gunardi, H., Pandia, V., & Magdalena, C. C. (2021). Mental Health Problems And Needs Among Transitional-Age Youth In Indonesia. *International Journal of Environmental Research and Public Health, 18(8)*. <https://doi.org/10.3390/ijerph18084046>
- Kementerian Kesehatan Republik Indonesia. (2022). PPKM di Indonesia Resmi Dicabut. Kementerian Kesehatan Republik Indonesia. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20221230/0042128/ppkm-di-indonesia-resmi-dicabut/>

- Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving Sentiment Analysis Accuracy With Emoji Embedding. *Journal of Safety Science and Resilience*, 2(4), 246–252. <https://doi.org/10.1016/j.jnlssr.2021.10.003>
- Nurrahman, I. (2021). Analisis Sentimen Penanganan Banjir DKI Jakarta.
- Pane, S. F., & Ramdan, J. (2022). Pemodelan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter. *Jurnal Sistem Cerdas*, 5(1), 12–20. <https://doi.org/10.37396/jsc.v5i1.191>
- PDSKJI. (2020). Masalah Psikologis di Era Pandemi Covid- 19. <http://www.pdskji.org/home>
- Rizki, A., & Sibaroni, Y. (2021). Analisis Sentimen Untuk Pengukuran Tingkat Depresi Pengguna. *E-Proceeding of Engineering*, 8(5), 11367–11375.
- Rokom. (2022). Indonesia Level Satu Transmisi Komunitas, Vaksinasi Tetap Gaspol. *Kemkes.Go.Id*. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20221203/2341917/indonesia-level-satu-transmisi-komunitas-vaksinasi-tetap-gaspol/>
- Rumelhart, David E., Geoffrey E. Hinton, and R. J. W. (1985). Learning internal representations by error propagation.
- Saxena, A., & Sukumar, T. R. (2018). Predicting bitcoin price using lstm And Compare its predictability with arima model. *International Journal of Pure and Applied Mathematics*, 119(17), 2591–2600. <https://doi.org/10.13140/RG.2.2.15847.57766>
- Setyo Nugroho, K., Akbar, I., & Nizar Suksmawati, A. (2021). Seminar Nasional Hasil Riset Prefix-Rtr Deteksi Depresi Dan Kecemasan Pengguna Twitter Menggunakan Bidirectional LSTM.
- Suwandi, G. R., Malinti, E., Fakultas, D., Keperawatan, I., & Uai, K. (2020). Hubungan Tingkat Pengetahuan Dengan Tingkat Kecemasan Terhadap Covid-19 Pada Remaja Di Sma Advent Balikpapan. In *Manuju: Malahayati Nursing Journal* (Vol. 2, Issue 4).
- Tavan, E., Rahmati, A., & Ali Keyvanrad, M. (2020). Persian Emoji Prediction Using Deep Learning and Emoji Embedding. *2020 10th International Conference on Computer and Knowledge Engineering, ICCKE 2020*, 350–355. <https://doi.org/10.1109/ICCKE50421.2020.9303639>
- Tosida, E. T., Erniyati, K., & Talib Bon, A. (2021). Sentiment Analysis Using the Support Vector Machine For Community Compliance Representation in The Covid-19 Pandemic Period.
- UGM, N. online. (2022). Hasil Penelitian: 2,4 Juta Remaja Indonesia Terkena Gangguan Jiwa. <https://www.nu.or.id/nasional/hasil-penelitian-2-4-juta-remaja-indonesia-terkena-gangguan-jiwa-gOeyb>
- Unicef, UNDP, Prospera, & SMERU. (2021). Analisis Dampak Sosial dan Ekonomi COVID-19 pada Rumah Tangga dan Rekomendasi Kebijakan Strategis untuk Indonesia. *SMERU Research Institute*, 1–7.

LAMPIRAN

Lampiran 1. SK Penelitian



YAYASAN PAKUAN SILIWANGI
Universitas Pakuan
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
Unggul, Mandiri & Berakhlak Dalam Bidang MIPA

KEPUTUSAN DEKAN
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PAKUAN
No.: 2089/D/FMIPA/VI/2022

T E N T A N G

PENGANGKATAN PEMBIMBING TUGAS AKHIR
PADA PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PAKUAN

DEKAN FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PAKUAN,

- Menimbang : a. bahwa setiap mahasiswa tingkat akhir Program Strata Satu (S1) harus melaksanakan Tugas Akhir sebagaimana tercantum di dalam kurikulum setiap Program Studi di lingkungan Fakultas MIPA Universitas Pakuan.
b. bahwa untuk pelaksanaan Tugas Akhir diperlukan pengawasan dari pembimbing.
c. bahwa sehubungan dengan point a dan b di atas perlu dituangkan dalam suatu Keputusan Dekan.
- Mengingat : 1. Undang-undang RI No.: 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah No.: 60 Tahun 1999 tentang Pendidikan Tinggi.
3. Statuta Universitas Pakuan Tahun 2019.
4. Surat Keputusan Rektor Nomor: 35/KEP/REK/VIII/2020 tanggal 03 Agustus 2020 tentang Pemberhentian Dekan dan Wakil Dekan Masa Bakti 2015-2020 serta Pengangkatan Dekan dan Wakil Dekan Masa Bakti 2020-2025 di lingkungan Universitas Pakuan.
5. Ketentuan Akademik yang tercantum dalam Buku Panduan Studi Fakultas MIPA, Universitas Pakuan Tahun 2021.
- Memperhatikan : Usulan dari Ketua Program Studi Ilmu Komputer FMIPA UNPAK.

M E M U T U S K A N

- Menetapkan :
Pertama : Mengangkat pembimbing yang namanya tersebut di bawah ini :
1. Pembimbing Utama : Eneng Tita Tosida, S.Tp., M.Si., M.Kom.
2. Pembimbing Pendamping : Adriana Sari Aryani, M.Cs

Untuk membimbing dalam rangka melaksanakan tugas akhir bagi mahasiswa :
Nama : Jenie Marcelina
NPM : 065118138
Program Studi : Ilmu Komputer
Judul Skripsi : Implementasi Algoritma Bilstm Dengan Analisis Sentimen
Emoji Embedded Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat

- Kedua : Kepada para pembimbing diharapkan dapat menjalankan tugasnya sebagai pembimbing dengan sebaik-baiknya.
- Ketiga : Dalam waktu 1 (satu) bulan setelah diterbitkannya SK ini, mahasiswa wajib melaksanakan Seminar Rencana Penelitian yang diselenggarakan oleh Program Studi Ilmu Komputer dengan dihadiri oleh Pembimbing dan Penguji.
- Keempat : Dana untuk honorarium pembimbing dibebankan kepada mahasiswa yang ketentuannya diatur oleh Fakultas MIPA.
- Kelima : Surat Keputusan ini berlaku untuk jangka waktu 1 (satu) tahun sejak tanggal ditetapkan sampai dengan mahasiswa tersebut Lulus Sidang/Ujian Skripsi, dengan ketentuan akan diadakan perubahan/perbaikan sebagaimana mestinya bila dikemudian hari terdapat kekeliruan dalam penetapannya.

Ditetapkan di : Bogor
Pada tanggal : 23 Juni 2022

Dekan,



Asep Denih, S.Kom., M.Sc., Ph.D.

Tembusan :

1. Yth. Ketua Program Studi Ilmu Komputer;
2. Yth. Eneng Tita Tosida, S.Tp., M.Si., M.Kom.;
3. Yth. Adriana Sari Aryani, M.Cs;
4. Arsip.

Lampiran 2. Kartu Bimbingan Mahasiswa

Lampiran 2. Kartu Bimbingan Mahasiswa

Kartu Bimbingan Mahasiswa Program Studi Ilmu Komputer FMIPA - UNPAK

Nama Mahasiswa : Jenie Marcelina
 NPM : 065118138
 Judul Skripsi : Implementasi Algoritma *Bidirectional Long Short Term Memory* (BiLSTM) Dengan Analisis Sentimen *Emoji Embedded* Terhadap Tingkat Kecemasan Covid-19 dan Sosial Ekonomi Masyarakat.
 Pembimbing Utama : **Dr. Eneng Tita Tosida, S.Tp., M.Si., M.Kom.**
 Pembimbing Pendamping : **Adriana Sari Aryani, M.Cs.**

No.	Tanggal	Catatan	Tanda Tangan	
			Pembimbing Utama	Pembimbing Pendamping
1.	13 Oktober 2022	Pentuan judul skripsi dan pengajuan SK Penelitian	1	Adriana
2.	24 Oktober 2022	Pengajuan rancangan proposal penelitian BAB 1 s/d BAB 4	2	Adriana
3.	2 November 2022	Revisi mengenai proses konversi emoji menjadi label dan pembobotan emoji, serta menambah jurnal penelitian terdahulu.	3	Adriana
4.	4 November 2022	Revisi bagian penulisan sitasi jurnal, penomoran persamaan, penulisan tabel emoji, serta penulisan daftar pustaka.	4	Adriana
5.	12 November 2022	Membuat PPT dan Revisi PPT untuk persiapan Seminar Proposal	5	Adriana
6.	14 November 2022	Pengajuan Final Draft proposal dan PPT untuk Seminar Proposal	6	Adriana
7.	24 Januari 2023	Koordinasi mengenai ketentuan kriteria pakar yang akan di minta kerjasamai dalam penelitian untuk penentuan sentiment.	7	Adriana
8.	7 November 2023	Bimbingan pengajuan draft Hasil berdasarkan revisi dari draft proposal pada seminar proposal sebelumnya.	8	Adriana
9.	8 November 2023	Pengajuan Revisi Hasil, perbaikan meliputi penulisan contoh data yang dipakai perbaikan proses perhitungan manualisasi di BAB 4 dan melanjutkan pengerjaan BAB 5 s/d BAB 6.	9	Adriana
10.	5 Januari 2024	Bimbingan terkait progress skripsi, pengajuan draft BAB 5	10	

11.	6 Januari 2024	Pengajuan revisi rancangan Bab 5 dan BAB 6, dengan revisi penambahan informasi jenis emoji yang teridentifikasi dalam teks, penambahan tabel emoji yang teridentifikasi, dijelaskan kembali bagaimana pengintegrasian emoji dalam teks, merapihkan lampiran, serta pembuatan PPT Seminar Hasil.	21	Adriana
12.	6 Januari 2024	Pengajuan draft dengan sebelumnya penambahan revisi untuk menghilangkan gambar yang tidak perlu, pemangkasan halaman draft laporan, dan masukkan untuk penambahan tuning parameter, penambahan pembahasan optimizer ADAM, perbaiki sitasi literatur di BAB 4 dan BAB 5, serta penambahan metode <i>K-Fold Cross Validation</i> .	23	Adriana
13.	8 Januari 2024	Pengajuan draft revisi laporan hasil berdasarkan revisi bimbingan sebelumnya dan PPT seminar hasil untuk segera daftar Seminar Hasil	25	Adriana
14.	14 Januari 2024	Pengajuan draft revisi laporan hasil berdasarkan revisi bimbingan sebelumnya dan Revisi PPT seminar hasil untuk segera daftar Seminar Hasil	27	Adriana
15.	25 Januari 2024	Pengajuan revisi draft laporan hasil berdasarkan revisi dari pelaksanaan Seminar Hasil dengan revisi penambahan metode <i>K-Fold Cross Validation</i> dan kemudian dibandingkan dengan analisis yang tanpa menggunakan metode <i>K-Fold Cross Validation</i> , serta memperbaiki penulisan Kesimpulan.	29	Adriana
16.	26 Januari 2024	Pengajuan revisi PPT untuk Sidang Skripsi dengan revisi menambahkan penjelasan teknis nilai parameter dan nilai akurasi dari tiap bidang di bagian kesimpulan.	31	Adriana

Bogor, 26 Februari 2024
Program Studi Ilmu Komputer
Fakultas MIPA – UNPAK
Ketua,

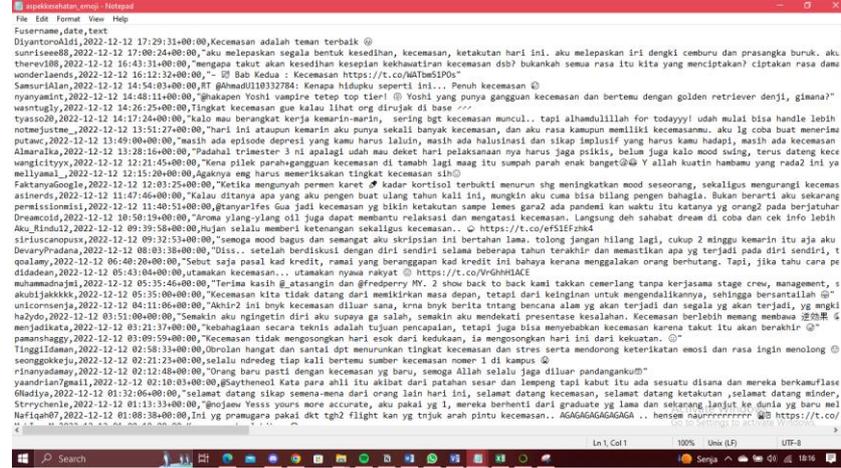


Arie Qur'ania, S.Kom., M.Kom.

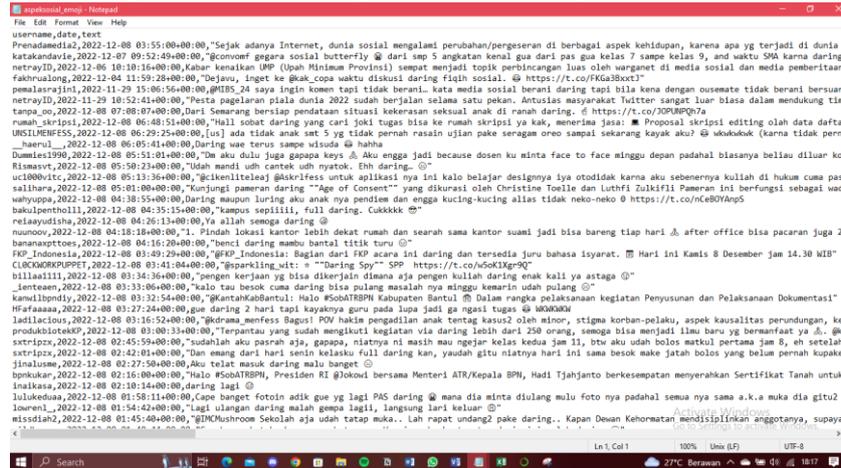
Lampiran 3. Gambar Proses Scrapping Tweet Kecemasan Dalam 3 Aspek

Pengambilan data atau *Scrapping* data pada media sosial Twitter dilakukan pada tanggal 1 s/d 31 Desember 2022.

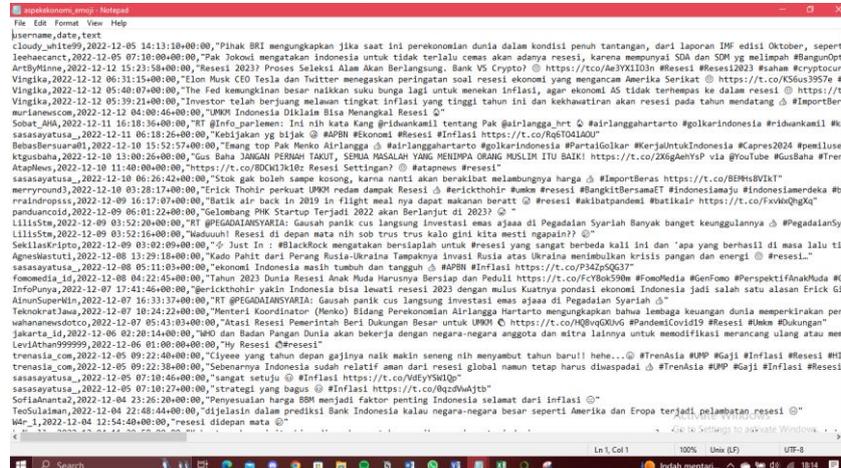
- Hasil Crawling Data Twitter Aspek Kesehatan + emoji



- Hasil Crawling Data Twitter Aspek Sosial + emoji



- Hasil Crawling Data Twitter Aspek Ekonomi + emoji



Lampiran 4. Hasil Wawancara Dengan Pakar Bahasa Indonesia

1. Dari data tweet pada file yang saya miliki untuk pelabelan sentimennya apakah sudah sesuai untuk konteks label negative, neutral, dan positive?

Jawab :

Sudah sesuai. saya telah membaca beberapa kali dan menurut pendapat saya ada beberapa faktor yang bisa dijadikan sebagai dasar untuk melabeli data tersebut, diantara lainnya seperti faktor usia ataupun gender dan juga konteks dari permasalahan yang ditanggapi oleh masyarakat, kedua hal tersebut dapat sangat mempengaruhi sentimen, namun dari data ini dapat terlihat dari secara Bahasa bahwa ada yang sangat terlihat jelas atau samar sentimen nya. Yang terlihat jelas bisa dikategorikan anantara *positive* dan *negative* tapi yang tidak jelas bisa dikategorikan *neutral* tapi bukan berarti tidak memiliki sentimen namun hanya saja samar. Dan untuk melabel data ini juga mimik wajah sangat berpengaruh karena terdapat imbuhan emoji di dalam tweet, emoji ini menjadi faktor yang sangat berpengaruh bahkan sangat besar untuk menyiratkan sesuatu daripada kalimat itu sendiri. Jadi emoji sangat berpengaruh dalam pelabelan sentimen.

2. Menurut bapak, apakah kata kunci / keywords yang saya gunakan seperti ‘kesehatan mental’, ‘cemas’, ‘covid-19’, ‘vaksin’, dan ‘pelayanan kesehatan’ pada aspek kesehatan, ‘sosial’, ‘psbb’, ‘interaksi’, ‘stigma’, ‘daring’ pada aspek sosial, dan ‘ekonomi’, ‘ppkm’, ‘resesi’, ‘inflasi’ pada aspek ekonomi, bisa merepresentasikan untuk menggali pengetahuan dalam melihat respon kecemasan masyarakat indonesia terhadap fenomena dan aktivitas masyarakat sehari-hari?

Jawab :

Cukup merepresentasikan, karna kata kunci menjadi sebuah trigger untuk masyarakat sehingga tanggapan itu menjadi mengarah ke dalam suatu hal yang mewakili rasa kecemasan masyarakat itu sendiri.

Lampiran 5. Biodata Pakar Bahasa Indonesia

DAFTAR RIWAYAT HIDUP

NAMA : MOHAMAD SOFYANDI
GELAR : SARJANA PENDIDIKAN
TEMPAT LAHIR : CIANJUR
TANGGAL LAHIR : 9 APRIL 1964
ALAMAT : PERUM PURI PERSADA INDAH JALAN GARUDA 1
BLOK R NOMOR 12 RT 04/RW 12 DESA SINDANG MULYA
KEC. CIBARUSAH KABUPATEN BEKASI
HP : 081321079070



RIWAYAT PENDIDIKAN
SEKOLAH DASAR : SD SWASTA DARUL HIKAM KOTA BANDUNG
LULUS TAHUN 1976
SEKOLAH MENENGAH PERTAMA : SMP NEGERI 7 KOTA BANDUNG
LULUS TAHUN 1980
SEKOLAH MENENGAH ATAS : SMA NEGERI 1 KOTA BANDUNG
LULUS TAHUN 1983
PESANTREN : PESANTREN MIFTAHUL HUDA TASIKMALAYA
MASUK TAHUN 1983 - KELUAR TAHUN 1984
PERGURUAN TINGGI : IKIP BANDUNG
PROGRAM D3 JURUSAN BAHASA DAN SASTRA INDONESIA
MASUK TAHUN 1984
LULUS TAHUN 1987
PERGURUAN TINGGI : UNIVERSITAS PENDIDIKAN INDONESIA (UPI) BANDUNG
PROGRAM S1 JURUSAN BAHASA DAN SASTRA INDONESIA
MASUK TAHUN 1991
LULUS TAHUN 1994
PERGURUAN TINGGI : UNIVERSITAS PENDIDIKAN INDONESIA (UPI) BANDUNG
PROGRAM S2 JURUSAN BAHASA DAN SASTRA INDONESIA
MASUK TAHUN 2001
TIDAK LULUS

RIWAYAT PEKERJAAN
GURU PNS : SMA NEGERI 1 CIBARUSAH KABUPATEN BEKASI
CPNS TAHUN 1988 DAN PNS TAHUN 1989
MENGAJAR BAHASA DAN SASTRA INDONESIA
SAMPAI SEKARANG

RIWAYAT JABATAN : **WAKIL KEPALA SEKOLAH BIDANG KESISWAAN**
(*SMA Negeri 1 Cibarusah Kabupaten Bekasi*)
- MASA KERJA TAHUN 1999 - TAHUN 2001
- MASA KERJA TAHUN 2007 - TAHUN 2014
- MASA KERJA TAHUN 2016 - TAHUN 2019
- MASA KERJA TAHUN 2020 - TAHUN 2021
WAKIL KEPALA SEKOLAH BIDANG HUMAS
- MASA KERJA TAHUN 2021 - TAHUN 2022
WAKIL KEPALA SEKOLAH BIDANG KURIKULUM
- MASA KERJA TAHUN 2022 - TAHUN 2023

RIWAYAT ORGANISASI : PELAJAR ISLAM INDONESIA (PII) KOTA BANDUNG
GERAKAN PRAMUKA KABUPATEN BEKASI

Lampiran 6. Deskripsi Label Nama Dan Volume Emoji Dalam Dataset Setiap Aspek

- Jenis dan volume emoji pada aspek kesehatan

No.	Label Nama Emoji	Jumlah
1	loudly crying face	138
2	face with tears of joy	97
3	rolling on the floor laughing	89
4	smiling face	71
5	thumbs up	67
6	folded hands	61
7	grinning face	48
8	relieved face	45
9	slightly smiling face	39
10	white heavy check mark	32
11	frowning face	32
12	red heart	30
13	winking face	29
14	smiling face with smiling eyes	29
15	pensive face	27
16	tired face	26
17	thumbs down	23
18	thumbs down	23
19	flexed biceps	22
20	angry face	22
21	grinning face with sweat	21
22	thinking face	21
23	downcast face with sweat	20
24	beaming face with smiling eyes	19
25	smiling face with 3 heart	18
26	sad but relieved face	18
27	hugging face	17
28	smiling face with halo	16
29	worried face	16
30	sparkles	15
31	grinning face with smiling eyes	13
32	index pointing up	13
33	slightly frowning face	12
34	face with steam from nose	12
35	face screaming in fear	11
36	ok hands	11
37	palms up together	11
38	expressionless face	11
39	crying face	11
40	person facepalming	11
41	smirking face	10

No.	Label Nama Emoji	Jumlah
70	smiling face with sunglasses	5
71	middle finger	5
72	person shrugging	5
73	dizzy	5
74	woozy face	4
75	dizzy face	4
76	speak no evil monkey	4
77	grinning face with big eyes	4
78	face with symbols on mouth	4
79	sleepy face	4
80	face blowing a kiss	4
81	kissing face	4
82	wilted flower	3
83	heart suit	3
84	confounded face	3
85	face without mouth	3
86	pouting face	2
87	revolving hearts	2
88	writing hand	2
89	t-rex	2
90	backhand index pointing down	2
91	no entry	2
92	face savoring food	2
93	hundred points	2
94	zany face	2
95	speech	2
96	kissing face with smiling eyes	2
97	exclamation mark	2
98	nauseated face	2
99	dissapointed face	1
100	police car light	1
101	play button	1
102	pushpin	1
103	up right arrow	1
104	candy	1
105	cloud with rain	1
106	money bag	1
107	shushing face	1
108	small blue diamond	1
109	eyes	1
110	heavy heart	1

42	weary face	10
43	backhand index pointing right	9
44	grimacing face	9
45	confused face	9
46	face with rolling eyes	9
47	grinning squinting face	9
48	oncoming fist	9
49	persevering face	9
50	broken heart	8
51	star struck	8
52	clown face	8
53	pleading face	8
54	anxious face with sweat	8
55	fearful face	8
56	skull	8
57	clapping hands	7
58	upside down face	6
59	warning	6
60	face vomiting	6
61	unamused face	6
62	face with raised eyebrow	6
63	fire	6
64	face with open mouth	6
65	smiling face with heart eyes	5
66	face with hand over mouth	5
67	sneezing face	5
68	cross mark	5
69	face with medical mask	5

111	exclamation sunflower	1
112	chart increasing	1
113	nail polish	1
114	growing heart	1
115	crying cat face	1
116	bar chart	1
117	chart decreasing	1
118	party popper	1
119	raising hands	1
120	anguished face	1
121	purple heart	1
122	victory hand	1
123	person gesturing ok	1
124	raised hand	1
125	face with monocle	1
126	two hearts	1
127	red triangle pointed up	1
128	face with head bandage	1
129	balloon	1
130	umbrella	1
131	person gesturing no	1
132	adult	1
133	cat	1
134	lying face	1
135	cold face	1
136	skull and crossbones	1
137	call me hand	1
138	person bowing	1

- Jenis dan volume emoji pada aspek sosial

No.	Label Nama Emoji	Jumlah
1	loudly crying face	127
2	smiling face	98
3	face with tears of joy	82
4	rolling on the floor laughing	64
5	frowning face	52
6	grinning face	50
7	folded hand	36
8	thumbs up	35
9	relieved face	35
10	grinning face with sweat	31
11	angry face	31
12	slightly smiling face	23
13	thinking face	23
14	smiling face with smiling eyes	21
15	pensive face	19

No.	Label Nama Emoji	Jumlah
78	raised fist	3
79	grimacing face	3
80	raised hand	3
81	purple heart	3
82	star-struck	3
83	cat face with wry smile	3
84	person getting massage	3
85	clock	3
86	face with tongue	3
87	person bowing	3
88	face with symbols on mouth	3
89	woman	3
90	blank note	2
91	bust in silhouette	2
92	Backhand index pointing down	2

16	beaming face with smiling eyes	16
17	hugging face	16
18	red heart	15
19	raising hands	14
20	upside-down face	14
21	winking face	13
22	smirking face	12
23	smiling face with heart-eyes	12
24	wearry face	12
25	tired face	12
26	index pointing right	12
27	smiling face with 3 hearts	11
28	grinning squinting face	11
29	person facepalming	11
30	exploding head	10
31	grinning face with smiling eyes	10
32	confused face	10
33	backhand index pointing right	9
34	pleading face	9
35	unamused face	9
36	worried face	9
37	face with hand over mouth	8
38	index pointing up	8
39	face with rolling eyes	8
40	face with steam from nose	8
41	anxious face with sweat	8
42	dissapointed face	7
43	slightly frowning face	7
44	ok hands	7
45	thumbs down	7
46	sparkles	7
47	person shrugging	7
48	expressionless face	7
49	palms up together	6
50	broken heart	6
51	sleepy face	6
52	handshake	6
53	fire	6
54	downcast face with sweat	6
55	smiling face with sunglasses	5
56	woozy face	5
57	shushing face	5
58	victory hand	5
59	eyes	5

93	hand with finger splayed	2
94	sleeping face	2
95	person gesturing ok	2
96	person gesturing no	2
97	clown face	2
98	zany face	2
99	heart suit	2
100	crescent moon	2
101	neutral face	2
102	blue heart	2
103	waving hand	2
104	party popper	2
105	button	2
106	cat face	2
107	pouting cat face	2
108	smiling with horns	2
109	middle finger	2
110	angry face with horns	2
111	police car light	2
112	confounded face	2
113	face without mouth	2
114	fearful face	2
115	play button	1
116	Face with monocle	1
117	Writing hand	1
118	star	1
119	drooling face	1
120	call me hand	1
121	woman in lotus position	1
122	right-facing fist	1
123	red envelope ribbon	1
124	calendar	1
125	face with thermometer	1
126	grinning cat face	1
127	flushed face	1
128	man facepalming	1
129	pouting face	1
130	man and woman holding hands	1
131	sun with face	1
132	palm tree	1
133	woman surfing sparkler	1
134	japanese dolls	2
135	adult	1
136	lying face	1

60	perserving face	5
61	face blowing a kiss	5
62	white heavy check mark	5
63	nauseated face	5
64	dizzy face	5
65	oncoming fist	5
66	sad but relieved face	5
67	backhand index pointing left	4
68	speak no evil monkey	4
69	heart with arrow	4
70	kissing face	4
71	clapping hands	4
72	face with raised eyebrow	4
73	flexed biceps	4
74	Spiral calendar	3
75	sneezing face	3
76	smiling face with halo	3
77	grinning face with big eyes	3

137	whale	1
138	see no evil monkey	1
139	kissing face with closed eyes	1
140	black heart	1
141	woman bowing	1
142	crying cat face	1
143	cold face	1
144	person in lotus position	1
145	face vomiting	1
146	seedling	1
147	next right arrow curving down	1
148	sweat droplets	1
149	face with medical mask	1
150	squinting face with tongue	1
151	frowning face with open mouth	1
152	heart with ribbon	1
153	santa claus	1

- Jenis dan volume emoji pada aspek ekonomi

No.	Label Nama Emoji	Jumlah
1	face with tears of joy	137
2	rolling on the floor laughing	107
3	smiling face	98
4	loudly crying face	90
5	grinning face	89
6	thumbs up	63
7	thumbs up	63
8	folded hands	52
9	grinning face with sweat	52
10	relieved face	38
11	smiling face with smiling eyes	30
12	thinking face	30
13	clapping hands	27
14	frowning face	26
15	face with hand over mouth	25
16	index pointing right	25
17	slightly smiling face	24

No.	Label Nama Emoji	Jumlah
66	face with steam from nose	5
67	person bowing	5
68	confused face	5
69	backhand index pointing left	5
70	face savoring food	4
71	pleading face	4
72	frowning face with open mouth	4
73	warning	4
74	oncoming fist	4
75	thumbs down	4
76	love you gesture	3
77	expressionless face	3
78	smiling face with sunglasses	3
79	victory hand	3
80	exploding head	3
81	clown face	3
82	hushed face	3

18	grinning face with smiling eyes	23
19	beaming face with smiling eyes	21
20	pensive face	20
21	winking face	20
22	crying face	19
23	weary face	19
24	flexed biceps	19
25	hugging face	19
26	waving hand	18
27	worried face	17
28	backhand index pointing right	16
29	smiling face with 3 hearts	16
30	smirking face	16
31	fire	16
32	index pointing up	16
33	angry face	15
34	person facepalming	15
35	face with rolling eyes	14
36	face screaming in fear	14
37	tired face	13
38	sparkles	13
39	dizzy	11
40	dizzy face	10
41	upside down face	10
42	palms up together	10
43	red heart	10
44	face with open mouth	10
45	smiling face with halo	10
46	sneezing face	9
47	sleepy face	9
48	raising hands	9
49	persevering face	9
50	ok hand	8
51	face with symbols on mouth	8
52	face with medical mask	8
53	handshake	8
54	neutral face	7
55	fearful face	7
56	anxious face with sweat	7
57	sad but relieved face	7
58	star-struck	6
59	downcast face with sweat	6
60	unamused face	6

83	chart decreasing	3
84	squinting face with tongue	3
85	face with monocle	3
86	skull	3
87	writing hand	3
88	chart increasing	3
89	index pointing down	2
90	zany face	2
91	walking	2
92	woman facepalming	2
93	woozy face	2
94	hundred points	2
95	adult	2
96	grinning face with big eyes	2
97	confounded face	2
98	middle finger	2
99	whitepaper	2
100	winking face with tongue	2
101	slightly frowning face	2
102	child	2
103	money with wings	2
104	memo	2
105	person getting massage	1
106	zombie	1
107	green heart	1
108	cyclone	1
109	cross mark button	1
110	white heavy check mark	1
111	person in lotus position	1
112	two hearts	1
113	anguished face	1
114	hear no evil monkey	1
115	police officer	1
116	police car	1
117	woman cartwheeling	1
118	woman walking	1
119	play or pause button	1
120	woman gesturing ok	1
121	sparkling heart	1
122	nauseated face	1
123	angry face with horns	1
124	graduation cap	1
125	grinning cat face with smiling eyes	1

61	raised first	5
62	broken heart	5
63	grimacing face	5
64	person shrugging	5
65	grinning squinting face	5

126	woman dancing	1
127	eyes	1
128	cold face	1
129	blue heart	1

Lampiran 7. Perhitungan skor sentimen contoh data dan perhitungan manualnya.

Data	Tweet	Positif	Negatif	Jumlah	Sentiment	Hasil Perhitungan
2.	Para pemimpin dan kabinet banyak yang mundur serta huru-hara di setiap negara dan kami sudah katakan tidak ada yang bisa keluar dari tekanan tatanan dunia baru atas kehancuran ekonomi. apalagi inflasi +62 termasuk lima besar se asia tenggara dan masuk 10 besar negara yang mengalami krisis pangan frowning face	Negara Dunia Ekonomi Masuk Alami	Mundur Tekan Hancur Inflasi Krisis Kerut	5 – 6 = -1	Negative	Positive
	['para', 'leader', 'cabinet', 'resign', 'riot', 'harass', 'state', 'press', 'order', 'world', 'destroy', 'economy', 'inflation', 'asia', 'southeast', 'entry', 'country', 'natural', 'crisis', 'food', 'worried', 'face']	State World Economy Entry Natural	Resign Press Destroy Inflation Crisis frowning			
3.	Stigma masyarakat kita selalu terlihat kalau menginjak umur 25 tahun belum kawin ini ketinggalan fase bahagia sedih sekali.. biarlah orang ingin kawin pada umur dan waktu kemampuan masing masing ingin ikut gaya hidup orang lain dengan kawin muda sudahlah akan ada yang struggle dengan masalah ekonomi dan keuangan	Masyarakat Bahagia Biar Orang Mampu Hidup Ekonomi Uang	Stigma Injak Sedih Struggle Frowning	8 – 5 = 3 3 > 0	Positive	Positive
	['stigma', 'society', 'trample', 'age', 'marry', 'stay', 'phase', 'happy', 'sad', 'let', 'people', 'marry', 'age', 'able', 'style', 'life', 'people', 'married', 'young', 'struggle', 'economy', 'money', 'frowning', 'face']	Society Happy Let People Able Life Economy Money	Stigma Trample Sad Struggle Frowning			
4.	pemulihan ekonomi global saat ini masih dibayangi sejumlah eskalasi risiko global dan tantangan yang harus disikapi dengan hati-hati seperti inflasi geopolitik dan potensi resesi. Meski begitu Indonesia	Pulih Potensi Negara Pulih Apresiasi	Risiko Inflasi Resesi Pandemi Gelap Exploding Screaming	5 – 8 = -3 -3 < 0	Negative	Negative

	menjadi satu di antara negara yang berhasil pulih melewati pandemi dan kita perlu mengapresiasi karna kita indonesia masih bisa bertahan sampai saat ini dan semoga ditahun depan tidak benar menjadi tahun yang gelap exploding head face screaming in fear		Fear			
	['recover', 'economy', 'global', 'shadow', 'escalation', 'risk', 'global', 'challenge', 'attitude', 'caution', 'inflation', 'geopolitics', 'potential', 'recession', 'indonesia', 'country', 'results', 'recovery', 'through', 'pandemic', 'appreciation', 'because', 'indonesia', 'hold on', 'hopefully', 'year', 'dark', 'exploding', 'head', 'face', 'screaming', 'in', 'fear']	Recover Potential Country Recovery Appreciation	Risk Inflation Recession Pandemic dark Exploding Screaming Fear			
5.	Untuk yang punya mental health issue saya saranin jangan terlalu spesifik nyebutinnya memang kultur sudah mulai terbuka tapi masih ada banyak stigma sama orang bias yang memandang hal itu secara sebelah mata dianggap lemah tidak kuat mental loudly crying face	Mental Health Kultur Orang Kuat Saran	Issue Stigma Bias Lemah Crying	6 – 5 = 1 1 > 0	Positive	Positive
	['mental', 'health', 'issue', 'me', 'saranin', 'specific', 'mention it', 'culture', 'open', 'stigma', 'people', 'bias', 'look', 'split', 'eye', 'perceive', 'weak', 'strong', 'mental', 'loudly', 'crying', 'face']	Mental Health Culture People Strong Advice	Issue Stigma Bias Weak Crying			

Lampiran 8. Proses Cleaning Data

- Proses preprocessing

Pada proses ini, data– data yang sudah didapat setelah melakukan Crawling untuk menyamakan karakteristik datanya, dilakukan processing dengan program sebagai berikut:

```
def convert_emoji(text):
    # Convert emojis to text representation
    text = emoji.demojize(text)

    # Add more specific replacements as needed
    emoji_replacements = {
        ':smile:': 'happy',
        ':cry:': 'sad',
        ':heart:': 'love',
        # Add more emoji representations here
    }

    for emoji_code, replacement in emoji_replacements.items():
        text = text.replace(emoji_code, f' {replacement} ')

    return text
df['text'] = df['text'].apply(convert_emoji)

print(df['text'])
```

Proses converting emoji

```
def remove_user_hashtag_url_mention(text):
    text = re.sub(r'@\S+|https?://\S+|www\.\S+|#\S+', '', text)
    return text

remove_user_hashtag = df['text'].apply(remove_user_hashtag_url_mention)

df['text'] = df['text'].apply(convert_emoji)
df['text_no_user_hashtag'] = df['text'].apply(remove_user_hashtag_url_mention)

print(df[['text_no_user_hashtag']])
```

Proses remove username, hashtag, url, & mention

```
def remove_punctuation(text):
    text = re.sub(r'^\W\s', '', text)
    return text

df['text_no_punctuation'] = df['text_no_user_hashtag'].apply(remove_punctuation)

print(df[['text', 'text_no_user_hashtag', 'text_no_punctuation']])
```

Proses punctuation

```
def case_folding(text):
    return text.lower()
df['text_lower'] = df['text_no_punctuation'].apply(case_folding)

print(df[['text', 'text_no_user_hashtag', 'text_no_punctuation', 'text_lower']])
```

Proses case folding

```
def tokenize(text):
    return word_tokenize(text)
df['tokens'] = df['text_lower'].apply(tokenize)
print(df[['text', 'text_no_user_hashtag', 'text_no_punctuation', 'text_lower', 'tokens']])
```

Proses tokenize

```

def words(text):
    return re.findall(r'\w+', text.lower())

WORDS = Counter(words(open('kata-dasar.txt').read()))

def P(word, N=sum(WORDS.values())):
    return WORDS[word] / N

def correction(word):
    return max(candidates(word), key=P)

def candidates(word):
    return (known([word]) or known(edits1(word)) or known(edits2(word)) or [word])

def known(words):
    return set(w for w in words if w in WORDS)

def edits1(word):
    letters = 'abcdefghijklmnopqrstuvwxyz'
    splits = [(word[:i], word[i:]) for i in range(len(word) + 1)]
    deletes = [L + R[1:] for L, R in splits if R]
    transposes = [L + R[1] + R[0] + R[2:] for L, R in splits if len(R) > 1]
    replaces = [L + c + R[1:] for L, R in splits if R for c in letters]
    inserts = [L + c + R for L, R in splits for c in letters]
    return set(deletes + transposes + replaces + inserts)

def edits2(word):
    return (e2 for e1 in edits1(word) for e2 in edits1(e1))

def correct_typos(tokens):
    return ' '.join([correction(word) for word in tokens])

df['text_corrected'] = df['tokens'].apply(correct_typos)

print(df[['text', 'text_no_user_hashtag', 'text_no_punctuation', 'text_lower', 'tokens', 'text_corrected']])

```

Proses koreksi text

```

def translate_to_english(tokens, max_retries=3):
    for _ in range(max_retries):
        try:
            translated_text = mtranslate_translate(' '.join(tokens), 'en', 'id')
            return translated_text.split()
        except Exception as e:
            print(f"Translation error: {e}. Retrying...")
            time.sleep(2)

    print("Max retries reached. Translation failed.")
    return tokens

# Example usage:
# Assuming you have a DataFrame 'df' with a column 'tokens', you can apply the translation as follows:
df['tokens_english'] = df['tokens'].apply(translate_to_english)

print(df[['tokens', 'tokens_english']])

```

Proses translate bahasa Indonesia ke bahasa inggris

```

nltk.data.path.append("aspekkesehatan_extract_emoji_labelling.csv")

def clean_text(text):
    text = convert_emoji(text)
    text = remove_user_hashtag_url_mention(text)
    text = remove_punctuation(text)
    text = case_folding(text)

    tokens = word_tokenize(text)

    stop_words_id = set(stopwords.words('indonesian'))
    tokens = [word for word in tokens if word.lower() not in stop_words_id]

    translated_tokens = translate_to_english(tokens)

    stemmer = StemmerFactory().create_stemmer()
    stemmed_tokens = [stemmer.stem(word) for word in translated_tokens]

    return ' '.join(stemmed_tokens)

with ThreadPoolExecutor(max_workers=4) as executor:
    tqdm.pandas()
    df['text'] = list(tqdm(executor.map(clean_text, df['text_corrected']), total=len(df)))

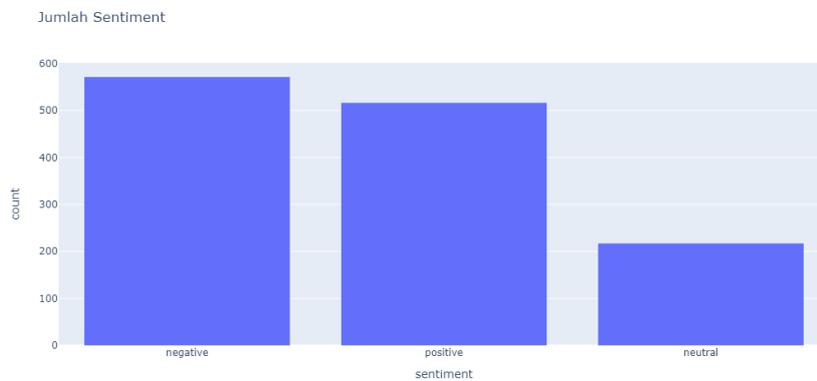
df.to_csv('pre_kesehatan.csv', columns=['text', 'sentiment'], index=False)

```

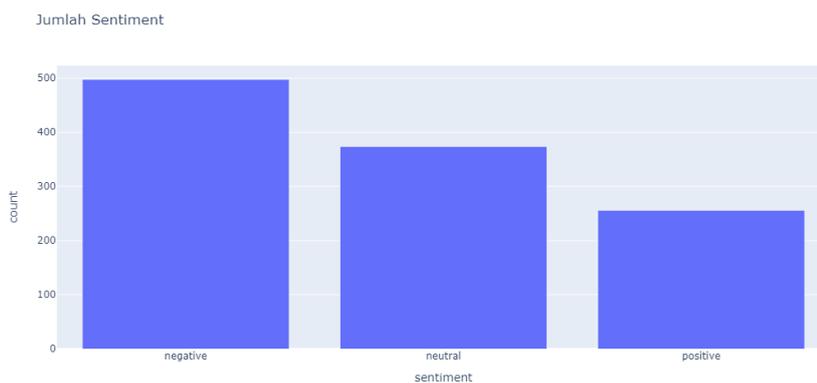
Proses cleaning data secara keseluruhan

Lampiran 9. Hasil Klasifikasi Aspek Tanpa Penyematan Emoji

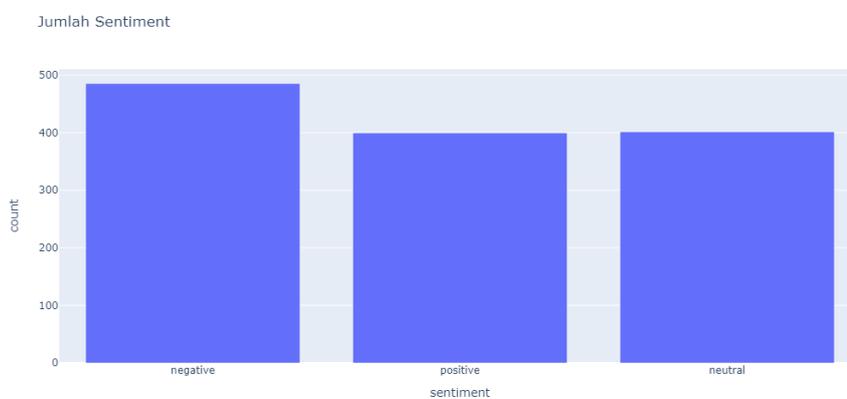
- Klasifikasi Aspek Kesehatan Tanpa Emoji dengan jumlah sentiment negatif sebanyak 571, sentiment positif sebanyak 516, dan sentiment neutral sebanyak 217.



- Klasifikasi Aspek Sosial Tanpa Emoji dengan jumlah sentiment negatif sebanyak 497, sentiment positif sebanyak 255, dan sentiment neutral sebanyak 373.

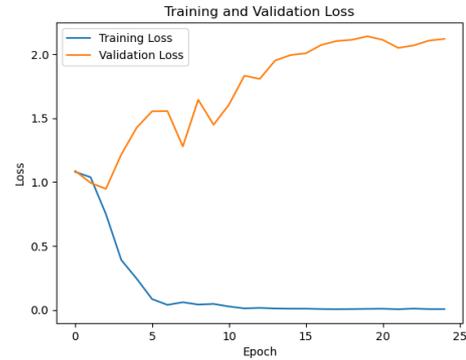
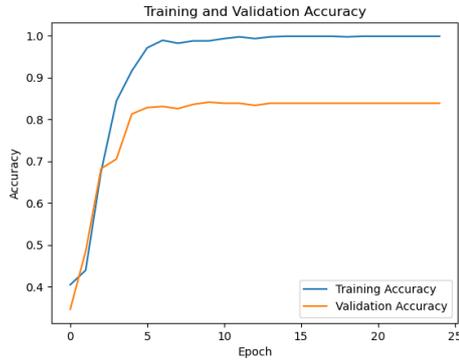


- Klasifikasi Aspek Ekonomi Tanpa Emoji dengan jumlah sentiment negatif sebanyak 485, sentiment positif sebanyak 399, dan sentiment neutral sebanyak 401.

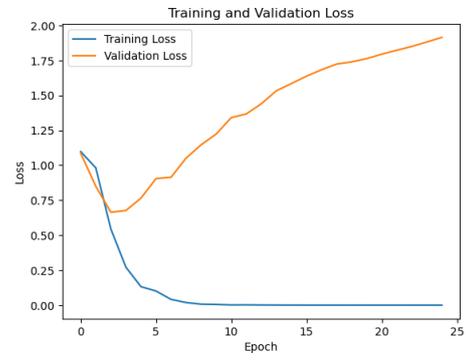
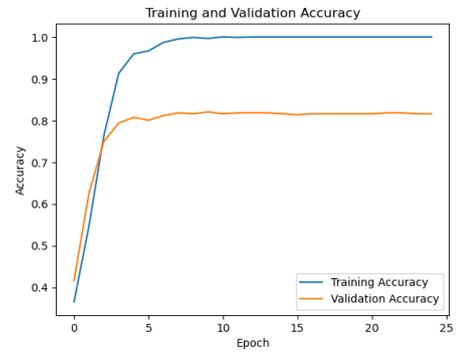


Lampiran 10. Hasil Training Dan Validation Seluruh Aspek Dengan Atau Tanpa Emoji

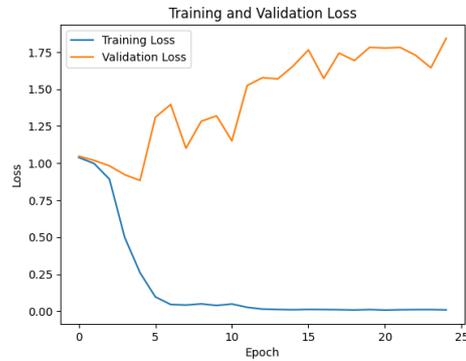
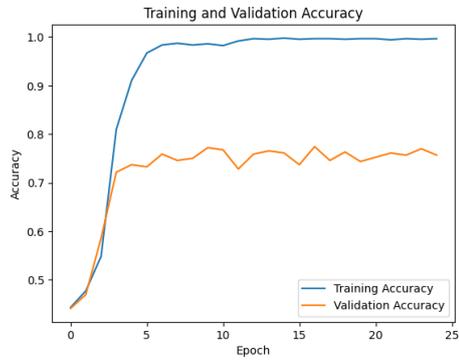
- Hasil training dan validation accuracy & loss aspek sosial dengan emoji



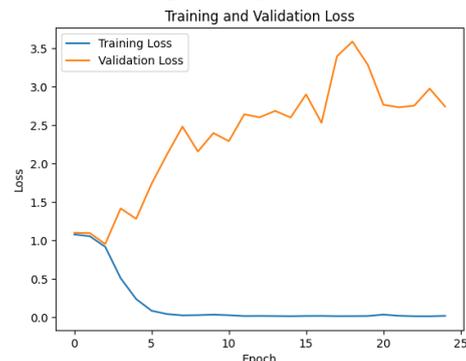
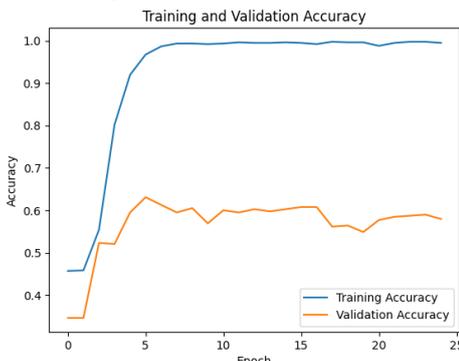
- Hasil training dan validation accuracy & loss aspek ekonomi dengan emoji



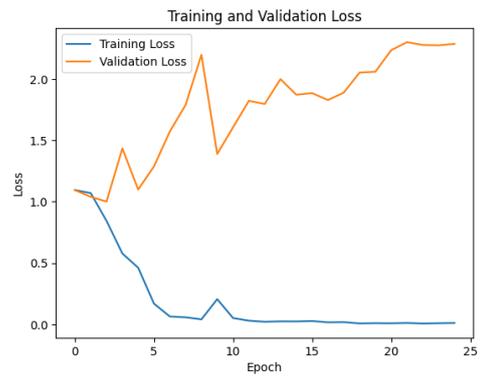
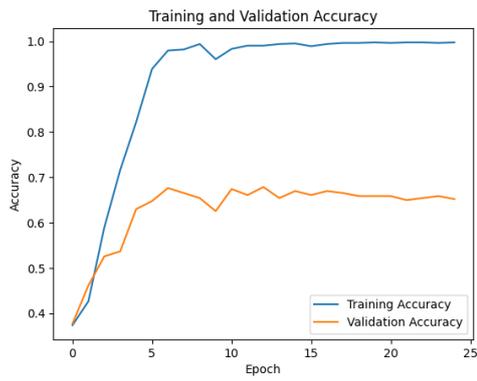
- Hasil training dan validation accuracy & loss aspek kesehatan tanpa emoji



- Hasil training dan validation accuracy & loss aspek sosial tanpa emoji

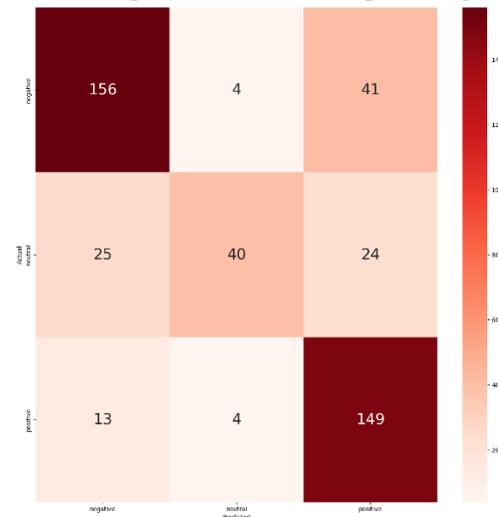


- Hasil training dan validation accuracy & loss aspek ekonomi tanpa emoji

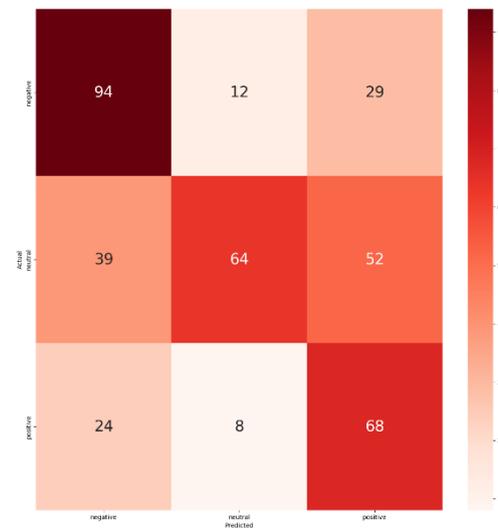


Lampiran 11. Hasil Confussion Matrix Seluruh Aspek Tanpa Penyematan Emoji

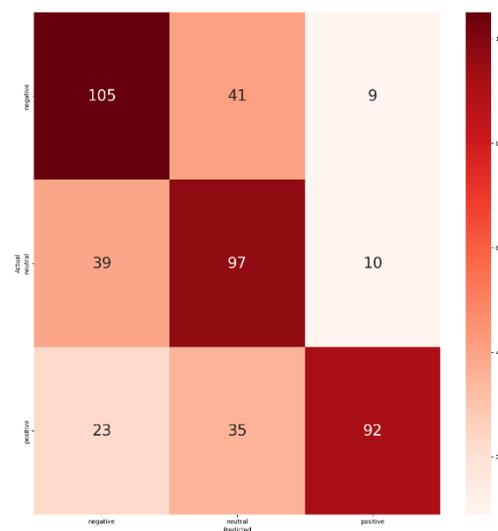
- Hasil *Confussion Matrix* Aspek Kesehatan tanpa Emoji



- Hasil *Confussion Matrix* Aspek Sosial tanpa Emoji



- Hasil *Confussion Matrix* Aspek Ekonomi tanpa Emoji



Lampiran 12. Hasil Wordcloud Seluruh Aspek Tanpa Penyematan Emoji.

- Hasil *Wordcloud* Aspek Kesehatan tanpa Emoji



- Hasil *Wordcloud* Aspek Sosial tanpa Emoji



- Hasil *Wordcloud* Aspek Ekonomi tanpa Emoji

