

Analisis Sentimen Transfer Pemain Sepak Bola di Twitter Menggunakan *Deep Learning* RoBERTa

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



Oleh:

M Ali Ridho
NIM : 09021281823065

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023

LEMBAR PENGESAHAN SKRIPSI

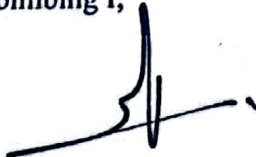
**ANALISIS SENTIMEN TRANSFER PEMAIN SEPAK BOLA
DI TWITTER MENGGUNAKAN
DEEP LEARNING ROBERTA**

Oleh:

M Ali Ridho
NIM : 09021281823065

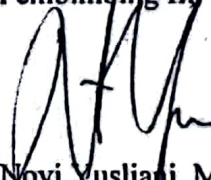
Indralaya, 4 Agustus 2023

Pembimbing I,



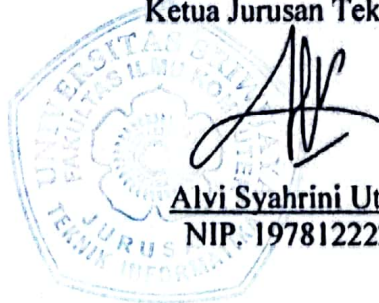
Dr. Abdiansah, M.Cs.
NIP. 198410012009121005

Pembimbing II,



Novi Yusliani, M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF

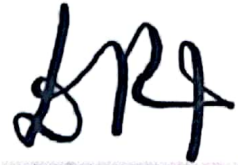
Pada hari kamis, tanggal 27 juli 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : M Ali Ridho
NIM : 09021281823065
Judul : Analisis Sentimen Transfer Pemain Sepak Bola di Twitter Menggunakan Deep Learning RoBERTa

dan dinyatakan LULUS.

1. Ketua

Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



2. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



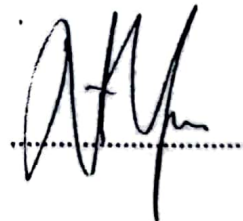
3. Pembimbing I

Dr. Abdiansah, M.Cs.
NIP. 198410012009121005



4. Pembimbing II

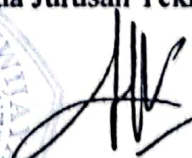
Novi Yusliani, M.T.
NIP. 198211082012122001



Mengetahui
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah :

Nama : M Ali Ridho

NIM : 09021281823065

Program Studi : Teknik Informatika

Judul Skripsi : Analisis Sentimen Transfer Pemain Sepak Bola di Twitter
Menggunakan Deep Learning RoBERTa

Hasil pengecekan software *iThenticate/Turnitin* : 16 %

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, 4 Agustus 2023



M Ali Ridho
NIM. 09021281823065

MOTO DAN PERSEMBAHAN

Moto:

“Time changes everything”. That’s what people say, it’s not true. Doing things changes things. Not doing things leaves things exactly as they were.

(Gregory House)

People don’t get what they deserve. They just get what they get. There’s nothing any of us can do about it.

(Gregory House)

Kupersembahkan karya tulis ini kepada:

- Kedua orang tua
- Kedua adik
- Sahabat terdekat
- Teman-teman
- Para guru dan dosen
- Universitas Sriwijaya

ABSTRACT

Football transfer can impact performance of a club, both on or off the pitch. By analyzing sentiment surrounding player signing, clubs can enhance their marketing strategy to achieve higher earnings. Using 1200 data from comment across some of world class player signing announcements on Twitter, this study aims to perform sentiment analysis task using RoBERTa Model on said data. To adapt RoBERTa model for sentiment analysis, the pretrained RoBERTa base model is modified with additional output layer and then the model is fine-tuned. Testing model revealed that the best accuracy is produced by using epoch value 6 and learning rate value 10^{-5} , resulting in 88,00% accuracy, 89,58% precision, 86,00% recall and 87,75% f-measure. However, there is sign of overfitting during fine-tuning model with learning rate value 10^{-5} that started to appear at epoch 4. Meanwhile, even though model with learning rate value 10^{-6} did not indicate any overfit sign, this configuration started slower even in third epoch their accuracy is still only at 50,00%. Eventually, model with learning rate value 10^{-6} caught up and achieved its best accuracy 85,50% with epoch value 8. Although it is the best for learning rate value 10^{-6} , this accuracy score is still 1,50% behind the lowest accuracy for learning rate value 10^{-5} that achieve 86,00% accuracy with only 3 epochs. Therefore, it can be concluded that using higher learning rate will result in much faster learning process though it can also drive the model to overfit. Additionally, increase epoch also can increase accuracy but the rate of improvement may vary depending on the learning rate.

Keyword: Sentiment Analysis, Football Player, Transfer, Deep Learning, RoBERTa

ABSTRAK

Transfer sepakbola dapat memengaruhi performa sebuah klub, baik di dalam maupun di luar lapangan. Dengan menganalisis sentimen seputar penandatanganan pemain, klub dapat meningkatkan strategi pemasaran mereka untuk mencapai pendapatan yang lebih tinggi. Menggunakan 1200 data dari komentar di beberapa pengumuman penandatanganan pemain kelas dunia di Twitter, penelitian ini bertujuan untuk melakukan tugas analisis sentimen menggunakan Model RoBERTa pada data tersebut. Untuk menyesuaikan model RoBERTa untuk analisis sentimen, model dasar RoBERTa yang telah dilatih sebelumnya dimodifikasi dengan lapisan keluaran tambahan, kemudian model tersebut disesuaikan kembali. Pengujian model menunjukkan bahwa akurasi terbaik dihasilkan dengan menggunakan nilai *epoch* 6 dan nilai *learning rate* 10^{-5} , menghasilkan akurasi 88,00%, presisi 89,58%, *recall* 86,00% dan *f-measure* 87,75%. Namun pada saat proses penyesuaian kembali model dengan nilai *learning rate* 10^{-5} terdapat pertanda *overfit* yang mulai terlihat pada *epoch* 4. Sementara itu, meskipun model dengan nilai *learning rate* 10^{-6} tidak menunjukkan adanya tanda *overfit*, konfigurasi ini memulai secara lambat bahkan pada *epoch* ketiga akurasinya hanya di 50,00%. Akhirnya, model dengan nilai *learning rate* 10^{-6} berhasil mengejar dan mencapai akurasi terbaiknya 85,50% dengan nilai *epoch* 8. Meskipun terbaik untuk nilai *learning rate* 10^{-6} , skor akurasi ini masih tertinggal 1,50% dari akurasi terendah untuk nilai *learning rate* 10^{-5} yang mencapai akurasi 86,00% dengan hanya 3 *epoch*. Oleh karena itu, dapat disimpulkan bahwa penggunaan *learning rate* yang lebih tinggi akan menghasilkan proses pembelajaran yang jauh lebih cepat meskipun juga dapat mendorong model untuk *overfit*. Selain itu, peningkatan *epoch* juga dapat meningkatkan akurasi tetapi tingkat peningkatan dapat bervariasi tergantung pada *learning rate*.

Kata kunci: Analisis Sentimen, Pemain Sepak Bola, Kepindahan, Pembelajaran Mendalam, RoBERTa

KATA PENGANTAR

Puji dan syukur kepada Allah SWT yang telah memberikan nikmat untuk dapat menyelesaikan laporan penelitian ini. Begitu pula shalawat dan salam kepada Nabi Muhammad saw. Pada bagian pengantar ini, peneliti akan menggunakan kesempatan ini untuk menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa dalam selesainya laporan ini, termasuk di antaranya kepada yang tersebut pada daftar berikut.

1. Kedua orang tua dan kedua adikku serta keluarga besar yang telah mendukung selama proses pengerjaan untuk menyelesaikan skripsi ini.
2. Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M. Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya dan Penguji Tugas Akhir yang telah memberikan kritik dan saran yang bermanfaat.
4. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Dosen Pembimbing I dan Ibu Novi Yusliani, M.T. selaku dosen Pembimbing II yang telah mengarahkan, memberikan masukan dan memotivasi penulis.
5. Ibu Mastura Diana Marieska, M.T. selaku Pembimbing Akademik selama di Universitas Sriwijaya yang telah membimbing selama proses perkuliahan, kerja praktik dan tugas akhir.
6. Ibu Dian Palupi Rini, M.Kom., Ph.D. selaku Ketua Sidang Tugas Akhir yang telah membantu dan memudahkan berjalannya proses sidang tugas akhir.

7. Seluruh dosen program studi yang telah memberikan ilmu dan pengetahuan serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Habibati yang telah menemani dalam setiap proses.
9. Pihak-pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Pada laporan ini tentunya akan didapati kesalahan dan kekurangan di sana sini. Atas dasar itu, kritik dan saran sangat diharapkan untuk meningkatkan hal-hal tersebut. Meskipun demikian, penulis juga tetap berharap agar hasil karya tulis ini dapat membawa manfaat bagi kita semua. Terima kasih.

Indralaya, 28 Juli 2023

Penulis,

M Ali Ridho

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF.....	iii
HALAMAN PERNYATAAN	iv
MOTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-5
1.5 Manfaat Penelitian.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-6
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Transfer Pemain Sepak Bola	II-1
2.2.2 Balasan Twitter	II-2
2.2.3 Analisis Sentimen.....	II-3
2.2.4 Transformer	II-3
2.2.5 BERT.....	II-6
2.2.6 RoBERTa	II-7
2.2.7 Pengukuran Hasil Klasifikasi	II-10

2.2.8 <i>Rational Unified Process (RUP)</i>	II-11
2.3 Penelitian Lain yang Relevan.....	II-13
2.4 Kesimpulan.....	II-18
BAB III METODE 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-2
3.3.1 Mengumpulkan Data	III-3
3.3.2 Kerangka Kerja Penelitian	III-3
3.3.3 Menentukan Kriteria Pengujian	III-4
3.3.4 Menentukan Format Data Pengujian.....	III-5
3.3.5 Menentukan Alat Penelitian	III-6
3.3.6 Pengujian Penelitian.....	III-6
3.3.7 Menganalisis dan Menarik Kesimpulan.....	III-6
3.4 Metode Pengembangan Perangkat Lunak	III-7
3.5 Manajemen Proyek Perangkat Lunak	III-7
3.6 Kesimpulan.....	III-9
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
4.2.3 Analisis dan Perancangan.....	IV-4
4.2.4 Implementasi	IV-10
4.3 Fase Elaborasi	IV-13
4.3.1 Pemodelan Bisnis	IV-13
4.3.2 Kebutuhan	IV-14
4.3.3 Analisis dan Perancangan.....	IV-15
4.4 Fase Konstruksi	IV-19
4.4.1 Kebutuhan	IV-19
4.4.2 Implementasi	IV-20

4.5	Fase Transisi.....	IV-22
4.5.1	Pemodelan Bisnis	IV-22
4.5.2	Kebutuhan Sistem	IV-23
4.5.3	Analisis Perancangan	IV-23
4.6	Kesimpulan.....	IV-27
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-12
5.3	Analisis Hasil Penelitian	V-15
5.3.1	Analisis Hasil Pelatihan Model	V-15
5.3.2	Analisis Hasil Pengujian Model	V-16
5.4	Kesimpulan.....	V-17
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-1
DAFTAR PUSTAKA		xvii

DAFTAR TABEL

Tabel II-1. Skor F1 SQuAD dan akurasi MNLI-m dan SST-2 RoBERTa	II-9
Tabel II-2. Hasil Evaluasi BERT, ALBERT dan RoBERTa	II-16
Tabel II-3. Performa Ensembling Model BERT, ALBERT dan RoBERTa	II-17
Tabel II-4. Performa Ensembling Model yang Dikompres.....	II-17
Tabel III-1. Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi	III-5
Tabel III-2. Rancangan Tabel Hasil Pengujian	III-5
Tabel III-3. Tabel Rencana Jadwal Penelitian	III-8
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 Pelatihan.....	IV-4
Tabel IV-4. Kebutuhan Non-Fungsional Perangkat Lunak Pengujian	IV-4
Tabel IV-5. Data Komentar.....	IV-6
Tabel IV-6. Hasil Pra-Pengolahan <i>Case Folding</i>	IV-7
Tabel IV-7. Hasil Pra-Pengolahan Pembersihan.....	IV-8
Tabel IV-8. Hasil Pra-Pengolahan <i>Tokenizing</i>	IV-9
Tabel IV-9. Definisi <i>Use Case</i> Perangkat Lunak Pengujian.....	IV-11
Tabel IV-10. Definisi Aktor	IV-11
Tabel IV-11. Skenario <i>Use Case</i> Melakukan Pengujian.....	IV-11
Tabel IV-12. Keterangan Implementasi Kelas	IV-21
Tabel IV-13. Rencana Pengujian Perangkat Lunak Pelatihan Model RoBERTa.....	IV-24
Tabel IV-14. Rencana Pengujian Perangkat Lunak Pengujian Model RoBERTa.....	IV-24
Tabel IV-15. Pengujian Perangkat Lunak Pelatihan Model RoBERTa.....	IV-25
Tabel IV-16. Pengujian Perangkat Lunak Pengujian Model RoBERTa.....	IV-26
Tabel V-1. <i>Confusion Matrix</i> Hasil Pengujian.....	V-13
Tabel V-2. Hasil Evaluasi Analisis Sentimen Model RoBERTa.....	V-14

DAFTAR GAMBAR

Gambar II-1. Arsitektur Model Transformer	II-4
Gambar II-2. <i>Static Masking</i> dan <i>Dynamic Masking</i>	II-8
Gambar II-3. Arsitektur <i>Rational Unified Process</i>	II-12
Gambar III-1. Alur Tahapan Penelitian	III-2
Gambar III-2. Diagram Alir Perangkat Lunak	III-3
Gambar IV-1. Diagram <i>Use Case</i> Perangkat Lunak Pengujian	IV-10
Gambar IV-2. Racangan Antarmuka Perangkat Lunak	IV-14
Gambar IV-3. Diagram Aktivitas Melakukan Pengujian	IV-15
Gambar IV-4. Diagram Alur Melakukan <i>Fine-Tuning</i> Model RoBERTa	IV-17
Gambar IV-5. Diagram Alur Melakukan Pengujian Model RoBERTa	IV-18
Gambar IV-6. Diagram Kelas Perangkat Lunak Pelatihan	IV-19
Gambar IV-7. Diagram Kelas Perangkat Lunak Pengujian	IV-20
Gambar IV-8. Tampilan Antarmuka Perangkat Lunak Pengujian	IV-22
Gambar V-1. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 3</i>	V-2
Gambar V-2. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 4</i>	V-2
Gambar V-3. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 5</i>	V-3
Gambar V-4. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 6</i>	V-4
Gambar V-5. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 7</i>	V-4
Gambar V-6. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 8</i>	V-5
Gambar V-7. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 9</i>	V-6
Gambar V-8. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-5} dan Epoch = 10</i>	V-6
Gambar V-9. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch</i> <i>Size = 8, Learning Rate = 10^{-6} dan Epoch = 3</i>	V-7
Gambar V-10. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size = 8, Learning Rate = 10^{-6} dan Epoch = 4</i>	V-8
Gambar V-11. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size = 8, Learning Rate = 10^{-6} dan Epoch = 5</i>	V-8
Gambar V-12. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size = 8, Learning Rate = 10^{-6} dan Epoch = 6</i>	V-9
Gambar V-13. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size = 8, Learning Rate = 10^{-6} dan Epoch = 7</i>	V-10

Gambar V-14. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size</i> = 8, <i>Learning Rate</i> = 10^{-6} dan <i>Epoch</i> = 8.....	V-10
Gambar V-15. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size</i> = 8, <i>Learning Rate</i> = 10^{-6} dan <i>Epoch</i> = 9.....	V-11
Gambar V-16. Plot Akurasi dan <i>Loss Fine-Tuning</i> Model RoBERTa <i>Batch Size</i> = 8, <i>Learning Rate</i> = 10^{-6} dan <i>Epoch</i> = 10.....	V-12

DAFTAR LAMPIRAN

1. Kode Program

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian. Bab ini juga membahas batasan masalah dan sistematika penulisan. Penjelasan umum terkait penelitian juga akan dibahas di bab ini.

1.2 Latar Belakang

Saat ini, sepak bola bukan hanya sebatas olahraga saja. Sepak bola sudah berkembang menjadi industri yang besar. Laporan dari Deloitte pada 2021 menunjukkan bahwa pada musim kompetisi 2019/2020, total pendapatan kotor 20 klub sepak bola dengan pendapatan tertinggi jika digabungkan adalah 8,2 milyar euro atau sekitar 124,5 triliun rupiah.¹ Hasil penelitian oleh Rohde dan Breuer (2016) menunjukkan bahwa kesuksesan finansial sebuah klub ditentukan oleh prestasi pada kompetisi nasional dan internasional serta nilai merek. Perolehan prestasi dipengaruhi oleh investasi terhadap pemain yang tepat.

Petríková & Soroková (2015) menjelaskan bahwa performa di lapangan, citra pemain sepak bola, kepribadian dan popularitas mereka dapat meningkatkan pendapatan klub melalui penjualan lisensi maupun barang dagangan. Citra dan popularitas pemain yang didatangkan dapat diketahui dengan memperhatikan

¹ <https://www2.deloitte.com>

tanggapan dari penggemar sepak bola. Data tanggapan terhadap pemain bisa diambil dari fitur *reply* di Twitter pada unggahan peresmian transfer seorang pemain.

Aloufi dan Saddik (2018) melakukan sentimen analisis terhadap tweet yang membahas sepak bola selama FIFA World Cup 2014 dan UCL 2016/2017 menggunakan metode *Support Vector Machine* (SVM), *Multinomial Naïve Bayes* (MNB) dan *Random Forest*. Skor akurasi tertinggi adalah 0,79. Angka ini berada beberapa persen di bawah skor BERT_{LARGE} yang mendapatkan angka 0,82 pada dataset GLUE dan 0,93 pada dataset SST-2 (Devlin et al., 2019).

Wunderlich dan Memmert (2020) menggunakan metode berbasis leksikon untuk melakukan sentimen analisis tweet terhadap 10 pertandingan menggunakan tiga alat. Akurasi tertinggi yang didapat adalah 0,67 dengan menggabungkan ketiganya. Kelemahan dari metode berbasis leksikon adalah tidak dapat melakukan klasifikasi jika data uji tidak mengandung kata dalam kamus. Paling sedikit 7,8% dari data uji tidak dapat diklasifikasi dikarenakan data uji tidak mengandung kata yang terdaftar dalam kamus dan paling banyak adalah 43,4%..

Pada 2021, Sarsam et al. mendapatkan akurasi 0,81 dan nilai Hamming *loss* 0,16 dalam melakukan analisis sentimen terhadap tweet saat berlangsungnya final sepak bola di Olimpiade 2016. Metode yang digunakan untuk klasifikasi teks adalah *Naïve Bayes* Menurut Kotelnikova et al. (2021), performa pada metode-metode sentimen analisis telah meningkat pesat selama beberapa tahun terakhir.

Hal ini terjadi berkat penggunaan berbagai macam model yang berdasarkan pada arsitektur Transformer, khususnya BERT.

Wolf et al., (2020) menyebutkan bahwa arsitektur Transformer yang diperkenalkan oleh Vaswani et al., (2017) telah menjadi arsitektur yang dominan dalam pemrosesan bahasa alami secara cepat. Performa Transformer dalam memahami dan menghasilkan (*generating*) bahasa alami mengungguli model jaringan syaraf tiruan yang lain seperti Recurrent Neural Network (RNN) dan Convolutional Neural Network (CNN). Arsitektur Transformer sangat mendukung proses pra-pelatihan pada korpora teks besar sehingga menghasilkan skor akurasi yang tinggi pada tugas turunan pemrosesan bahasa alami termasuk klasifikasi teks (Yang et al., 2019).

BERT (*Bidirectional Encoder Representations from Transformer*) adalah model Pemrosesan Bahasa Alami yang dibangun di atas Transformer (Devlin et al., 2019). Menurut survei yang dilakukan Rogers, Kovaleka dan Rumshisky (2020) terhadap 150 publikasi penelitian yang menganalisis dan meningkatkan model, disimpulkan bahwa dalam kurun waktu kurang dari setahun, BERT telah menjadi dasar dimana-mana dalam penelitian pemrosesan bahasa alami.

Pada tahun 2019, peneliti dari University of Washington dan Facebook AI mempublikasikan model yang merupakan peningkatan dari BERT yang diberi nama RoBERTa (*A Robustly Optimized BERT Pretraining Approach*). RoBERTa adalah BERT yang dilatih ulang namun dengan metodologi pelatihan yang ditingkatkan, 1000% lebih banyak data dan peningkatan daya komputasi (Liu et al., 2019). Hasilnya, RoBERTa mengungguli akurasi BERT pada dataset

SQUAD, MNLI-m dan SST-2 saat dilakukan pra-pelatihan dengan dataset yang sama dengan BERT dan dengan jumlah langkah sepuluh kali lebih sedikit dibanding BERT.

Ghasiya dan Okamura (2021) menggunakan arsitektur RoBERTa untuk menganalisis sentimen judul artikel terkait pandemi COVID-19. Akurasi yang didapat adalah sebesar 90%. Adoma, Henry dan Chen (2020) menunjukkan bahwa RoBERTa mengungguli BERT, DistilBERT dan XLNet dalam tugas klasifikasi emosi pada dataset ISEAR. Dataset ISEAR terdiri dari tujuh jenis emosi. Masing-masing emosi memiliki data antara 1093 sampai 1096 contoh. Jumlah ini relatif kecil dibanding dataset seperti SST-2 (Socher et al., 2013).

Pada penelitian kali ini, data latih yang digunakan juga relatif kecil dibanding dataset seperti SST-2. Berdasarkan cara kerja RoBERTa yang menggunakan mekanisme perhatian dan dua arah untuk meningkatkan pemahaman, performa akurasi RoBERTa dalam tugas analisis sentimen pada beberapa dataset, referensi dari penelitian sebelumnya terkait analisis sentimen di bidang sepak bola dan jumlah data yang dipakai pada penelitian ini, model RoBERTa diajukan sebagai metode dalam melakukan analisis sentimen terhadap unggahan transfer pemain sepak bola.

1.3 Rumusan Masalah

Berikut ini adalah perumusan masalah berdasarkan apa yang sudah diuraikan pada latar belakang

1. Bagaimana melakukan analisis sentimen terhadap komentar penggemar pada unggahan peresmian kedatangan suatu pemain?

2. Bagaimana performa *Deep Learning* RoBERTa dalam analisis sentimen terhadap unggahan transfer pemain?

1.4 Tujuan Penelitian

Penelitian ini dilakukan bertujuan untuk

1. Melakukan analisis sentimen terhadap komentar penggemar pada unggahan peresmian kedatangan suatu pemain.
2. Mengetahui performa *Deep Learning* RoBERTa dalam analisis sentimen pada unggahan transfer pemain.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah

1. Mendapatkan sentimen dari penggemar terhadap kedatangan suatu pemain.
2. Menjadi rujukan dalam pengambilan keputusan terhadap pemain tersebut, terutama terhadap sisi pemasaran.
3. Hasil penelitian bisa menjadi rujukan bagi penelitian lain yang relevan.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan merupakan *reply* atau Balasan dari penggemar sepak bola pada tweet atau unggahan di Twitter yang berisi peresmian kedatangan seorang pemain. Dalam hal ini adalah unggahan dari akun resmi Manchester United yang berisikan berita mengenai resminya Cristiano Ronaldo membela klub tersebut.

2. Data komentar yang dikumpulkan adalah sebanyak 1200 komentar berbahasa Inggris berjumlah 600 komentar positif dan 600 komentar negatif.
3. Klasifikasi sentimen terbagi menjadi 2 kelas, yaitu positif dan negatif.
4. Nilai variabel *Learning Rate* antara 10^{-5} dan 10^{-6} .
5. Nilai *Epoch* antara 3 sampai 10.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti tata cara penulisan tugas akhir jurusan Teknik Informatika fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut.

BAB I. PENDAHULUAN

Pada bab ini membahas ide pokok yang melandasi penelitian, meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas dasar teori yang digunakan dalam penelitian. Pembahasan meliputi analisis sentimen, *deep learning* dan RoBERTa serta beberapa literatur terkait yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas mengenai tahapan-tahapan yang akan dilaksanakan dalam melakukan penelitian. Tahapan-tahapan yang dimaksud meliputi pengumpulan data, analisis data dan perancangan pembangunan sistem. Setiap

tahapan pada penelitian dibahas secara terperinci berdasarkan kerangka kerja yang dibuat.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas mengenai analisis dan rancangan perangkat lunak yang akan dibangun. Diawali dengan menganalisis kebutuhan, melakukan perancangan dan mengonstruksi perangkat lunak. Kemudian akan dilakukan mengenai pengujian sehingga hasil dari konstruksi perangkat lunak memenuhi kebutuhan dan sesuai dengan rancangan awal.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini membahas mengenai hasil pengujian sesuai yang sudah dirancang sebelumnya. Hasil penelitian disajikan dalam bentuk tabel dan analisis yang dijadikan dasar dalam pengambilan kesimpulan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini membahas mengenai kesimpulan yang ditarik dari semua uraian pada bab-bab sebelumnya. Bab ini juga menyajikan saran yang didapat setelah melakukan penelitian.

1.8 Kesimpulan

Pada bab ini telah dibahas mengenai ide pokok terkait penelitian yang akan dilakukan. Ide pokok tersebut meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian, serta batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

- Adoma, A. F., Henry, N. M., & Chen, W. (2020). Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*, 117–121. <https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379>
- Agarwal, B., Mittal, N., Bansal, P., & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience, 2015*. <https://doi.org/10.1155/2015/715730>
- Aloufi, S., & Saddik, A. El. (2018). Sentiment Identification in Football-Specific Tweets. *IEEE Access*, 6(December), 78609–78621. <https://doi.org/10.1109/ACCESS.2018.2885117>
- Anwar, A. (2014). A Review of RUP (Rational Unified Process). *International Journal of Software Engineering*, 5(2), 8–24. <http://www.cscjournals.org/library/manuscriptinfo.php?mc=IJSE-142>
- Batra, H., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2021). BERT-Based Sentiment Analysis: A Software Engineering Perspective. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12923 LNCS, 138–148. https://doi.org/10.1007/978-3-030-86472-9_13
- Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *Unified Modeling Language*

User Guide, The Second Edition.

European Commission. (2002). Commission closes investigations into FIFA regulations on international football transfers. Brussels . Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/IP_02_824

Deloitte. (2021). Testing times - Football Money League. In *Deloitte Sports Business Group* (Issue January).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1*(Mlm), 4171–4186.

Ghasiya, P., & Okamura, K. (2021). Investigating COVID-19 News across Four Nations: A Topic Modeling and Sentiment Analysis Approach. *IEEE Access*, 9, 36645–36656. <https://doi.org/10.1109/ACCESS.2021.3062875>

Gong, X. R., Jin, J. X., & Zhang, T. (2019). Sentiment Analysis Using Autoregressive Language Modeling and Broad Learning System. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, 1130–1134. <https://doi.org/10.1109/BIBM47256.2019.8983025>

Hoey, S., Peeters, T., & Principe, F. (2021). The transfer system in European football: A pro-competitive no-poaching agreement? *International Journal of Industrial Organization*, 75. <https://doi.org/10.1016/j.ijindorg.2020.102695>

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges.

- Journal of King Saud University - Engineering Sciences*, 30(4), 330–338.
<https://doi.org/10.1016/j.jksues.2016.04.002>
- Kotelnikova, A., Paschenko, D., Bochenina, K., & Kotelnikov, E. (2021). *Lexicon-based Methods vs. BERT for Text Sentiment Analysis*. 95.
<http://arxiv.org/abs/2111.10097>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 1. <http://arxiv.org/abs/1907.11692>
- Narayanaswamy, G. R. (2021). *Exploiting BERT and RoBERTa to Improve Performance for Aspect Based Sentiment Analysis* Gagan Reddy Narayanaswamy. <https://doi.org/10.21427/3w9n-we77>
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1, 2227–2237.
<https://doi.org/10.18653/v1/n18-1202>
- Petríková, D., & Soroková, T. (2015). Economic and Psychological Aspects of Importance of Real Madrid Brand in Marketing Communication. *Mediterranean Journal of Social Sciences*, 6(3), 504–514.
<https://doi.org/10.5901/mjss.2015.v6n3p504>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*. 1. <https://s3-us-west->

2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

- Rajpurkar, P., Jia, R., & Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2, 784–789. <https://doi.org/10.18653/v1/p18-2124>
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuad: 100,000+ questions for machine comprehension of text. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings, ii*, 2383–2392. <https://doi.org/10.18653/v1/d16-1264>
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics*, 8, 842–866. https://doi.org/10.1162/tacl_a_00349
- Rohde, M., & Breuer, C. (2016). Europe's elite football: Financial growth, sporting success, transfer investment, and private majority investors. *International Journal of Financial Studies*, 4(2), 10–12. <https://doi.org/10.3390/ijfs4020012>
- Sarsam, S. M., Al-Samarraie, H., Bahar, N., Shibghatullah, A. S., Eldenfria, A., & Al-Sa'Di, A. (2021). Detecting Real-Time Correlated Simultaneous Events in Microblogs: The Case of Men's Olympic Football. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12789 LNCS (Issue July)*. Springer International Publishing. <https://doi.org/10.1007/978-3-030->

- Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1631–1642.
- Stephen, V. S. (1997). Selecting and Interpreting Measure of Thematic Classification Accuracy. *Remote Sensing of Environment*, 62(1), 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Williams, A., Nangia, N., & Bowman, S. R. (2018). A broad-coverage challenge corpus for sentence understanding through inference. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1*, 1112–1122. <https://doi.org/10.18653/v1/n18-1101>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. Von, Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., Drame, M., ... Rush, A. M. (2020). *Transformers : State-of-the-Art Natural Language Processing*. 38–45.
- Wunderlich, F., & Memmert, D. (2020). Innovative approaches in sports science-

Lexicon-based sentiment analysis as a tool to analyze sports-related twitter communication. *Applied Sciences (Switzerland)*, 10(2).
<https://doi.org/10.3390/app10020431>

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32(NeurIPS), 1–18.

Zhang, X., Zhao, J., & Lecun, Y. (2015). Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems, 2015-Janua*, 649–657.

Zhao, L., Li, L., Zheng, X., & Zhang, J. (2021). A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts. *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021*, 1233–1238.
<https://doi.org/10.1109/CSCWD49262.2021.9437616>