

ANALISIS SENTIMEN KOMENTAR VAKSINASI COVID-19 DI INSTAGRAM MENGGUNAKAN DEEP LEARNING XLNET

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Muhammad Rafi
NIM: 09021281823163

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022

LEMBAR PENGESAHAN SKRIPSI

ANALISIS SENTIMEN KOMENTAR VAKSINASI COVID-19 DI INSTAGRAM MENGGUNAKAN DEEP LEARNING XLNET

Oleh:

Muhammad Rafi
NIM: 09021281823163

Inderalaya, 25 Agustus 2022

Pembimbing I



Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Pembimbing II,



Novi Yusliani, M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 5 Agustus 2022 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhammad Rafi
NIM : 09021281823163
Judul : Analisis Sentimen Komentar Vaksinasi Covid-19 di Instagram Menggunakan Deep Learning XLNet

dan dinyatakan **LULUS**.

1. Ketua

Osvari Arsalan, M.T.
NIP 198806282018031001



2. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003



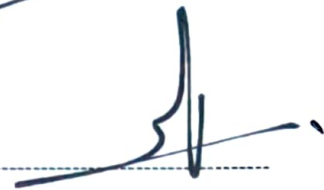
3. Penguji II

Desty Rodiah, M.T.
NIP 198912212020122011



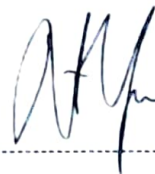
4. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.
NIP 198410012009121005



5. Pembimbing II

Novi Yusliani, M.T.
NIP 198211082012122001



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Muhammad Rafi

NIM : 09021281823163

Program Studi : Teknik Informatika

Judul Skripsi : Analisis Sentimen Komentar Vaksinasi Covid-19 di Instagram
menggunakan Deep Learning XLNet

Hasil pengecekan software *iThenticate/Turnitin* : 7%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Indralaya, 25 Agustus 2022



Muhammad Rafi
NIM. 09021281823163

MOTTO DAN PERSEMBAHAN

Motto:

I am not what happened to me, I am what I choose to become.

(Carl Gustav Jung)

Love what you do; Do what you love.

(Wayne W. Dyer)

Kupersembahkan Karya Tulis ini kepada :

- Allah SWT
- Kedua orang tua dan kedua adik saya
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

The Covid-19 vaccination activity is one of the most discussed things on social media. Comments on social media can be used to look at public sentiment on the Covid-19 vaccination. This study aims to perform sentiment analysis on Instagram comments about Covid-19 vaccination using XLNet and look at its performance. This study used Wikipedia corpus data and 2.000 Indonesian comments from Instagram. This study uses two software, namely training and test software. Training software is used to pre-train and fine-tune the XLNet model. Test software used for testing sentiment analysis using XLNet. The results of sentiment analysis on Instagram comments using XLNet get the best accuracy value on the model using epochs value: 7 and batch size: 16 in the fine-tuning process with accuracy: 74,25%, precision: 70,73%, recall: 82,75%, and f-measure: 76,27%. This study found that epoch value and batch size changes in the fine-tuning process affect the model performance. The epoch value addition and the selection of a smaller batch size value in the fine-tuning process almost always increase the model accuracy. The epochs value addition reduces accuracy in model with epochs value: 8 and batch size: 16, also in model with epochs value: 7 and batch size: 50. Meanwhile, the selection of a smaller batch size reduces accuracy in model with epochs value: 3 and batch size: 32 as well as a model with epochs value: 8 and batch size: 16.

Keyword: Sentiment Analysis, Covid-19 Vaccination, Deep Learning, XLNet

ABSTRAK

Kegiatan vaksinasi Covid-19 menjadi salah satu hal yang sering dibicarakan di media sosial. Komentar pada media sosial dapat digunakan untuk mengetahui sentimen masyarakat terhadap kegiatan vaksinasi Covid-19. Penelitian ini membahas bagaimana melakukan analisis sentimen pada komentar di Instagram mengenai vaksinasi Covid-19 menggunakan *deep learning* XLNet dan mengetahui kinerjanya. Data yang digunakan dalam penelitian ini berupa data korpus Wikipedia dan 2.000 komentar berbahasa Indonesia dari aplikasi Instagram. Pada penelitian ini digunakan dua perangkat lunak, yaitu perangkat lunak pelatihan untuk melakukan pra-pelatihan dan *fine-tuning deep learning* XLNet, serta perangkat lunak pengujian yang digunakan untuk melakukan pengujian analisis sentimen menggunakan *deep learning* XLNet. Hasil penelitian analisis sentimen komentar menggunakan *deep learning* XLNet memperoleh nilai akurasi terbaik pada model yang menggunakan nilai parameter *epoch*: 7 dan *batch size*: 16 pada proses *fine-tuning* dengan hasil akurasi: 74,25%, presisi: 70,73%, *recall*: 82,75%, dan *f-measure*: 76,27%. Dari penelitian yang telah dilakukan juga diketahui bahwa perubahan nilai *epoch* dan *batch size* pada proses *fine-tuning* berpengaruh terhadap kinerja model. Penambahan nilai *epoch* dan pemilihan nilai *batch size* yang lebih kecil pada proses *fine-tuning* hampir selalu meningkatkan akurasi model. Penambahan nilai *epoch* yang menyebabkan penurunan nilai akurasi terjadi pada model dengan konfigurasi *epoch*: 8 dan *batch size*: 16 serta model dengan konfigurasi *epoch*: 7 dan *batch size*: 50. Sementara pemilihan nilai *batch size* yang lebih kecil menurunkan akurasi pada model dengan konfigurasi *epoch*: 3 dan *batch size*: 32 dan model dengan konfigurasi *epoch*: 8 dan *batch size*: 16.

Kata Kunci: Analisis Sentimen, Vaksinasi Covid-19, *Deep Learning*, XLNet

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan nikmat-Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat menyelesaikan pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan skripsi ini, penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua, adik-adikku serta keluarga besar yang telah mendoakan, memberi semangat, motivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya dan Dosen Penguji I Tugas Akhir yang telah memberikan ilmu, nasihat serta saran yang membangun.
5. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Dosen Pembimbing I dan Ibu Novi Yusliani, M.T. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
6. Ibu Rizki Kurniati, M.T. selaku Pembimbing Akademik selama di Universitas Sriwijaya.
7. Ibu Desty Rodiah, M.T. selaku Dosen Penguji II Tugas Akhir yang telah

memberikan ilmu, nasihat serta saran yang membangun.

8. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Pihak-pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima kasih.

Indralaya, 25 Agustus 2022

Penulis

Muhammad Rafi

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
 BAB I PENDAHULUAN	 I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-5
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-6
1.6 Batasan Masalah	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan.....	I-8
 BAB II KAJIAN LITERATUR.....	 II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Analisis Sentimen	II-1
2.2.2 XLNet.....	II-3
2.2.2.1 Pra-Pelatihan	II-4
2.2.2.2 <i>Fine-Tuning</i>	II-5
2.2.3 Pengukuran Hasil Klasifikasi	II-5
2.2.4 Vaksinasi Covid-19	II-7
2.2.5 <i>Rational Unified Process</i>	II-7
2.3 Penelitian Lain yang Relevan	II-9
2.4 Kesimpulan.....	II-10
 BAB III METODOLOGI PENELITIAN.....	 III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data	III-1
3.2.2 Metode Pengumpulan Data	III-2
3.3 Tahapan Penelitian.....	III-4
3.3.1 Mengumpulkan Data	III-4
3.3.2 Menentukan Kerangka Kerja Penelitian	III-5
3.3.3 Menentukan Kriteria Pengujian	III-7

3.3.4	Menentukan Format Data Pengujian.....	III-9
3.3.5	Menentukan Alat Bantu Penelitian	III-10
3.3.6	Melakukan Pengujian Penelitian.....	III-10
3.3.7	Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-11
3.4	Metode Pengembangan Perangkat Lunak	III-11
3.4.1	Fase Insepsi	III-12
3.4.2	Fase Elaborasi	III-12
3.4.3	Fase Konstruksi.....	III-12
3.4.4	Fase Transisi.....	III-13
3.5	Kesimpulan.....	III-13
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-2
4.2.2.1	Sistem Pelatihan	IV-2
4.2.2.2	Sistem Pengujian	IV-3
4.2.3	Analisis dan Perancangan	IV-4
4.2.3.1	Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.1.1	Perangkat Lunak Pelatihan.....	IV-5
4.2.3.1.2	Perangkat Lunak Pengujian.....	IV-5
4.2.3.2	Analisis Pra-Pengolahan Data	IV-6
4.2.3.3	Analisis Proses Pelatihan	IV-9
4.2.3.4	Analisis Proses Klasifikasi	IV-10
4.2.3.5	Analisis Hasil Klasifikasi	IV-10
4.2.4	Implementasi	IV-10
4.3	Fase Elaborasi.....	IV-16
4.3.1	Pemodelan Bisnis	IV-16
4.3.1.1	Perancangan Data	IV-16
4.3.1.2	Perancangan Antarmuka.....	IV-16
4.3.2	Kebutuhan	IV-17
4.3.3	Analisis dan Perancangan	IV-18
4.3.3.1	Diagram Aktivitas	IV-18
4.3.3.2	Diagram Alur.....	IV-21
4.4	Fase Konstruksi	IV-25
4.4.1	Kebutuhan	IV-26
4.4.2	Implementasi	IV-27
4.4.2.1	Implementasi Kelas	IV-28
4.4.2.2	Implementasi Antarmuka	IV-29
4.5	Fase Transisi	IV-30
4.5.1	Pemodelan Bisnis	IV-30
4.5.2	Kebutuhan	IV-31
4.5.3	Analisis dan Perancangan	IV-31
4.5.3.1	Rencana Pengujian	IV-31
4.5.3.2	Implementasi	IV-34

4.6 Kesimpulan.....	IV-40
BAB V HASIL DAN ANALISIS	V-1
5.1 Pendahuluan	V-1
5.2 Hasil Penelitian.....	V-1
5.2.1 Hasil Pelatihan Model	V-1
5.2.2 Hasil Pengujian Model	V-21
5.3 Analisis Hasil Penelitian.....	V-23
5.3.1 Analisis Pelatihan Model	V-23
5.3.2 Analisis Hasil Pengujian Model	V-25
5.4 Kesimpulan.....	V-25
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran	VI-1
DAFTAR PUSTAKA	xviii
LAMPIRAN.....	xviii

DAFTAR TABEL

Tabel II-1. Confusion Matrix.....	II-6
Tabel III-1. Contoh Komentar yang Dikumpulkan.....	III-2
Tabel III-2. Parameter Pra-Pelatihan XLNet.....	III-8
Tabel III-3. Parameter Fine-Tuning XLNet.....	III-9
Tabel III-4. Rancangan Tabel Confusion Matrix Hasil Klasifikasi.....	III-9
Tabel III-5. Rancangan Tabel Hasil Pengujian.....	III-10
Tabel III-6. Hasil Analisis Klasifikasi.....	III-11
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak Pelatihan.....	IV-3
Tabel IV-2. Kebutuhan Non-Fungsional Perangkat Lunak Pelatihan.....	IV-3
Tabel IV-3. Kebutuhan Fungsional Perangkat Lunak Pengujian.....	IV-4
Tabel IV-4. Kebutuhan Non-Fungsional Perangkat Lunak Pengujian.....	IV-4
Tabel IV-3. Data Komentar.....	IV-6
Tabel IV-4. Hasil Pra-Pengolahan <i>Case Folding</i>	IV-7
Tabel IV-5. Hasil Pra-Pengolahan <i>Cleaning</i>	IV-8
Tabel IV-6. Hasil Pra-Pengolahan <i>Tokenizing</i>	IV-9
Tabel IV-7. Tabel Definisi <i>Use Case</i> Perangkat Lunak Pelatihan.....	IV-12
Tabel IV-8. Tabel Definisi <i>Use Case</i> Perangkat Lunak Pengujian.....	IV-12
Tabel IV-8. Tabel Definisi Aktor.....	IV-12
Tabel IV-9. Skenario <i>Use Case</i> Melakukan Pra-Pengolahan Data Masukan.....	IV-13
Tabel IV-10. Skenario <i>Use Case</i> Melakukan Pra-Pelatihan Model XLNet.....	IV-14
Tabel IV-11. Skenario <i>Use Case</i> Melakukan Proses <i>Fine-Tuning</i> Model XLNet.....	IV-14
Tabel IV-12. Skenario <i>Use Case</i> Melakukan Proses Klasifikasi Data.....	IV-15
Tabel IV-13. Keterangan Implementasi Kelas.....	IV-28
Tabel IV-14. Rencana Pengujian <i>Use Case</i> Melakukan Pra-Pengolahan Data pada Sistem Pengujian.....	IV-32
Tabel IV-15. Rencana Pengujian <i>Use Case</i> Melakukan Pra-Pengolahan Data pada Sistem Pelatihan.....	IV-32
Tabel IV-16. Rencana Pengujian <i>Use Case</i> Melakukan Pra-Pelatihan Model XLNet.....	IV-33
Tabel IV-17. Rencana Pengujian <i>Use Case</i> Melakukan <i>Fine-Tuning</i> Model XLNet.....	IV-33
Tabel IV-18. Rencana Pengujian <i>Use Case</i> Proses Klasifikasi XLNet.....	IV-33
Tabel IV-19. Pengujian <i>Use Case</i> Melakukan Pra-Pengolahan Data pada Sistem Pengujian.....	IV-35
Tabel IV-20. Pengujian <i>Use Case</i> Melakukan Pra-Pengolahan Data pada Sistem Pelatihan.....	IV-36
Tabel IV-21. Pengujian <i>Use Case</i> Proses Pra-Pelatihan Model XLNet.....	IV-37
Tabel IV-22. Pengujian <i>Use Case</i> Proses <i>Fine-Tuning</i> Model XLNet.....	IV-38
Tabel IV-23. Pengujian <i>Use Case</i> Proses Klasifikasi XLNet.....	IV-39

Tabel V-1. <i>Confusion Matrix</i> Hasil Pengujian.....	V-21
Tabel V-2. Hasil Evaluasi Analisis Sentimen Model XLNet.....	V-22

DAFTAR GAMBAR

Gambar II-1. Contoh Case Folding.....	II-2
Gambar II-2. Contoh Cleaning Text.....	II-2
Gambar II-3. Contoh Tokenizing	II-3
Gambar II-4. Arsitektur Two-Stream Self-Attention XLNet	II-4
Gambar II-5. Tahapan Rational Unified Process	II-8
 Gambar III-1. Proses mengunjungi halaman Wikimedia Downloads.....	III-3
Gambar III-2. Tautan artikel Wikipedia	III-3
Gambar III-3. Diagram Tahapan Penelitian	III-4
Gambar III-4. Diagram Alur Proses Umum Perangkat Lunak.....	III-5
 Gambar IV-1. Diagram <i>Use Case</i> Perangkat Lunak Pelatihan	IV-11
Gambar IV-2. Diagram <i>Use Case</i> Perangkat Lunak Pengujian	IV-11
Gambar IV-3. Rancangan Antarmuka Perangkat Lunak	IV-17
Gambar IV-4. Diagram Aktivitas Melakukan Pra-Pengolahan Data pada Sistem Pengujian.....	IV-19
Gambar IV-5. Diagram Aktivitas Melakukan Pra-Pengolahan Data pada Sistem Pelatihan.....	IV-19
Gambar IV-6. Diagram Aktivitas Melakukan Pra-Pelatihan Model XLNet	IV-20
Gambar IV-7. Diagram Aktivitas Melakukan <i>Fine-Tuning</i> Model XLNet.....	IV-20
Gambar IV-8. Diagram Aktivitas Melakukan Klasifikasi	IV-21
Gambar IV-9. Diagram Alur Melakukan Pra-Pengolahan Data pada Sistem Pengujian.....	IV-22
Gambar IV-10. Diagram Alur Melakukan Pra-Pengolahan Data pada Sistem Pelatihan.....	IV-22
Gambar IV-11. Diagram Alur Melakukan Proses Pra-Pelatihan Model XLNet	IV-23
Gambar IV-12. Diagram Alur Melakukan Proses <i>Fine-Tuning</i> Model XLNet	IV-24
Gambar IV-13. Diagram Alur Melakukan Proses Klasifikasi XLNet.....	IV-25
Gambar IV-14. Diagram Kelas Perangkat Lunak Pelatihan	IV-26
Gambar IV-15. Diagram Kelas Perangkat Lunak Pengujian	IV-27
Gambar IV-16. Implementasi Antarmuka Perangkat Lunak Pengujian...	IV-30
 Gambar V-1. Plot <i>Loss</i> Pra-Pelatihan Model XLNet	V-2
Gambar V-2. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =3 & <i>Batch Size</i> =16.....	V-3
Gambar V-3. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =4 & <i>Batch Size</i> =16.....	V-4
Gambar V-8. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =3 & <i>Batch Size</i> =32.....	V-9
Gambar V-9. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =4 & <i>Batch Size</i> =32.....	V-10

Gambar V-10. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =5 & <i>Batch Size</i> =32.....	V-11
Gambar V-11. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =6 & <i>Batch Size</i> =32.....	V-12
Gambar V-12. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =7 & <i>Batch Size</i> =32.....	V-13
Gambar V-13. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =8 & <i>Batch Size</i> =32.....	V-14
Gambar V-14. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =3 & <i>Batch Size</i> =50.....	V-15
Gambar V-15. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =4 & <i>Batch Size</i> =50.....	V-16
Gambar V-16. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =5 & <i>Batch Size</i> =50.....	V-17
Gambar V-17. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =6 & <i>Batch Size</i> =50.....	V-18
Gambar V-18. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =7 & <i>Batch Size</i> =50.....	V-19
Gambar V-19. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model XLNet <i>Epoch</i> =8 & <i>Batch Size</i> =50.....	V-20
Gambar V-8. Pesan <i>Error</i> Kekurangan Memori Selama Pelatihan.....	V-24

DAFTAR LAMPIRAN

1. Kode Program

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah. Bab ini akan berisi penjelasan umum mengenai keseluruhan penelitian.

1.2 Latar Belakang Masalah

Sentimen seseorang dapat ditentukan polaritasnya, positif atau negatif, dengan teknik analisis sentimen. Sentimen tersebut juga dapat berisi suasana hati, penilaian serta evaluasinya (Afif dan Pratama, 2021). Analisis sentimen merupakan pembelajaran komputasional mengenai pendapat, penilaian, sikap dan emosi seseorang terhadap suatu peristiwa, masalah, individu, topik dan atributnya (Liu and Zhang, 2012). Analisis sentimen telah meningkatkan minat komunitas ilmiah dalam berbagai bidang seperti psikologi kognitif dan sosial, pemrosesan sinyal, dan pemrosesan bahasa alami (Bouchekeif et al., 2019).

Analisis sentimen telah dibahas pada banyak literatur menggunakan metode-metode yang bervariasi seperti *Naïve Bayes Classifier* (Yulita, Nugroho dan Algifari, 2021), *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM) dan *bidirectional LSTM* (bi-LSTM) (Ali et al., 2021). *Naïve Bayes Classifier* merupakan algoritma yang cukup populer untuk klasifikasi teks, dan memiliki kinerja cukup baik pada banyak domain (Ye, Zhang and Law, 2009).

Akan tetapi, data dengan jumlah fitur terlalu banyak dapat meningkatkan waktu komputasi dan mengurangi akurasi metode tersebut (Negara, Muhandi dan Putri, 2020). CNN dapat digunakan untuk memperoleh informasi kontekstual lokal pada teks (Ali et al., 2021) akan tetapi tidak dapat digunakan untuk memperoleh informasi kontekstual secara keseluruhan (Wang et al., 2021). LSTM memiliki kelebihan yaitu dapat memperoleh informasi kontekstual secara keseluruhan dari teks (Bouhekif et al., 2019), akan tetapi LSTM hanya dapat menganalisis informasi kontekstual dari satu arah (Al-Omari, Abdullah and Shaikh, 2020). Metode bi-LSTM merupakan pengembangan dari LSTM untuk menganalisis informasi dari dua arah (Bouhekif et al., 2019), namun semakin panjang kalimat, maka kinerja model akan menurun secara eksponensial (Topal, Bas and Heerden, 2021).

Metode lain yang dapat digunakan untuk melakukan analisis sentimen yaitu dengan menggunakan model bahasa pra-pelatihan. Model bahasa pra-pelatihan dan *transformers* akhir-akhir ini menunjukkan peningkatan kemampuan dalam beberapa tugas pemrosesan bahasa alami dan sekarang dianggap sebagai *state-of-the-art* pada representasi kata kontekstual. *Generative Pre-Trained Transformer* (GPT) dan *Bidirectional Encoder Representations from Transformers* (BERT) secara bertahap menggantikan penyisipan kata statis dalam pemrosesan bahasa alami dan dapat disempurnakan (*fine-tuned*) pada tugas yang berbeda (Alexandridis et al., 2021).

GPT merupakan model bahasa pra-pelatihan *autoregressive* (AR) searah. Model bahasa searah dapat membatasi kemampuan representasi pra-pelatihan,

terutama pada pendekatan *fine-tuning*. Keterbatasan utama pada model bahasa searah ialah terbatasnya pilihan arsitektur yang dapat digunakan dalam proses pra-pelatihan. Pada GPT, digunakan arsitektur kiri ke kanan dalam proses pra-pelatihan, sehingga setiap token hanya dapat mengikuti token sebelumnya pada *layer self-attention* transformer. Pembatasan ini menyebabkan model menjadi kurang optimal pada tugas tingkat kalimat dan sangat berbahaya ketika menerapkan pendekatan *fine-tuning* pada tugas yang membutuhkan penggabungan konteks dari dua arah (Devlin et al., 2018).

BERT menggunakan dua tugas *unsupervised* pada proses pelatihan, untuk memahami bahasa dan konteks, yaitu *masked language modeling* (MLM) dan *next sentence prediction* (NSP). MLM digunakan untuk mendapatkan konteks pada kalimat dari dua arah, sementara NSP digunakan untuk mengklasifikasi apakah sebuah kalimat mengikuti kalimat lainnya, sehingga memungkinkan BERT untuk memahami ketergantungan jangka panjang di seluruh kalimat (Qian, Xie and Bruckmann, 2022).

Akan tetapi, simbol buatan seperti [MASK] yang digunakan BERT pada proses pra-pelatihan tidak muncul pada data asli selama proses *fine-tuning*, sehingga menyebabkan ketidaksesuaian antara pra-pelatihan dan *fine-tuning*. Selain itu, karena token yang diprediksi tertutup dalam input, BERT tidak dapat memodelkan probabilitas gabungan menggunakan *product rule* seperti pada model bahasa AR. BERT mengasumsikan token yang diprediksi bersifat independen terhadap token lain yang tidak tertutup. Hal ini terlalu disederhanakan karena ketergantungan jarak jauh dan tingkat tinggi merupakan hal lazim pada bahasa alami (Yang et al.,

2019).

XLNet merupakan model bahasa permutasi yang dapat digunakan untuk menganalisis informasi kontekstual teks secara keseluruhan (Yang et al., 2019). XLNet menggunakan pendekatan permutasi untuk memodelkan konteks dari dua arah dan Transformer-XL sebagai intinya untuk meningkatkan pemahaman pada teks yang panjang (Topal, Bas and Heerden, 2021).

Penggunaan operasi permutasi bertujuan agar model bahasa mendapatkan pemahaman kontekstual dari dua arah. Selain itu, penggunaan operasi permutasi menyebabkan model XLNet tidak bergantung pada korupsi data, sehingga tidak terjadi ketidaksesuaian antara pra-pelatihan dan *fine-tuning*. Sementara itu, XLNet juga menyediakan jalan untuk menggunakan *product rule* untuk memodelkan probabilitas gabungan token yang diprediksi, guna mengatasi asumsi independen yang dibuat oleh BERT. Secara empiris, jika menggunakan pengaturan yang sebanding, XLNet mampu mengungguli BERT pada 20 tugas, termasuk menjawab pertanyaan, inferensi bahasa alami, analisis sentimen dan peringkat dokumen (Yang et al., 2019).

Komunikasi adalah hal yang penting bagi manusia. Setiap manusia pasti melakukan kegiatan komunikasi dalam hidupnya. Saat ini komunikasi dapat dengan mudah dilakukan melalui berbagai media sosial. Media sosial juga menjadi sumber informasi utama masyarakat di dunia. Instagram menjadi salah satu dari 10 media sosial yang terpopuler di Indonesia (Ayutiani dan Putri, 2018).

Salah satu informasi yang banyak dibicarakan oleh pengguna media sosial adalah pemberitaan mengenai kegiatan vaksinasi Covid-19. Masyarakat memiliki

pandangan yang berbeda mengenai kegiatan vaksinasi. Sebagian masyarakat bersedia untuk menerima vaksin Covid-19, ada juga yang ragu bahkan menolak untuk mengikuti kegiatan vaksinasi Covid-19. Mereka yang ragu atau menolak kegiatan vaksinasi Covid-19 mempertimbangkan tingkat keamanan dan efektivitas vaksin Covid-19 (Izmi, Hajrah dan Aj, 2021). Analisis sentimen mengenai vaksinasi Covid-19 dapat membantu penyelenggara kegiatan vaksinasi untuk meningkatkan efektivitas vaksinasi sehingga masyarakat percaya pada proses vaksinasi (Alam et al., 2021).

Pada penelitian ini XLNet akan digunakan untuk menganalisis sentimen komentar mengenai vaksinasi Covid-19 di Instagram.

1.3 Rumusan Masalah

Berdasarkan permasalahan yang telah dijelaskan pada latar belakang, rumusan masalah penelitian ini adalah

1. Bagaimana mengembangkan analisis sentimen mengenai kegiatan vaksinasi Covid-19 menggunakan model XLNet?
2. Bagaimana kinerja analisis sentimen mengenai kegiatan vaksinasi Covid-19 menggunakan model XLNet?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah

1. Menghasilkan perangkat lunak yang dapat melakukan analisis sentimen mengenai kegiatan vaksinasi Covid-19 menggunakan model XLNet.
2. Mengetahui akurasi analisis sentimen mengenai kegiatan vaksinasi Covid-19

menggunakan model XLNet.

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah

1. Hasil analisis sentimen dapat digunakan sebagai dasar evaluasi pada pelaksanaan kegiatan vaksinasi Covid-19.
2. Hasil penelitian dapat dijadikan sebagai rujukan penelitian terkait.

1.6 Batasan Masalah

Batasan masalah penelitian ini adalah

1. Data yang digunakan adalah data korpus Wikipedia dan data komentar berbahasa Indonesia mengenai kegiatan vaksinasi Covid-19 yang diambil dari Instagram¹.
2. Klasifikasi terdiri atas dua kelas, yaitu positif dan negatif.
3. Data komentar yang digunakan adalah 2.000 komentar Instagram terdiri dari 1.000 komentar positif dan 1.000 komentar negatif.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penelitian yang akan dijadikan sebagai

¹ <https://www.instagram.com>

pokok pikiran penelitian ini.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas landasan teori yang digunakan dalam penelitian, seperti definisi analisis sentimen dan model XLNet, serta beberapa literatur yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas proses yang akan dilaksanakan selama penelitian, Seperti pengumpulan data, analisis data dan perancangan perangkat lunak. Setiap tahap akan dijelaskan berdasarkan kerangka kerja yang dibuat.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas analisis dan rancangan perangkat lunak yang akan dikembangkan. Diawali dari analisis kebutuhan, perancangan dan konstruksi perangkat lunak, dan diakhiri dengan pengujian untuk memastikan sistem yang dikembangkan sudah sesuai dengan rancangan dan kebutuhan penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini menyajikan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis diberikan sebagai dasar kesimpulan yang akan diambil dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini membahas kesimpulan yang diambil berdasarkan uraian dalam bab sebelumnya serta saran yang diberikan berdasarkan penelitian yang telah dilakukan.

1.8 Kesimpulan

Pada bab ini telah dijelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penelitian yang akan dijadikan sebagai pokok pikiran penelitian ini.

DAFTAR PUSTAKA

- Acheampong, F. A., Nunoo-Mensah, H., & Wenyu, C. Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-based Emotion Recognition.
- Afif, A. S., & Pratama, A. R. I. (2021). Analisis Sentimen Kebijakan Pendidikan di Masa Pandemi COVID-19 dengan CrowdTangle di Instagram. *AUTOMATA*, 2(2).
- Ahmed, S. S. (2021, June). Classification of Censored Tweets in Chinese Language using XLNet. In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda* (pp. 136-139).
- Al-Omari, H., Abdullah, M. A., & Shaikh, S. (2020, April). Emodet2: Emotion detection in english textual dialogue using bert and bilstm models. In *2020 11th International Conference on Information and Communication Systems (ICICS)* (pp. 226-232). IEEE.
- Alam, K. N., Khan, M. S., Dhruba, A. R., Khan, M. M., Al-Amri, J. F., Masud, M., & Rawashdeh, M. (2021). Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. *Computational and Mathematical Methods in Medicine*, 2021.
- Alexandridis, G., Varlamis, I., Koroveis, K., Caridakis, G., & Tsantilas, P. (2021). A survey on sentiment analysis and opinion mining in greek social media. *Information*, 12(8), 331.
- Ali, M. Z., Javed, K., & Tariq, A. (2021). Sentiment and Emotion Classification of Epidemic Related Bilingual data from Social Media. *arXiv preprint arXiv:2105.01468*.
- Alshahrani, A., Ghaffari, M., Amirizirtol, K., & Liu, X. (2020, July). Identifying Optimism and Pessimism in Twitter Messages Using XLNet and Deep Consensus. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Ayunda, R., Kosasih, V., & Disemadi, H. S. (2021). Perlindungan hukum bagi masyarakat terhadap efek samping pasca pelaksanaan vaksinasi covid-19 di Indonesia. *Nusantara: Jurnal Ilmu Pengetahuan Sosial*, 8(3), 194-206.
- Ayutiani, D. N., & Putri, B. P. S. (2018). Penggunaan akun instagram sebagai media informasi wisata kuliner. *PRofesi Humas*, 3(1), 39-59.
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based*

Systems, 226, 107134.

- Boucekif, A., Joshi, P., Boucekif, L., & Afli, H. (2019, June). EPITA-ADAPT at SemEval-2019 Task 3: Detecting emotions in textual conversations using deep learning models combination. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 215-219).
- Chouikhi, H., Chniter, H., & Jarray, F. (2021, September). Arabic sentiment analysis using BERT model. In *International Conference on Computational Collective Intelligence* (pp. 621-632). Springer, Cham.
- Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483.
- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan teknik supervised untuk klasifikasi teks bahasa indonesia. *Jurnal Bandung Text Mining*, 16(1), 5-1.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Gornik, D.: IBM Rational Unified Process: best practices for software development teams. Technical Report TP026B, Rev 11/01, IBM.
- Izmi, A. N., Hajrah, H., & Aj, A. A. (2021). Indonesian Netizen Reponse to Covid-19 Vaccine News in Instagram IDN Times. *INSIGHT: Indonesian Journal Social Studies and Humanities*, 1(1).
- Lestandy, M., Abdurrahim, A., & Syafa'ah, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802-808.
- Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA.
- Negara, A. B. P., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naïve Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(3).
- Pradha, S., Halgamuge, M. N., & Vinh, N. T. Q. (2019, October). Effective text data preprocessing technique for sentiment analysis in social media data. In *2019 11th international conference on knowledge and systems engineering (KSE)* (pp. 1-8). IEEE.
- Prasanti, A. A., Fauzi, M. A., & Furqon, M. T. (2018). Neighbor Weighted K-Nearest Neighbor for Sambat Online Classification. *Indonesian Journal of*

Electrical Engineering and Computer Science, 12(1), 150-160.

- Qian, T., Xie, A., & Bruckmann, C. (2022). Sensitivity Analysis on Transferred Neural Architectures of BERT and GPT-2 for Financial Sentiment Analysis. *arXiv preprint arXiv:2207.03037*.
- Suprawoto, T. (2016). Klasifikasi data mahasiswa menggunakan metode k-means untuk menunjang pemilihan strategi pemasaran. *JIKO (Jurnal Informatika dan Komputer)*, 1(1).
- Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring transformers in natural language generation: Gpt, bert, and xlnet. *arXiv preprint arXiv:2102.08036*.
- Wang, Y., Zheng, J., Li, Q., Wang, C., Zhang, H., & Gong, J. (2021). XLNet-Caps: Personality Classification from Textual Posts. *Electronics*, 10(11), 1360
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.
- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3), 6527-6535.
- Yulita, W. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1-9.
- Zhang, C., Kudo, M., & Yamana, H. Evaluation of BERT and XLNet Models on Irony Detection in English Tweets.
- Zheng, J., Cai, F., Chen, H., & de Rijke, M. (2020). Pre-train, Interact, Fine-tune: a novel interaction representation for text classification. *Information Processing & Management*, 57(6), 102215.