APLIKASI DETEKSI EMOSI PADA TEKS MENGGUNAKAN NATURAL LANGUAGE PROCESSING

SKRIPSI

DIAJUKAN SEBAGAI SALAG SATU SYARAT
UNTUK MENYUSUN SKRIPSI PROGRAM STUDI TEKNIK INFORMATIKA



Oleh:

NPM	**: 200210038**
NAMA	**: MUHAMMAD HAMBALI**
JENJANG STUDI	**: STRATA SATU (S1)**
PROGRAM STUDI	**: TEKNIK INFIRMATIKA**

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM KUANTAN SINGINGI
2025

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama	: Muhammad Hambali
NPM	: 200210038
Tempat/Tgl Lahir	: Bangka, 05 Maret 2000
Alamat	: Seberang Taluk

Saya menyatakan bahwa dalam skripsi yang berjudul "**Aplikasi Deteksi Emosi Pada Teks Menggunakan *Natural Language Processing***" tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana komputer disuatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naska ini dan disebutkan dalam daftar pustaka. Atas pernyataan ini dibuat saya siap menanggung segala resiko dan sanksi apabila dikemudian hari ditemukan palanggaran terhadap etika keilmuan dalam skripsi saya ini.

TelukKuantan, 06 Februari 2026



Muhammad Hambali
NPM. 200210038

PERSETUJUAN SKRIPSI

NPM	:	200210038
Nama	:	Muhammad Hambali
Jenjang Studi	:	Strata Satu (S1)
Program Studi	:	Teknik Informatika
Judul Skripsi	:	Aplikasi Deteksi Emosi Pada Teks Menggunakan *Natural Language Processing*

Teluk Kuantan, 06 Februari 2026

Menyetujui,

Pembimbing I,

Harlanja, S.Pd., M. Kom Tanggal, 06 Februari 2026
NIDN. 1017057702

Pembimbing II,

Erlinda, M. Kom Tanggal, 06 Februari 2026
NIDN. 1006039301

Mengetahui,
Ketua Program Studi Teknik Informatika

Helpi Nopriandi, S. Kom., M. Kom Tanggal, 06 Februari 2026
NIDN. 1030118303

TANDA TANGAN PENGESAHAN SKRIPSI

NPM	:	200210038
Nama	:	Muhammad Hambali
Jenjang Studi	:	Strata Satu (S1)
Program Studi	:	Teknik Informatika
Judul Skripsi	:	Aplikasi Deteksi Emosi Pada Teks Menggunakan *Natural Language Processing*

Dipertahankan di depan Tim Penguji Skripsi

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Islam Kuantan Singingi

Pada Tanggal :

Dewan Penguji

No.	Nama	Jabatan	Tanda Tangan
1.	Agus Candra, ST, M.Si	Ketua	
2.	Harianja, S. Pd., M. Kom	Pembimbing I	
3.	Erlinda, M. Kom	Pembimbing II	
4.	Elgamar, Ph. D	Penguji I	
5.	Nofri Wandi Al-Hafiz, S.Kom., M.Kom	Penguji II	

Mengetahui,

Dekan,
Fakultas Teknik

AGUS CANDRA, ST, M.Si
NIDN. 1020088701

Ketua,
Prodi Teknik Informatika

HELPI NOPRIANDI, M.Kom
NIDN. 1030118303

DAFTAR ISI

Halaman

KATA PENGANTAR ..... **i**
DAFTAR ISI ..... **ii**
DAFTAR TABEL ..... **iii**
DAFTAR GAMBAR ..... **iv**

I. PENDAHULUAN
1.1 Latar Belakang..... 1
1.2 Identifikasi Masalah 7
1.3 Rumusan Masalah..... 8
1.4 Ruang Lingkup Penelitian 8
1.5 Tujuan dan Manfaat Penelitian..... 8
1.6 Sistematika Penelitian..... 9

II. TINJAUAN PUSTAKA
2.1 Pengertian *Phyton* 10
2.2 *Chatbot* 10
2.2.1 Pengertian Emo**si** 11
2.2.2 Klasifikasi Emosi 12
2.2.3 Natural Language Processing 13
2.2.4 Cakupan Natural Language Processing 16
2.2.5 Machine Translation 16
2.2.6 Attention 17
2.3 Identifikasi Emosi Pada Teks..... 19
2.4 Neural Network 20
2.5 Kajian Terdahulu 23

III. METODE PENELITIAN
3.1 Lokasi dan Waktu Penelitian 25

3.2 Kerangka Penelitian 25
3.3 Metode Penelitian 28
3.4 Teknik Mengumpulkan Data 29

IV. ANALISA DAN HASIL PERANCANGAN SISTEM

4.1 Analisis Sistem 31
4.2 Leksikon Emosi 32
4.3 Pengujian Kekayaan Leksikon Emosi 33
4.4 Problem Scoping 34
4.5 Data Acquisition 35
4.6 Data Exploration 36
4.7 Menghilangkan Punktuasi 36
4.8 Lemmatization (Lematisasi) 36
4.9 Menyortir Data Kelas Tags 36
4.10 Mencari Jumlah Keseluruhan *Data* Teks 36
4.11 Encoding Label Atau Tag 36
4.12 Modeling 37
4.13 Arsitektur Neural Network 38
4.14 Pelatihan Model 38
4.15 Evaluation 39
4.16 Metodologi Pengembangan Sistem 40
4.17 *Problem Scoping* 40
4.18 *Data Acquisition* 41
4.19 Data Exploration 41

V. HASIL DAN PEMBAHASAN

5.1 Pengujian Kekayaan Leksikon Emosi 43
5.2 Deteksi Emosi dengan Leksikon Emosi 45
5.2.1 Pengambilan data 45
5.2.2 Pengujian Dengan Leksikon Emosi 45
5.2.3 Validasi Hasil Pengujian 46

VI. PENUTUP

6.1 Kesimpulan 47

6.2 Saran 47

DAFTAR PUSTAKA 48

DAFTAR GAMBAR

Gambar 2.1	Model Emosi Ekman.....	13
Gambar 2.2	Contoh tata bahasa .....……..	16
Gambar 2.3	*Pharse tree* .....……	16
Gambar 2.4	*Feed-forward neural network* dan *Recurrent neural network* .....…….	22
Gambar 3.1	Tahapan penelitian .....……	26
Gambar 4.1	Lemmatization (Lematisasi) .....…	36
Gambar 4.2	Menyortir data kelas tags	37
Gambar 4.3	Mencari jumlah keseluruhan data teks.....	37
Gambar 4.4	Melakukan padding	37
Gambar 4.5	Modeling LSTM	38
Gambar 4.6	Arsitektur Neural Network	38
Gambar 4.7	Flowchart AI Project Cycle	40

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu ...…….……	23
Tabel 4.1	Contoh Emolex .....…….....……	31
Tabel 4.2	Contoh leksikon emosi penggabungan EmoLex dan sinonimnya.....……	32
Tabel 4.3	Pembagian kata dalam 8 emosi dan sentimen…..……	32
Tabel 4.4	Blackbox Testing .....…….……	50
Tabel 5.1	Contoh *update status* yang diperoleh ……...	45
Tabel 5.2	Emosi dan akumulasi bobotnya .....……....	46

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kehidupan manusia erat kaitannya dengan emosi. Emosi merupakan reaksi terhadap seseorang atau suatu kejadian yang sedang terjadi. Emosi memiliki peranan penting terhadap kehidupan sehari-hari. Emosi dibagi menjadi dua luaran yaitu positif dan negatif. Dari dua luaran tersebut, ada beberapa kategori di dalamnya seperti marah, senang, sedih, takut, dan terkejut. Analisis emosi sudah menjadi pembahasan dalam berbagai disiplin ilmu seperti kognitif, psikologi, bahkan media sosial[1].

Indentifikasi emosi sendiri terbagi menjadi dua bentuk yaitu verbal dan non-verbal. Emosi verbal merupakan emosi dalam bentuk lisan ataupun tulisan, sedangkan non-verbal adalah emosi yang menggunakan bahasa tubuh untuk mengekspresikannya, seperti raut wajah, tindakan, gerak-gerik tangan, maupun kaki. Mendeteksi emosi dalam dialog tekstual adalah masalah yang menantang karena tidak adanya ekspresi wajah dan modulasi suara [2]. Untuk mendeteksi emosi tersebut perlu dilakukan klasifikasi emosi. Klasifikasi sendiri merupakan proses pengelompokan beberapa data sesuai kriteria. Mendeteksi emosi memberikan peranan penting dalam berbagai aspek karena dapat diterapkan dalam berbagai bidang, seperti pengambilan keputusan dalam lingkungan sosial maupun lingkungan bisnis [3].

Dewasa ini, dengan berkembang pesatnya teknologi, komunikasi teks merupakan salah satu bentuk komunikasi yang paling sering kita temui. Dengan

adanya internet, manusia disuguhkan banyak sekali komunikasi dalam bentuk teks, mulai dari berita di media massa, materi-materi pelajaran sekolah, dan contoh yang paling mudah ditemui adalah aplikasi online chat. Penggunaan aplikasi online chat sudah hal yang lazim dipakai dimasa kini. Bahkan di Indonesia, berdasarkan hasil survey APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pada tahun 2017 terhadap 2.500 responden yang tersebar di seluruh Indonesia, jenis aplikasi layanan internet yang paling digemari adalah pesan instan atau chatting. Terdapat banyak sekali aplikasi online chat yang tersedia di internet, contohnya Facebook Messenger, WhatsApp, WeChat, Telegram, dll. Diantara aplikasi online chat yang lazim digunakan di Indonesia adalah Whatsapp dan Telegram, masing-masing memiliki 88,7% dan 62,8% per Februari 2022, dari rentang umur 16 - 64 tahun.

Meningkatnya preferensi untuk melakukan komunikasi berbasis teks pada aplikasi online chat menyebabkan jumlah interaksi sosial meningkat pesat, akan tetapi menurunkan kualitas interaksi itu sendiri yang bisa merusak hubungan [4]. Laporan-laporan dari para individu juga mengklaim bahwa komunikasi berbasis teks menghasilkan banyak miskomunikasi dibandingkan dengan komunikasi tatap muka [5]. Kombinasi antara ketergantungan yang besar dengan komunikasi berbasis teks ditambah seringnya terjadi miskomunikasi dapat merusak hubungan karena individu sering mengalami perasaan negatif ketika kesalahpahaman terjadi [6]. Wiliandari [7] dalam artikelnya yang berjudul “Aplikasi Kecerdasan Emosional dalam Perilaku Organisasi” menyebutkan bahwa masalah yang terbesar dalam komunikasi adalah kesalahpahaman yang terjadi karena orang gagal menyampaikan apa

yang mereka pikirkan dan yang dirasakan sehingga dapat menyebabkan adanya konflik.

Kebahagiaan merupakan sebuah konsep yang mengacu pada bentuk emosi positif yang dirasakan individu serta aktivitas positif yang disukai oleh seorang individu. Emosi positif diklasifikasikan menjadi tiga kategori yaitu emosi yang berhubungan dengan masa lalu, emosi yang berhubungan dengan masa kini dan emosi yang berhubungan dengan masa depan [8]. Natural Language Processing (NLP) mencari tahu penggunaan komputer untuk memproses atau memahami bahasa manusia dengan tujuan untuk melakukan tugas tertentu. NLP adalah bidang interdisipliner yang menggabungkan komputasi linguistik, ilmu kognitif dan kecerdasan buatan [9]. Dari perspektif ilmiah, NLP bertujuan unruk membuat model yang memiliki mekanisme kognitif dengan dasar fungsi untuk memahami dan mengetahui asal produksi dari sebuah bahasa manusia.

Dari perspektif teknik, NLP memiliki kaitan dengan bagaimana mengembangkan aplikasi praktis baru untuk memfasilitasi interaksi antara sebuah komputer dan bahasa manusia. NLP adalah bidang untuk memahami bahasa manusia menggunakan komputer. NLP melibatkan analisis dari data bahasa manusia dengan volume yang besar menggunakan komputer untuk mencari arti demi keperluan sebuah aplikasi [10]. Permasalahan yang dihadapi oleh masyarakat dewasa ini adalah tingginya tingkat orang yang membutuhkan bantuan tenaga profesional dalam bidang psikologi. Pada masa pandemi ini masyarakat dibatasi aksesnya untuk pergi ke layanan psikologi. Selain itu, meskipun animo masyarakat Indonesia akan pelayanan psikologi sudah tinggi,

tetapi jumlah orang yang sudah mendapatkan bantuan masih sedikit karena gangguan kesehatan mental pada masyarakat Indonesia masih merupakan hal yang tabu untuk dibicarakan, sebagian besar dari para penderita kesehatan mental memilih untuk tidak mengkonsultasikan kesehatan mereka ke pihak profesional.

Aplikasi ini berbasis mobile (Android) dengan sistem operasi minimal Android 5.0 Aplikasi ini akan terdiri 3 bagian menu utama untuk digunakan oleh user yang terdiri dari “Home Page”, ”Help Nearby”, ”About Us” dan dioperasikan menggunakan tombol layar dan navigasi pada handphone. Aplikasi ini menggunakan bahasa Inggris dan menggunakan REST API yang dibuat sendiri untuk databasenya dan Natural Language Processing sebagai Implementasi Machine Learning dan aplikasi ini bisa berjalan dengan Internet.

Mendeteksi emosi memberikan peranan penting dalam berbagai aspek karena dapat diterapkan dalam berbagai bidang, seperti pengambilan keputusan dalam lingkungan sosial maupun lingkungan bisnis [11]. Pada era pandemi seperti ini, segala aktivitas yang biasanya tatap muka beralih ke daring (dalam jaringan), mulai dari bidang politik, ekonomi, hingga pendidikan. Di dalam dunia pendidikan, tenaga pengajar biasanya kesulitan untuk menyampaikan materinya dikarenakan *mood* murid atau siswa yang berbeda-beda, dan hasilnya akan buruk. Dengan demikian analisis emosi menjadi sangat penting untuk dilakukan. Analisis emosi dapat digunakan untuk mengetahui keadaan emosi siswa. Dari emosi tersebut tenaga pengajar dapat mengetahui *mood* siswanya. Hal tersebut berguna untuk meningkatkan kualitas belajar mengajar.

Dalam penelitian tentang klasifikasi emosi pada teks, banyak peneliti menggunakan beberapa metode, seperti LSTM, BiLSTM, word2vec, SVM, dan CNN. Dari penelitian- penelitian tersebut dihasilkan akurasi dari 70% hingga 90%. Diantaranya adalah klasifikasi emosi pada teks menggunakan metode LSTM oleh Riza [12] penelitian tersebut mendapatkan nilai akurasi sebesaar 73,15%. Penelitian selanjutnya oleh Chiorrini et al, 2021 [13] dengan penelitian yang berjudul *emotion and sentiment analysis of tweet using BERT,* pada penelitian ini peneliti mendapatkan akurasi sebesar 92%. Hasil akurasi tersebut dipengaruhi oleh banyak data dan perbedaan label emosi yang dipakai. Pada penelitian ini, peneliti menggunakan sembilan label emosi diantaranya adalah marah, senang, sedih, takut, jijik, netral, tertarik, kaget, dan percaya.

Analisis sentimen melakukan deteksi emosi melalui teknologi kecerdasan buatan menggunakan NLP yang menggabungkan berbagai teknik linguistik dan komputasi untuk membantu computer memahami dan menghasilkan teks dan ucapan/suara dalam bentuk bahasa manusia. Deteksi emosi merupakan turunan yang diperluas dari analisis sentimen. Deteksi emosi menganalisis emosi- emosi yang lebih halus seperti kemarahan, kebahagiaan, kesedihan, kecemasan, depresi, dan lain-lain, dan menerapkan umpan balik ini untuk pengambilan keputusan di masa depan. Analisis sentimen menginterpretasikan pandangan dan perasaan untuk memberikan polaritas seperti positif, negatif, atau netral. Namun, deteksi emosi bertujuan untuk menemukan nuansa sentimen yang lebih spesifik seperti kebahagiaan, kesedihan, depresi, kecemasan, dan lainnya. Analisis emosi teks seperti ini membantu dalam memahami umpan balik yang diberikan melalui

media sosial online, ulasan pelanggan, ulasan produk, forum diskusi, sistem rekomendasi online, agen percakapan, email, blog web, dan lain-lain.

Dalam penerapan teknik NLP, salah satunya yang paling penting adalah analisis sentimen. Analisis sentimen memungkinkan sistem untuk memahami sikap, pendapat, dan emosi melalui teks serta membantu dalam memprediksi informasi tersembunyi yang ada dalam teks (Shindu & Vavidu, 2019) [14]. Singkatnya analisis sentimen memungkinkan sistem untuk mengevaluasi dan memahami sentimen dari sebuah teks. Analisis sentimen digunakan untuk menentukan status sebuah data (Prasasti, 2023) [15], seperti apakah data tersebut mengandung konotasi positif atau negatif. Pada perusahaan Kata.ai, teknologi NLP dapat digunakan untuk menyediakan solusi analisis sentimen yang berguna bagi proses bisnis. Dengan teknologi ini, suatu bisnis dapat dengan mudah memantau dan menganalisis sentimen misalnya *tweet*, postingan pada media sosial, hingga artikel berita, yang mana hal ini tentunya dapat membantu mereka memahami pendapat pelanggan mengenai produk maupun layanan mereka. Dampak positifnya akan sangat besar bagi perusahaan yang memahami pendapat pelanggan mengenai produk atau layanan yang dimilikinya. Akhirnya perusahaan dapat melakukan pengembangan produk atau layanan sesuai dengan selera pasar dengan mengidentifikasi tren serta pola perilaku konsumen yang terdapat pada sumber data yang telah dilakukan analisis sentimen.

Perlu dipahami bahwa analisis sentimen oleh kecerdasan buatan merupakan teknologi yang masih terus dikembangkan. Keakuratan informasi hasil analisis sentimen dapat dipengaruhi oleh berbagai macam faktor seperti, kualitas data, metode yang dipakai, serta kompleksitasnya. Namun, strategi

pemasaran digital dan analisis data berbasis kecerdasan buatan jauh lebih efisien dan akurat dibandingkan dengan kemampuan manusia. Hal ini memungkinkan kecerdasan buatan melakukan personalisasi pengalaman pengguna, meningkatkan kepuasan *customer experience*, dan meningkatkan penjualan. Melalui berbagai informasi yang telah dikumpulkan dan dianalisis, teknologi ini menargetkan iklan berdasarkan preferensi personal pengguna. Kecerdasan buatan menjadi penunjang yang sangat besar dalam pemasaran dan tidak menutup kemungkinan akan menjadi pusat dari kegiatan pemasaran digital di masa depan. Meskipun saat ini, kebanyakan kecerdasan buatan belum dapat berfungsi mandiri tanpa campur tangan manusia, bantuan kecerdasan buatan telah memudahkan pekerjaan dan secara signifikan mengurangi biaya, meningkatkan efisiensi, serta produktivitas pemasaran bisnis.

Berdasarkan pertimbangan diatas penulis tertarik untuk mengangkat permasalahan Aplikasi Deteksi Emosi ini dengan judul **"Aplikasi Deteksi Emosi Pada Teks Menggunakan *Natural Language Processing*"**

1.2. Identifikasi Masalah

1. Bagaimana menerapkan metode *Natural Language Processing* dapat meningkatkan pelayanan online di Pemda Kuantan Singingi.
2. Bagaimana membangun aplikasi dalam membantu kebijakan-kebijakan Pemda Kuantan Singingi dalam meningkatkan pelayanan secara online.

1.3. Rumusan Masalah

Berdasarkan uraian latar belakang diatas maka perumusan masalah pada penelitian ini adalah "Bagaimana membangun Aplikasi Deteksi Emosi Pada Teks Menggunakan Natural Language Processing" agar mempermudah pelayanan terhadap masyarakat secara online yang ada di lingkungan Kabupaten Kuantan Singingi.

1.4. Tujuan Penelitian

Tujuan penelitian ini yaitu untuk menganalisis emosi pada teks, terutama untuk melakukan klasifikasi emosi pada teks dengan menggunakan *Natural Language Processing*.

1. Menghasilkan analisis emosi dalam pelayanan masyarakat yang ada di Kabupaten Kuantan Singingi guna meningkatkan pelayanan.
2. Menghasilkan analisis emosi yang dapat menyebarkan informasi terkait pelayanan ada di Kabupaten Kuantan Singingi

1.5. Manfaat Penelitian

1. Memberikan kemudahan terhadap Pemda Kabupaten Kuantan Singingi dalam meningkatkan pelayanan online kepada masyarakat
2. Memberikan pengetahuan yang lebih baik terhadap masyarakat dengan mudahnya menemukan informasi bidang *Natural Language Processing* khususnya dalam tugas klasifikasi emosi yang ada di Kabupaten Kuantan Singingi

1.6. Ruang Lingkup Penelitian

Agar tidak menyimpang dari topik permasalahan maka perlu adanya batasan masalah, sebagai berikut:

1. Sistem ini dibangun untuk memberikan kemudahan terhadap pelayanan masyarakat dalam melakukan pengaduan sehingga masyarakat bisa langsung melakukan pengaduan secara online yang ada di Kuantan Singingi.
2. Sistem yang akan diusulkan nanti yaitu sebuah perencanaan Aplikasi Deteksi Emosi Pada Teks menggunakan *Natural Language Processing*.

1.7. Sistematika Penelitian

Supaya penelitian yang di usulkan ini lebih mudah untuk dipahami, maka dari itu penulis mengklasifikasikan penelitian ini kedalam enam bab, dimana bab yang satu dengan yang lain saling berhubungan, untuk lebih jelasnya lihat pada penjelasan bab-bab sebagai berikut:

BAB I : PENDAHULUAN

Bab ini akan membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian dan manfaat penelitian.

BAB II : TINJAUAN PUSTAKA

Bab ini akan menguraikan berbagai konsep dasar dan penerapan Metode yang berhubungan dengan penelitian ini, serta hal-hal yang diperlukan pada proses analisis masalah.

BAB III : METODE PENELITIAN

Bab ini akan menerangkan kerangka kerja penelitian, ruang lingkup permasalahan, analisa, pengumpulan data, dan menjelaskan metode penelitian.

BAB IV : JADWAL KERJA

Pada bab ini akan dijelaskan mengenai tahapan kerja hasil perancangan yang telah dijabarkan di Bab III.

BAB II

TINJAUAN PUSTAKA

2.1. Pengertian Emosi

Dalam Bahasa Latin, asal kata emosi adalah emovere, yang mana merupakan kata kerja yang berarti “menggerakkan, bergerak”, ditambah awalan “e-” untuk memberi arti “bergerak menjauh”, menyiratkan bahwa kecenderungan bertindak merupakan hal yang mutlak dalam emosi. Terdapat berbagai macam definisi pada emosi, seperti keadaan bergejolak, gangguan keseimbangan, respon kuat dan tidak beraturan terhadap stimulus [16]. Menurut Goleman, emosi adalah dorongan untuk bertindak, rencana seketika untuk mengatasi masalah yang telah ditanamkan secara berangsur-angsur oleh evolusi. Chaplin dalam bukunya yang berjudul Dictionary of Psychology menyebutkan bahwa emosi adalah sebagai suatu keadaan terangsang dari organisme mencakup perubahan-perubahan yang disadari, yang mendalam sifatnya dari perubahan perilaku.

Dari beberapa definisi diatas, dapat disimpulkan bahwa emosi adalah suatu respon berpola individu yang kompleks terhadap suatu keadaan, baik eksternal maupun internal, yang mencakup perubahan-perubahan disadari, yang juga ditunjukkan dari berbagai reaksi psikologis, yang bisa bermakna negatif atau positif bagi individu.

Sistem survei analisis emosi dilakukan pertama kali pada tahun 1966, ini bisa dikatakan fase awal dari analisis emosi pada teks. Robert Plutchik membagi emosi menjadi delapan kategori utama. Separuhnya emosi positif dan

separuhnya lagi emosi negatif. Setiap emosi memiliki subkelompok. Sebagai contoh pada emosi bahagia dapat menggambarkan keadaan kegembiraan atau kesenangan tergantung pada koteks yang diberikan. Robert Plutchik juga membuat roda emosi sebagai ilustrasi hubungan emosi satu dengan lainnya.

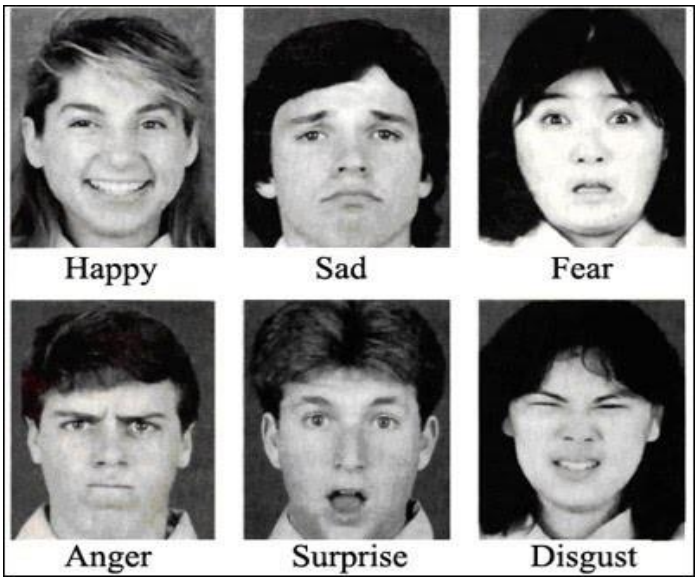
2.1.1. Klasifikasi Emosi

Salah satu bidang yang menjadi fokus dalam studi ilmu computational linguistic adalah klasifikasi emosi. Klasifikasi emosi mempelajari sentimen, opini, dan emosi yang ada pada teks. Hal utama yang menjadi tujuan dari klasifikasi emosi adalah untuk mengekstrak, memproses, merangkum, dan menganalisa informasi yang terdapat pada teks menggunakan berbagai metode sehingga dapat mendeteksi niat, sudut pandang, dan emosi yang diberikan oleh penulis teks.

Pada umumnya, klasifikasi emosi dapat dibagi menjadi tiga level: tingkat dokumen (document level), tingkat kalimat (sentence level), dan tingkat berbutir halus (fine-grained level). Sentence-level dan document-level juga dapat dikategorikan ke dalam coarse-grained level. Metode dalam klasifikasi emosi terbagi menjadi dua jenis, yaitu lexical-based dan learning-based. Metode learning-based menggunakan data training dan data testing, sedangkan lexical-based menggunakan kamus (opinion lexicon).

Model kategorikal yang paling populer dan banyak digunakan adalah model emosi Ekman, dimana membagi emosi dasar manusia menjadi: marah, menjijikkan, ketakutan, kebahagiaan, kesedihan, dan terkejut. Emosi-emosi ini dikarakteristikkan sebagai universal, sebagaimana mereka diekspresikan

dalam cara yang sama tanpa memandang perbedaan budaya dan era. Gambar 2.1 menunjukkan jenis-jenis emosi pada model emosi Ekman.



Gambar 2.1: Model Emosi Ekman

2.2. Natural Language Processing

Natural language processing (NLP) merupakan subbidang ilmu komputer yang menggunakan teknik komputasi untuk belajar, memahami, dan menghasilkan konten bahasa manusia. NLP memungkinkan pengguna untuk berkomunikasi dengan komputer secara alami. Proses memahami bahasa alami dapat diuraikan menjadi analisis sintaksis dan semantik. Sintaksis mengacu pada pengaturan kata-kata dalam kalimat sedemikian rupa sehingga masuk akal secara tata bahasa. Selain itu, analisis sintaksis mengubah urutan kata menjadi struktur yang menunjukkan bagaimana kata-kata ini terkait satu sama lain. Di sisi lain, semantik mengacu pada arti setiap kata dan kalimat. Analisis semantik dari konten bahasa alami menangkap makna sebenarnya, itu memproses struktur

kalimat logis untuk menemukan kesamaan antara kata-kata dan memahami topik yang dibahas dalam kalimat [17].

Natural language processing, atau disingkat NLP, adalah suatu cabang ilmu dari ilmu komputer, atau lebih spesifiknya NLP merupakan suatu cabang ilmu dari artificial intelligence (AI). NLP fokus pada bagaimana memberi komputer kemampuan untuk memahami teks dan ucapan sebagaimana kemampuan yang dimiliki manusia. NLP mengombinasikan ilmu computational linguistic dengan ilmu lainnya seperti statisika, machine learning, dan deep learning. Internasional Business Machines Corporation menyebutkan beberapa tugas yang termasuk fokus dalam studi NLP: speech recognition, part of speech tagging, word sense disambiguation, named entity recognition, co-reference resolution, sentiment analysis, dan natural language generation.

Natural language processing [18] atau pengolahan bahasa alami adalah salah satu cabang ilmu yang meneliti dan mengembangkan cara kerja komputer untuk memahami dan memproses bahasa alami sebagai kalimat atau ucapan. Pada natural language processing (NLP) digunakan untuk mengevaluasi bagaimana bahasa dalam teks sesuai dengan aturan tata bahasa. Dalam penerapannya ada tahapan yang dipakai untuk menganalisa bahasa yaitu:

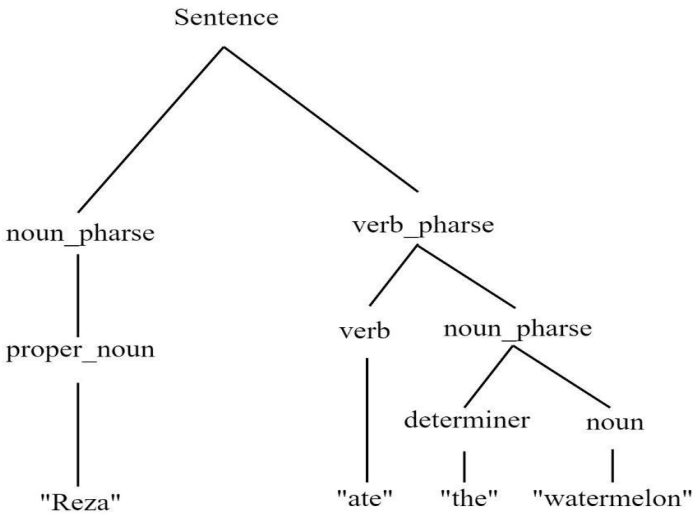
a. Stemming, proses pemotongan kata diawal atau diakhir untuk mendapatkan kata baku sesuai dengan aturan tata bahasa. Tujuan dari stemming ini adalah mengilangkan imbuhan pada kata. Contohnya: perumahan menjadi rumah.

b. Tokenization, proses pembagian sebagian kalimat menjadi kata per kata. “Rumah ini sebentar lagi akan ditinggalkan”. Setiap kata dalam kalimat tersebut akan dipisah satu per satu. “Rumah”, “ini”, “sebentar”, “lagi”, “akan”, “ditinggalkan”. Maka kalimat yang sudah dipisah perkata disebut token.

c. Parsing, proses menentukan struktur pada teks dengan menganalisa kata penyusnnya berdasarkan tata hasa yang mendasarinya. Contoh pengaplikasiannya pada gambar 2.1. Gambar 2.1 merupakan contoh dari proses penguraian kata, setelah proses tersebut akan menjadi sebuah pharse tree seperti pada gambar 2.2 dimana kalimat akan diurai kata per sesuai tata bahasanya pada sebuah pharse tree.

“Reza ate the watermelon”

Sentence > noun_pharse, verb_pharse
Noun_pharse > proper_noun
Noun_pharse > determiner, noun
Verb_pharse > verb, noun_pharse
Proper_noun > [Reza]
Noun> [watermelon]
Verb > [ate]
Determiner > [the]



Gambar 2.3 *Pharse tree*

2.2.1 Cakupan Natural Language Processing

NLP adalah salah satu cabang dari AI yang memiliki beberapa hubungan yang signifikan di antaranya mesin komputer dan interaksi orang. Dalam dunia digital ini, NLP memainkan peran penting yang menjembatani kesenjangan komunikasi antara manusia dan dunia digital. Lebih lanjut, bagian ini memberikan ruang lingkup dan aplikasi NLP yang digunakan di masa mendatang, berikut ini cakupan NLP.

2.3. Pengertian *Python*

Python adalah sebuah bahasa pemrograman berbasis objek yang dapat diinteraksi secara interaktif. Bahasa ini memiliki struktur data tingkat tinggi. *Python* merupakan sebuah bahasa pemrograman interpretatif yang memiliki banyak fungsi, dan didesain dengan fokus pada kejelasan dan kemudahan pemahaman kode. *Python* dianggap sebagai bahasa yang menggabungkan kemampuan dan kejelasan sintaks kode. Bahasa pemrograman Pythondirancang

khusus untuk memudahkan programmerdalam membuat program dengan efisiensi waktu, kemudahan pengembangan, dan kompatibilitas dengan sistem. Pythondapat digunakan untuk membuat aplikasi mandiri atau pemrograman skrip.

2.4. Chatbot

Chatbot atau *chatterbot* adalah program komputer yang mampu menyimulasikan dan memproses percakapan manusia (tertulis atau lisan). *Chatbot* memungkinkan manusia untuk berinteraksi dengan perangkat digital seperti layaknya mereka berbicara dengan manusia nyata. Bentuk *chatbot* dapat berupa program yang mampu merespon satu baris kalimat sederhana pada sebuah pertanyaan, atau bisa juga berbentuk asisten digital yang belajar dan berevolusi untuk memberikan tingkat personalisasi tingkat tinggi seiring *chatbot* mengumpulkan dan memproses informasi. *Chatbot* dengan kata lain merupakan komputer atau robot yang dirancang untuk berinteraksi atau bercakap-cakap denganmanusia [19].

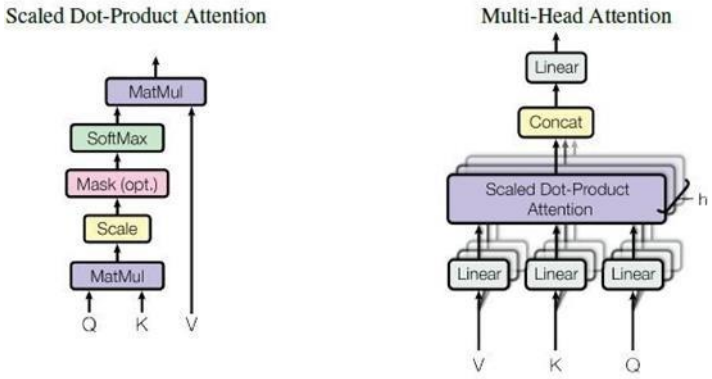
Chatbot dapat diimplementasikan untuk bidang komersial, pendidikan, *entertainment* dan sektor pelayanan publik [20]. Beberapa contoh pengimplementasian dari *chatbot* adalah asisten virtual seperti Amazon Alexa, Google Assistant, atau aplikasi perpesanan seperti WeChat, dan Facebook Messenger. *Chatbot* akan bekerja jika diberi masukan teks yang nantinya akan diproses oleh suatu *software* yang memiliki fungsi untuk klasifikasi maksud percakapan. Contoh kalimat seperti “Selamat malam!” mungkin akan direspon oleh *chatbot* dengan “Selamat malam!” atau “Selamat malam juga untuk anda”.

2.4.1. Machine Translation

Seiring dengan meningkatnya informasi global secara online, tantangan utama ialah menyediakan akses informasi kepada setiap orang dengan terjemahan bahasa yang dapat dimengerti. Google adalah perusahaan inovatif yang memiliki terjemahan mesin yang menerjemahkan informasi berdasarkan informasi statistik yang dimiliki Tantangan teknik penerjemahan mesin bukanlah menerjemahkan kata-kata tetapi memahami dan menyimpan makna kalimat untuk menawarkan terjemahan yang sebenarnya. Google adalah perusahaan inovatif yang memiliki terjemahan mesin yang menerjemahkan informasi berdasarkan informasi statistik yang dimiliki. Tantangan teknik penerjemahan mesin bukanlah menerjemahkan kata-kata tetapi memahami dan menyimpan makna kalimat untuk menawarkan terjemahan yang sebenarnya.

2.4.2. Attention

Attention dapat dideskripsikan sebagai sebuah fungsi yang memetakan sebuah *query* dan sebuah himpunan yang terdiri dari pasangan *key-value* ke *output*. *Attention* memiliki mekanisme yang digunakan pada arsitektur *transformer* yaitu *multi-head attention* yang terdiri dari mekanisme- mekanisme yang lebih sederhana yaitu *scaled dot-product attention*. Tipe mekanisme *attention* lain yang juga digunakan adalah *self-attention*.



Cara kerja dari mekanisme *scaled dot-product attention* yaitu skalar antara vektor *key* dan *query* dikalikan yang masing-masing memiliki dimensi dk. Hasil dari perkalian tersebut akan dibagi dengan √dk, yang nantinya akan dimasukkan ke sebuah fungsi *softmax* dan akan digunakan sebagai bobot dari sebuah matriks *value*.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{.....(2.1)}$$

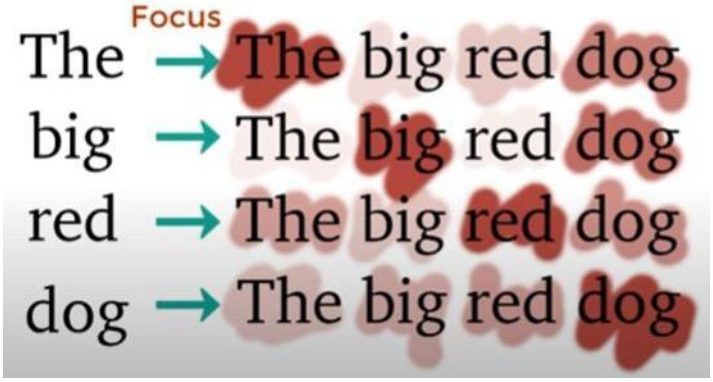
Multi-head attention merupakan *attention* yang dilakukan sebanyak h kali. Hal tersebut akan menghasilkan sejumlah h elemen dari perkalian matriks *key* dan *query*. Perkalian dengan matriks *value* sebanyak h akan dilakukan sebagai hasil dari perkalian matriks *key* dan *query* sejumlah h elemen di tahap sebelumnya. Selanjutnya, hasil dari perkalian matriks *value* sebanyak h tersebut akan menghasilkan sebanyak h vektor yang akan diteruskan kesebuah *layer concat* dan *dense* untuk menghasilkan sebuah vektor *output*. *Multi-head attention* menghasilkan keluaran vektor *output* yang bersifat lebih kontekstual dibandingkan keluaran vektor *output* yang dihasilkan mekanisme *attention* yang hanya dilakukansekali.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head1}, \ldots, \text{headh})\, W^o \quad \text{·····(2.2)}$$

$$\text{dimana head}i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Self-attention, atau bisa disebut *intra-attention*, adalah sebuah mekanisme untuk mengetahui *contextual relationships* antar kata yang ada di sebuah kalimat

maupun paragraf. *Self-attention* memungkinkan model untuk mengetahui input mana yang harus difokuskan.



2.5. Identifikasi Emosi Pada Teks

Emosi merupakan salah satu aspek dengan pengaruh besar terhadap sikap manusia dengan lingkungan. Emosi sendiri terbagi menjadi dua yaitu emosi positif dan negatif. Untuk mengidentifikasinya perlu mengetahui beberapa aspek pemicunya. Adanya media sosial diera digital ini, membuat seseorang sering berkomunikasi di dunia maya. Identifikasi emosi dilakukan dengan mengetahui pemicu yang muncul dan mengetahui respon seseorang tersebut. Emosi sendiri dibagi menjadi dua yaitu verbal dan nonverbal, emosi verbal dapat diartikan emosi langsung atau menggunakan lisan maupun tulisan sebagai medianya. Sedangkan nonverbal biasanya menggunakan bahasa tubuh untuk mengekspresikannya.

Emosi verbal khususnya pada tulisan atau teks perlu dilakukan identifikasi lebih lanjut untuk mengetahui emosi yang benar-benar terjadi agar tidak terjadi kesalahan identifikasi karena pada tulisan bisa bermaksud multimakna di dalamnya. Teks sendiri adalah media utama dalam komunikasi pada sosial media, contoh media sosial yang hampir semua proses komunikasinya menggunakan teks

adalah *twitter*. *Twitter* sendiri selain menjadi media komunikasi, pengguna juga dapat membagikan informasi, opini, curahan hati, dan berdiskusi. Identifikasi emosi pada teks dilakukan dengan menggunakan pendekatan klasifikasi emosi [21].

2.6. Klasifikasi Teks

Klasifikasi teks adalah proses pengelompokan suatu data ataupun dokumen ke dalam suatu kategori yang telah dibuat dan didefinisikan ke dalam kategori kelas. Tujuan dari klasifikasi teks adah untuk menganalisa, memproses, dan mengekstrak suatu informasi yang terkandung dalam teks. Dalam klasifikasi emosi pada teks, suatu teks akan diambil sebuah informasi yang terkandung di dalamnya untuk mengetahui emosi yang terkandung dalam teks opini tersebut. Pada klasifiksi teks, terdapat beberapa kategori seperti analisis sentiment, kategorisasi berita, dan klasifikasi topik [22].

Teks-teks opini cenderung berbentuk data yang tidak terstruktur, maka perlu dilakukan tahap *preprocessing* agar data teks tersebut lebih terstruktur dan dapat diserap informasi secara mudah. *Preprocessing* teks bertujuan untuk memecah setiap dokumen menjadi kata-kata individu dengan mewakili masing-masing sebagai vektor fitur. Untuk tujuan pengindeksan dokumen, proses pemilihan fitur dan fase preprocessing teks utama harus digunakan untuk memilih kata kunci. Tahap preprocessing teks, di sisi lain, membagi dokumen teks input menjadi fitur yang dikenal sebagai (*tokenization, words, terms* atau *attributes*).

2.7. Neural Network

Neural network adalah cabang penelitian yang meniru bagaimana otak

manusia memproses informasi dan menghasilkan hasil. Dengan meniru prinsip pembelajaran yang terinspirasi dari bagaimana sistem saraf manusia dan organisme biologis lainnya berfungsi. *Neural network*, juga dikenal sebagai jaringan saraf tiruan, adalah salah satu teknik pembelajaran mesin yang paling terkenal. *Neuron* adalah sel khusus yang menyusun sistem saraf. Akson dan dendrit berfungsi sebagai mekanisme penghubung antara *neuron-neuron*. Hubungan antara akson dan dendrit disebut sinapsis [23].

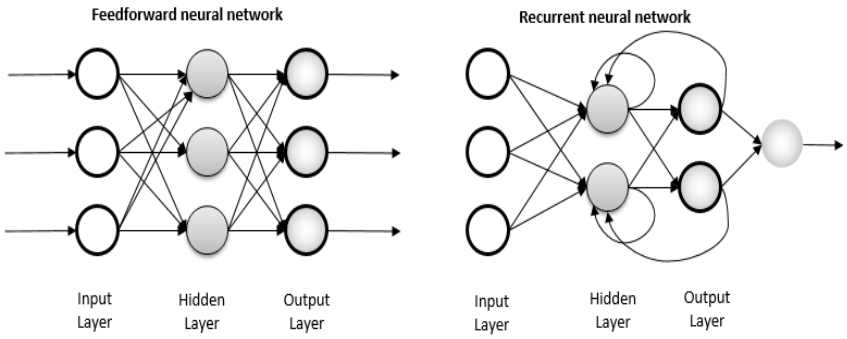
Arsitektur jaringan yang menghubungkan tingkat yang berbeda, disebut sebagai arsitektur jaringan. Lapisan antara lapisan input dan output dikenal sebagai *hidden layer*, sedangkan lapisan di atas lapisan output dikenal sebagai *hidden units*. Unit-unit ini tidak dapat langsung dilihat sebagai *input* atau *output* dari luar, oleh karena itu disebut dengan istilah "*hidden*". Lapisan tersembunyi yang terdiri dari unit tersembunyi, yang masing-masing merupakan unit saraf, merupakan inti dari jaringan saraf. Ini menerapkan non-linearitas ke jumlah tertimbang dari input di lapisan ini. Setiap unit lapisan berkomunikasi satu sama lain dengan menggunakan semua unit di lapisan bawah sebagai *input* dan *output*, serta koneksi antara setiap pasangan unit dalam dua lapisan yang berdekatan. Unit *input* ditambahkan bersama oleh setiap unit tersembunyi.

Secara umum terdapat dua jenis arsitektur neural network, yaitu feed-forward network dan recurrent/recursive network.

1. *Feed-forward Networ*k atau *Multi-Layer Perceptrons* (MLP) adalah jaringan di mana komponen tidak digilir dan *output* dikirim kembali ke tingkat bawah. Hal ini memungkinkan jaringan untuk beroperasi dengan *input* ukuran tetap

atau *input* panjang variabel dari serangkaian bagian yang dapat diabaikan. Jaringan belajar untuk mengintegrasikan komponen *input* saat mereka dimasukkan ke dalamnya. Data hanya mengalir dari *input* ke *output* dalam satu arah. Jaringan semacam ini digunakan untuk pengenalan pola karena lebih sederhana. Pengenalan gambar sering menggunakan jenis khusus jaringan umpan maju yang disebut *Convolutional Neural Network (CNN* atau *ConvNet).*

2. Dalam situasi di mana input berurutan. *Recurrent Neural Network (RNN)* sering digunakan ketika jaringan sedang memproses teks atau suara,. Urutan objek dimasukkan ke dalam RNN, dan menghasilkan vektor dengan ukuran tertentu yang berisi urutan tersebut. RNN memungkinkan *loop* dan memungkinkan data mengalir di kedua arah melalui jaringan. Jaringan ini lebih canggih dan kuat dari CNN. Untuk melihat prosesnya dapat dilihat pada gambar 2.3.



Gambar 2.4 *Feed-forward neural network* dan *Recurrent neural network*

Sumber: Pekel et al (2017)

Fungsi matematika yang disebut fungsi aktivasi, yang memetakan output dari satu lapisan ke input lapisan berikutnya, muncul setelah setiap lapisan dalam

tumpukan. Fungsi softmax adalah komponen aktivasi. Fungsi aktivasi softmax memastikan bahwa vektor keluaran dijumlahkan tepat satu kali atau 1. Kemudian, simpul keluaran dengan kemungkinan tertinggi dipilih sebagai label prediktif untuk frasa masukan [24]. Jaringan yang membutuhkan keluaran dengan label tertentu dapat menggunakan *activation function*. Misalnya, dalam jaringan yang telah dilatih untuk mengenali anjing dan ayam. Jaringan mengenali ayam dengan kepastian 75 persen jika vektor input untuk identifikasi ayam memiliki vektor output 0,75. Softmax hanya berfungsi ketika hanya ada satu solusi. Output node dengan probabilitas tertinggi kemudian dipilih sebagai label prediksi untuk kalimat yang menjadi input [25].

2.5. Kajian Terdahulu

Adapun beberapa penelitian terdahulu tentang Natural language processing (NLP) dapat dilihat pada tabel berikut ini:

Tabel 2.1
Penelitian Terdahulu

Nama	Judul	Metode	Hasil penelitian
Ayanouz dkk(2020)	*A Smart Chatbot Architecture based NLPand Machine Learning for Health Care Assistance*	Dalam penelitian ini penulis melakukan survei terperinci tentang literatur. Peneliti memeriksa banyak publikasi dari lima terakhir tahun, yang berkaitan dengan *Chatbot*.	Hasil penelitian ini menunjukkan *Chatbot* ini ramah pengguna dan *Chatbot* ini dengan sangat mudah. Khususnya di bidang kedokteran, *Chatbot* berbasis medis

Muktafin, E. H., Pramono, P., & Kusrini, K. (2021)	*Sentiments analysis of customer satisfaction inpublic services using K-nearest neighbors algorithm and natural language processing approach*	Penelitian ini menggunakan *K-nearest neighbors* (KNN) dan term *frequency-inverse documentfrequency* (TF-IDF) (TF-IDF) algoritma dengan pendekatan *natural language processing* (NLP) untuk mengklasifikasikan percakapan menjadi 2 kelas, "puas" dan "tidak puas".	Hasil penelitian ini mendapatkan akurasi 74,00%, presisi 76,00% dan recall 73,08%. Dalam percakapan dengan label "puas" menunjukkan pelanggan puas dengan pelayanan dan pemenuhankeinginan pelanggan, saat bercakap-cakap denganlabel"tidak puas" pelanggan kurang puas dengan waktu tunggu.
Rohman, A. N., Utami, E., & Raharjo, S. (2019)	Deteksi Emosi MediaSosial Menggunaka nPendekatan Leksikon dan *Natural language processing*	Penelitian ini menggunakan EmoLex sebagai leksikon yang digunakan untuk mendeteksi emosi pada suatu teks. Kosa kata pada EmoLex diperluas dengan pencarian sinonim menggunakan Kateglo API. EmoLex digunakan sebagai leksikon 8 kategori emosi Plutchik dan sentimen.	Pengujian menunjukkan bahwa leksikon emosi mampu mendeteksi 55.45% atau 15.357 dari 27.696 kata yang diperoleh dari update status pengguna Facebook. Selanjutnya update status tersebut diperbaiki menggunakan *Natural language processing* (NLP). Hasil perbaikannya dinilai dengan leksikon emosi yang telah dibuat sebelumnya. 26 dari 100 update status dapat

BAB III

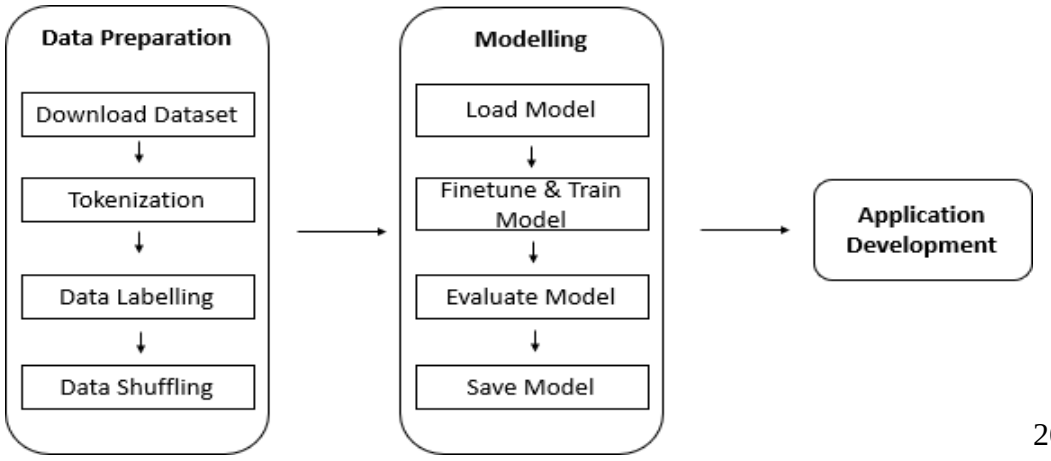
METODELOGI PENELITIAN

3.1. Uraian Tempat Penelitian

Tempat penelitian ini dilaksanakan di Kabupaten Kuantan Singingi, Kabupaten Kuantan Singingi (Kuansing) adalah sebuah kabupaten di Provinsi Riau, Indonesia, dengan ibu kota Teluk Kuantan. Kuansing merupakan pemekaran dari Kabupaten Indragiri Hulu dan terletak di bagian barat daya Provinsi Riau. Kabupaten ini memiliki luas wilayah 7.656,03 km² dan berbatasan dengan beberapa kabupaten/provinsi lain di sekitarnya.

3.2. Kerangka Penelitian

Kerangka penelitian adalah model struktur yang digambarkan dalam bentuk alur penelitian, mulai dari mengidentifikasi masalah hingga menyusun laporan akhir. Kemudian akan menghasilkan suatu sistem yang bisa mengatasi masalah yang diangkat pada awal penelitian. Tahapan-tahapan yang ada pada kerangka penelitian dibuat dalam bentuk diagram alur agar dalam penyelesaian penelitian tertata dengan baik dan menghasilkan output sistem yang dapat mengatasi permasalahan yang diangkat pada penelitian ini.



Gambar 3.1 Tahapan penelitian

3.3. Metode Penelitian

Metode waterfall adalah pengerjaan dari suatu sistem dilakukan secara berurutan. Jadi setiap tahap harus diselesaikan terlebih dahulu secara penuh atau detail sebelum diteruskan ke tahap berikutnya untuk menghindari terjadinya pengulangan tahapan.

Metode *Waterfall* ini dipilih karena pada metode ini tahapan ini analisis kebutuhan perangkat lunak harus didefinisikan dengan baik agar menghasilkan rancangan dan implementasi yang baik pula. Pada tahap analisi, penelitian harus benar-benar memahami tentang rancangan Sistem Informasi Manajemen Presensi kepuasan pelanggan, untuk selanjutnya menentukan fungsional sistem. Adapun langkah pengembangan sistem Waterfall yang digunakan.

1. Analisis kebutuhan perangkat lunak. Tahap ini merupakan proses pengumpulan kebutuhan yang dilakukan secara intensif untuk menspesifikasikan kebutuhan perangkat lunak agar dapat dipahami sesuai seperti yang dibutuhkan oleh pengguna. Spesifikasi kebutuhan perangkat lunak pada tahap ini perlu untuk didokumentasikan.
2. Desain. Tahap ini merupakan proses yang fokus pada desain pembuatan perangka lunak termasuk perangkat lunak, representasi struktur data, arsitektur perangkat lunak, representasi antar muka dan prosedur pengkodean.
3. Pembuatan kode program. Hasil tahap ini adalah program komputer sesuai dengan desain yang telah dibuat pada tahap sebelumnya.

4. Pengujian kode program. Hasil tahap ini adalah program komputer sesuai desai yang telah dibuat pada tahap sebelumnya.
5. Pengujian. Pengujian fokus pada perangkat lunak dari segi logik dan fungsional dan memastikan bahwa semua bagian sudah diuji. Hal ini dilakukan untuk meminimalisis kesalhan dan memastikan keluaran yang dihasilkan sesuai yang diinginkan.

3.4. Teknik Mengumpulkan Data

Tahapan ini dilakukan dengan cara studi literatur dan pustaka. Referensi-referensi yang berkaitan dengan penelitian dicari secara daring melalui internet. Literatur-literatur relevan yang didapatkan berupa jurnal ilmiah, skripsi, serta artikel di internet yang kemudian dikaji untuk menambah wilayah keilmuan pada penelitian ini serta menjadi pelengkap dan kebaruan dari penelitian yang sudah ada.

No	Judul	Penulis	Chatbot	IndoBERT	Klasifikasi Emosi	Aplikasi *Online Chat*

-	Pengembangan Chatbot untuk Aplikasi Online Chat dengan Pendekatan Klasifikasi Emosi pada Teks Menggunakan Metode IndoBERT Lite	Faishal Rayyan	✓	✓	✓	✓
1	IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding	Willie et al.	-	✓	✓	-
2	Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Support Vector Machine	Abirawa et al.	-	-	✓	-

3	Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Naïve Bayes Classifier	Noviardini et al.	-	-	✓	-
4	Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers	Putri	-	-	-	-
5	Web-based Application for Detecting Indonesian Clickbait Headlines using IndoBERT	Fakhruzzaman & Gunawan	-	✓	-	-

6	Emotion Classification on Indonesian Twitter Dataset	Saputri et al.	-	-	✓	-
7	A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation	Oh et al.	✓	-	✓	-

3.5. Data Preparation

Sesuai namanya, dalam tahap ini data dipersiapkan sebelum masuk ke tahap selanjutnya seperti *training*, *testing*, dan *production*. *Data preparation* adalah tahapan dimana data mentah dibersihkan untuk keperluan analisis. Keuntungan yang didapat setelah melakukan tahap ini adalah untuk mengatasi galat, dan menghasilkan data yang berkualitas yang siap dipakai.

3.5.1. Download Dataset

Penelitian ini menggunakan *dataset* publik milik Mei Silviana Saputri dkk. [26]. *Dataset* ini bisa diunduh di akun GitHub Mei Silviana Saputri. Namun,

penulis mengunduh *dataset* tersebut dari repositori GitHub bernama indonlu. Alasan penulis melakukannya karena *dataset* pada repositori indonlu sudah dibagi menjadi 3 set, yaitu *training*, *validation*, dan *test*. *Training*, *validation*, dan *test* set tersebut masing-masing memiliki 3.523, 440, dan 440 baris. Total ukuran *dataset* adalah 822 KB dengan tipe *file* berupa *comma-separated values* (CSV).

3.5.2. Tokenization

Tokenization adalah proses untuk memecah teks menjadi potongan kata-kata atau token-token. Pada tahap ini, penulis menggunakan pustaka Transformers untuk mengunduh *tokenizer* IndoBERT-liteLARGE fase pertama yang sudah dilatih. Contoh proses tokenisasi pada teks dapat dilihat pada gambar beriku.



3.5.3. Data Labelling

Metode *supervised learning* dengan *task* seperti klasifikasi emosi, membutuhkan *dataset* yang sudah diberi label. Hal tersebut perlu dilakukan karena dalam *supervised learning*, algoritma belajar dari *training dataset* dengan cara memprediksi terhadap data (label) dan memperbaiki hasil prediksi tersebut secara iteratif (Delua, 2021). Sehingga, algoritma dapat memahami, melihat, dan mengerti bagaimana komentar yang memiliki emosi.

Dalam penelitian ini, *dataset* yang akan dipakai sudah diberi label untuk mengkatogerikan teks ke dalam kelas-kelas. Kelas-kelas tersebut berupa emosi yang akan dipahami oleh model, yaitu cinta (*love*), gembira (*happy*), marah (*anger*), sedih (*sadness*), dan takut (*fear*). Nantinya teks dengan emosi sedih akan diberi nilai 0, emosi marah akan diberi nilai 1, emosi cinta akan diberi nilai 2, emosi takut akan diberi nilai 3, dan emosi senang akan diberi nilai 4.

3.5.4. Data Shuffling

Data Shuffling (mengacak data) umum dilakukan sebelum melakukan *training* pada model *machine learning*. Hal ini dilakukan untuk meminimalisasi varians dan memastikan model akan mempelajari represntasi data dengan baik. Dalam penelitian ini, penulis akan melakukan *data shuffling* pada *validation set*, dan *test set* supaya model bisa menggeneralisasi data dengan baik dan mengurangi *overfitting* pada model.

BAB IV

ANALISA DAN HASIL PERANCANGAN SISTEM

4.1. Analisis Sistem

Analisa kebutuhan implementasi diperlukan untuk mengetahui perangkat yang digunakan untuk penelitian. Perangkat yang digunakan meliputi perangkat lunak dan perangkat keras. Spesifikasi perangkat lunak dan perangkat keras yang dipakai untuk penelitian ini adalah sebagai berikut [27].

Tabel 4.1 Spesifikasi Perangkat Keras

Perangkat	Lenovo B40
Prosesor	Intel(R) Core(TM) i3-4005U CPU @ 1.70GHz 1.70 GHz
Hard Disk Drive	238 GB
RAM	4 GB

Terdapat spesifikasi perangkat keras penunjang menggunakan *cloud service* gratis yang disediakan oleh Google yaitu Google Colaboratory. Hal tersebut perlu dilakukan karena penelitian ini menggunakan model *pre-trained* yang cukup besar.

Tabel 4.2 Spesifikasi Perangkat Keras Berbasis Cloud

Perangkat	Colaboratory: Python 3 Google Compute Engine
Prosesor	Intel(R) Xeon(R) CPU @ 2.20GHz
Memori	12.23 GB
GPU	NVIDIA Tesla T4
VRAM	13.62 GB

Sedangkan untuk spesifikasi perangkat lunak yang digunakan adalah sebagai berikut:

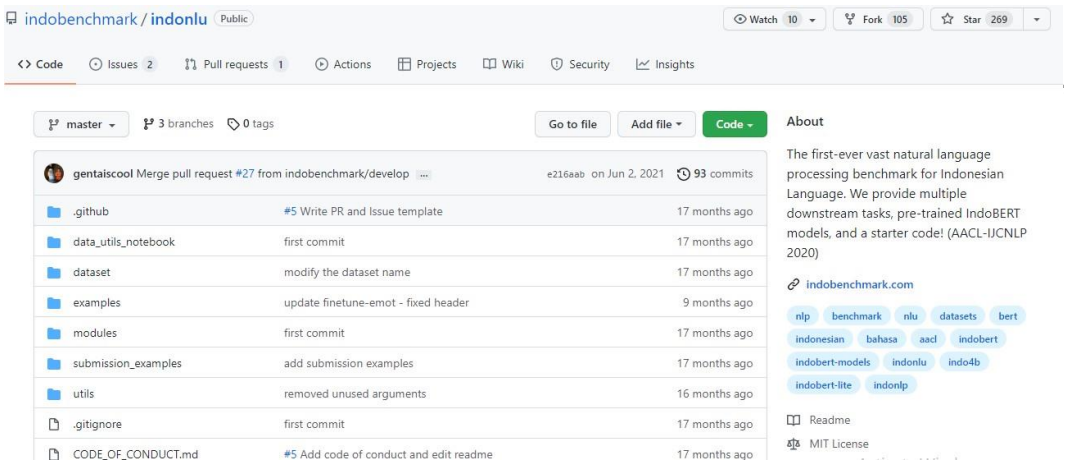
Tabel 4.3 Spesifikasi Perangkat Lunak

Sistem Operasi	Windows 10 Home Single Language 64-bit
Bahasa Pemrograman	Python
Tools	Colaboratory, Heroku CLI, Git, Visual Studio Code
Library	flask, transformers, tokenizers, torch, numpy, pandas, telegram, tqdm, nltk, python-telegram-bot, gunicorn

4.2. Data Preparation

4.2.1. Download Dataset

Dataset yang dibutuhkan pada penelitian ini adalah kumpulan teks berbahasa Indonesia yang nantinya akan dipakai untuk melakukan *finetuning* dan *train* pada model IndoBERT-lite. *Dataset* pada penelitian ini dapat diunduh repositori GitHub bernama indonlu.



Gambar 4.1 Download Dataset

Dataset ini memiliki tipe *file* csv (*comma-separated values*). Dataset ini terdiri dari 3 set, yaitu *training*, *validation*, dan *test*. *Training set* memiliki 3.523 baris data, *validation set* memiliki 440 baris data, dan *test set* memiliki 440 baris

data. Baris-baris data tersebut merupakan cuitan-cuitan warganet berbahasa Indonesia yang diambil dari Twitter.

4.2.2. Tokenization

Langkah selanjutnya setelah mendapatkan *dataset* adalah melakukan tokenisasi. Tahap awal dalam proses tokenisasi yaitu mengunduh token-token yang sudah dilatih (*pre-trained tokenizer*). Penelitian ini menggunakan *pre-trained tokenizer* khusus milik model IndoBERT-liteLARGE fase pertama. Hal ini dilakukan karena *pretrained-model* seperti IndoBERT, hanya akan bekerja dengan baik jika diberi input token-token yang sama seperti token-token yang digunakan untuk tokenisasi *pre-trained model* tersebut. Akan tetapi, sebelum memulai proses tokenisasi dan proses-proses lainnya, pustaka-pustaka (*library*) yang diperlukan untuk keperluan penelitian ini harus dimuat terlebih dahulu. Pustaka-pustaka yang diperlukan bermacam-macam, seperti pustaka untuk mengatur *file path* (os, sys), membuat *neural network* (torch), mengunduh *pre-trained model* & *tokenizer* (transformers), menghitung nilai *loss* & *load data* (utils), dan fungsi-fungsi tambahan untuk melatih model. Pustaka-pustaka tersebut dapat dilihat pada *Kode 4.1* di bawah

```
# library untuk mengatur file path
import os, sys
sys.path.append('../')

os.chdir('../')


# library untuk membuat fungsi seed
import random

import numpy as np
```

```
# library untuk membuat neural network
import torch

from torch import optim
import torch.nn.functional as F
from tqdm import tqdm


# library untuk load pre-trained model, konfigurasi, dan tokenizer IndoBert
from transformers import BertTokenizer, AlbertConfig,

AlbertForSequenceClassification


# library untuk training, load data, & menghitung nilai loss
from utils.forward_fn import forward_sequence_classification
from utils.metrics import document_sentiment_metrics_fn
```

Kode 4.1 Kode untuk Import Library

4.2.3. Data *Shuffing*

Tiap-tiap *Dataset* harus diacak setelah melalui tahap *tokenization* dan *data labelling*, untuk meminimalisasi varians dan memastikan model akan mempelajari representasi data dengan baik. Oleh karena itu, tahap *data shuffling* diperlukan. Berikut merupakan potongan kode untuk *data shuffling*:

```
train_loader =
EmotionDetectionDataLoader(dataset=train_dataset, max_seq_len=512, batc
h_size=16, num_workers=16, shuffle=True)


valid_loader =
EmotionDetectionDataLoader(dataset=valid_dataset, max_seq_len=512, batc
h_size=16, num_workers=16, shuffle=False)


test_loader =
EmotionDetectionDataLoader(dataset=test_dataset, max_seq_len=512, batch
```

Fungsi Emotion Detection DataLoader () diatas memiliki parameter *shuffle* yang berguna untuk mengacak data. Selain itu, terdapat parameter-parameter lain seperti max_seq_len dan batch_size, yang berfungsi untuk menentukan panjang maksimal pada kalimat dan memuat data per *batch*.

4.3. Modelling

Dataset siap dipakai untuk tahap *modelling* setelah melalui proses *data preparation*. Dalam tahap *modelling* ini dijelaskan dari cara mempersiapkan model, melatih model, hingga menyimpan model.

4.3.1. Load Model

Proses *load model* dilakukan dengan menggunakan fungsi dari pustaka Transformers, yaitu AlbertForSequenceClassification.from_pretrained(). Selain itu, agar model IndoBERT-lite yang diunduh bekerja lebih baik, penting untuk menggunakan konfigurasi yang sudah diuji coba sebelumnya. Konfigurasi yang dimaksud adalah seperti *embedding size*, *vocab size*, dll. Untuk mengunduh konfigurasi yang sudah diuji coba, digunakan fungsi AlbertConfig.from_pretrained (), yang merupakan pustaka dari Transformers. Berikut ini potongan kode untuk *load model.*

```
config =

AlbertConfig.from_pretrained("indobenchmark/indobert-lite-large-p1")

model =
AlbertForSequenceClassification.from_pretrained("indobenchmark/indobert

-lite-large-p1", config=config)
```

Kode 4.6 Load Model

4.3.2. Finetune & Train Model

Model *pre-trained* seperti IndoBERT tidak bisa langsung dipakai untuk mengklasifikasi emosi. Diperlukan proses *finetuning* dan *training*

pada model supaya model mempunyai akurasi yang bagus sehingga bisa digunakan dan masuk ke tahap produksi. Berikut ini potongan kode untuk tahapan *finetune* dan *train model*

```
optimizer = optim.Adam(model.parameters(), lr=3e-6)
model = model.cuda()
```

Kode 4.7 Kode untuk Menggunakan Optimizer dan Pengecekan GPU

Tahap pertama dalam *finetuning* yang dilakukan adalah memilih *optimizer* yaitu ADAM dan menentukan *learning rate* bernilai 0,000003. Setelah itu, langkah selanjutnya adalah memberitahu model bahwa nantinya model akan dilatih menggunakan GPU.

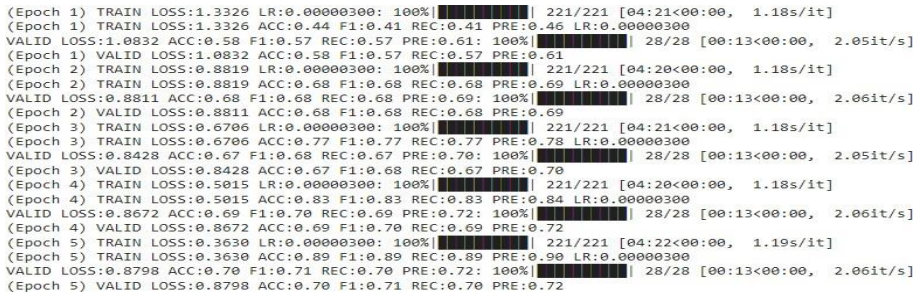
```
# Train
n_epochs = 5

for epoch in range(n_epochs):
    model.train()
    torch.set_grad_enabled(True)


    total_train_loss = 0
    list_hyp, list_label = [], []


    train_pbar = tqdm(train_loader, leave=True, total=len(train_loader)

)

    for i, batch_data in enumerate(train_pbar):
        # Forward model

        loss, batch_hyp, batch_label = forward_sequence_classification(
model, batch_data[:-1], i2w=i2w, device='cuda')


        # Update model
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()


        tr_loss = loss.item()

        total_train_loss = total_train_loss + tr_loss
```

4.3.3. Evaluate Model

```
# Calculate metrics
        list_hyp += batch_hyp
        list_label += batch_label

        train_pbar.set_description("(Epoch {}) TRAIN LOSS:{:.4f} LR:{:.
8f}".format((epoch+1),
            total_train_loss/(i+1), get_lr(optimizer)))

    # Calculate train metric
    metrics = document_sentiment_metrics_fn(list_hyp, list_label)
    print("(Epoch {}) TRAIN LOSS:{:.4f} {} LR:{:.8f}".format((epoch+1),
        total_train_loss/(i+1), metrics_to_string(metrics), get_lr(opti
mizer)))

    # Evaluate on validation
    model.eval()
    torch.set_grad_enabled(False)

    total_loss, total_correct, total_labels = 0, 0, 0
    list_hyp, list_label = [], []

    pbar = tqdm(valid_loader, leave=True, total=len(valid_loader))
    for i, batch_data in enumerate(pbar):
        batch_seq = batch_data[-1]
        loss, batch_hyp, batch_label = forward_sequence_classification(
model, batch_data[:-1], i2w=i2w, device='cuda')

        # Calculate total loss
        valid_loss = loss.item()
        total_loss = total_loss + valid_loss
```

Tahap *evaluate model* diatas berguna untuk mengetahui tingkat keakuratan pada model. Penulis menggunakan nilai-nilai di dalam

confusion matrix, yaitu *accuracy*, *precision*, *recall*, dan F1. Gambar dibawah memperlihatkan nilai-nilai *confusion matrix* yang didapat



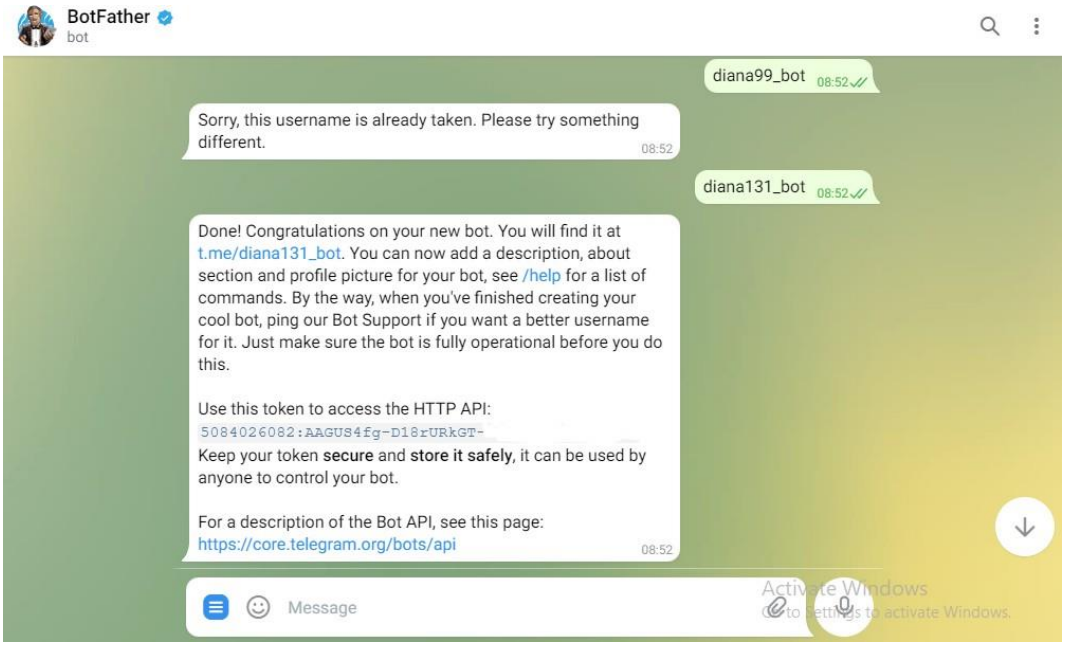
4.4. Application Development

4.4.1. Pembuatan *Bot* di Telegram

Cara membuat bot di Telegram sangatlah mudah karena Telegram mempunyai *bot* bernama BotFather yang berguna untuk membuat *bot* di Telegram menjadi cepat. Namun sebelumnya, terdapat 2 langkah yang harus dilakukan, yaitu mengunduh aplikasi Telegram dan membuat akun Telegram. Setelah itu, mencari BotFather di kolom *search*. Kemudian langkah-langkah berikut dilakukan untuk membuat *bot.*

1. Mengetikkan *command* "/start" untuk memulai interaksi ke BotFathe.
2. Mengetikkan *command* "/newbot" untuk membuat *bot* baru.
3. Mengetikkan nama untuk *bot* yang akan dibuat.
4. Mengetikkan *username* untuk *bot.*

Apabila muncul tulisan "Done! Congratulations on your new bot." artinya *bot* telah berhasil dibuat. Token untuk mengakses HTTP API Telegram yang didapat nantinya akan dipakai dalam pembuatan *bot engine*.



Gambar 4.7 BotFather

BAB VI

PENUTUP

5.1. Kesimpulan

Perhitungan emosi dari teks adalah bidang baru-baru ini. Dalam linguistik, perhitungan otomatis dan deteksi emosi dari teks menjadi salah satu topik yang sangat banyak diteliti. Salah satu pendekatan mendeteksi emosi dari teks adalah leksikon, EmoLex merupakan leksikon emosi yang tersedia dalam 105 bahasa di dunia salah satunya Indonesia yang memuat 14.182 kata. Kateglo API sebuah tesaurus bahasa Indonesia dapat dimanfaatkan untuk memperluas EmoLex dengan mencarikan sinonim tiap kata yang ada pada EmoLex menjadi 20.690 kata. Hasil pengujian menunjukkan bahwa leksikon emosi mampu mendeteksi 55.45% atau 15.357 dari 27.696 kata yang. NLP dapat dimanfaatkan untuk memperbaiki teks yang berasal dari *update* status. Hasil perbaikan dari NLP dicocokkan dengan leksikon yang telah dibuat untuk mengetahui label emosi dari suatu *update* status. Sebanyak 26 *update status* dapat terdeteksi label emosinya dan 61,53% di antaranya akurat.

5.2. Saran

Peneliti berharap pada penelitian selanjutnya dapat meningkatkan nilai akurasi dari pelabelan emosi. Meningkatkan prapemrosesan dengan membuang nama orang, nama tempat dan kata sambung, serta kata yang tidak baku dalam bahasa Indonesia.

DAFTAR PUSTAKA

[1] Aleedy, M., Shaiba, H., & Bezbradica, M. (2019). Generating and Analyzing Chatbot Responses using Natural Language Processing. Internasional Journal of Advanced Computer Science and Applications, 10(9). doi:10.14569/ijacsa.2019.0100910.

[2] Auslander, P. (2002). Live from Cyberspace: Or, I Was Sitting at My Computer This Guy Appeared He Thought I Was a Bot. *PAJ: A Journal of Performanceand Art*, *24*(1), 16–21. http://www.jstor.org/stable/3246456.

[3] Chaplin, J. P. (1989). *Kamus Lengkap Psikologi*. *(Dr. Kartini Kartono, Terjemahan)*. Jakarta: PT. Raja Grafindo Persada.

[4] Chatterjee, A., Narahari, K. N., Joshi, M., & Agrawal, P. (2019). *SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text*. 39–48. https://doi.org/10.18653/v1/s19-2005.

[5] Diamantini, C., Mircoli, A., & Potena, D. (2021). Emotion and sentiment analysis of tweets using BERT. *CEUR Workshop Proceedings*, *2841*.

[6] Chowdhary, K. R. (2020). Fundamentals of artificial intelligence. In *Fundamentals of Artificial Intelligence*. Springer India. https://doi.org/10.1007/978-81-322-3972-7.

[7] Dun, T. (2005, Mei). *Troublesome differences: A typology of lay views of misunderstanding*. Paper presented at the meeting of the International Communication Association Conference, New York, NY.

[8] Fera Fanesya, Randy Cahya Wihandika, I. (2019). Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(7), 3.

[9] Goleman, D. (2002). *Working With Emotional Intelligence*. Jakarta: PT. GramediaPustaka Utama.

[10] Johnson, A. J., Bostwick, E., & Anderson, C. (2016). How do computer-mediated channels negatively impact existing interpersonal relationships? In E. S. Gilchrist-Petty & S. D. Long (Eds.), *Contexts of the dark side of communication* (pp. 241–252). New York, NY: Peter Lang.

[11] Kelly, L., Keaten, J. A., Becker, B., Cole, J., Littleford, L., Rothe, B. (2012). “It's the American Lifestyle!”: An Investigation of Text Messaging by

College Students. *Qualitative Research Reports in Communication*, *13*(1), 1–9. doi:10.1080/17459435.2012.719203.

[12] Kerlyl, A., Hall, P., & Bull, S. (2007). Bringing Chatbots into education: Towards Natural Language Negotiation of Open Learner Models. In: Ellis R., Allen T., Tuson A. (eds) Applications and Innovations in Intelligent Systems XIV. SGAI 2006. Springer, London. https://doi.org/10.1007/978-1-84628-666- 7_14.

[13] M. Seligman, (2011). "Authentic Happiness". Britania Raya. Nicholas Brealey Publishing.

[14] Mahmud, M.D. (1990). *Psikologi: suatu pengantar*. Yogyakarta: BPFE

[15] Mulqueen, M. (2019, Desember 07). Texting really is ruining personal relationships. *NBC News*. https://www.nbcnews.com/think/opinion/texting-really-ruining-personal-relationships-ncna1097461.

[16] R. Arumugam, R.. Shanmugamani, (2018). "Hands-On Natural Language Processing with Python: A Practical Guide to Applying Deep Learning Architectures to Your NLP Applications." Britania Raya. Packt Publishing.

[17] Riza, M. A., & Charibaldi, N. (2021). Emotion Detection in Twitter Social Media Using Long Short-Term Memory (LSTM) and Fast Text. *International Journal of Artificial Intelligence & Robotics (IJAIR)*, *3*(1), 15–26. https://doi.org/10.25139/ijair.v3i1.3827

[18] Rohman, A. N., Utami, E., & Raharjo, S. (2019). Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing. *Eksplora Informatika*, *9* (1), 70-76.

[19] Rohman, A. N., Utami, E., & Raharjo, S. (2019). Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing. *Eksplora Informatika*, *9*(1), 70–76. https://doi.org/10.30864/eksplora.v9i1.277.

[20] Vig, J., & Belinkov, Y. (2019). *Analyzing the Structure of Attention in a Transformer Language Model*. http://arxiv.org/abs/1906.04284.

[21] Wiliandari, Y. (2013). Aplikasi Kecerdasan Emosional dalam Prilaku Organisasi. *Society, Jurnal Jurusan Pendidikan IPS Ekonomi*.

[22] Wiliandari, Y. (2013). Aplikasi Kecerdasan Emosional dalam Prilaku Organisasi. *Society, Jurnal Jurusan Pendidikan IPS Ekonomi*.

[23] Y. Liu, (2018). “Deep Learning in Natural Language Processing.” Jerman. Springer Singapore.