

**ANALISIS SENTIMEN REVIEW PENGGUNA TERHADAP
APLIKASI RUANG GURU PADA PLAY STORE
MENGUNAKAN ALGORITMA BERT DENGAN METODE
NLP**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ROMANSA YUDHARISTYA

20.11.3395

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN REVIEW PENGGUNA TERHADAP
APLIKASI RUANG GURU PADA PLAY STORE
MENGUNAKAN ALGORITMA BERT DENGAN METODE
NLP**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ROMANSA YUDHARISTYA

20.11.3395

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN REVIEW PENGGUNA TERHADAP
APLIKASI RUANG GURU PADA PLAY STORE
MENGUNAKAN ALGORITMA BERT DENGAN METODE
NLP**

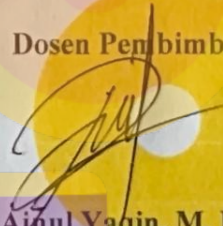
yang disusun dan diajukan oleh

Romansa Yudharistya

20.11.3395

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Agustus 2025

Dosen Pembimbing,



Ajiul Yaqin, M. Kom

NIK. 190302255

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN REVIEW PENGGUNA TERHADAP
APLIKASI RUANG GURU PADA PLAY STORE
MENGUNAKAN ALGORITMA BERT DENGAN METODE
NLP**

yang disusun dan diajukan oleh

Romansa Yudharistya

20.11.3395

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

Agung Nugroho, S.Kom., M.Kom.
NIK. 190302242

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M. Kom,
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Romansa Yudharistya
NIM : 20.11.3395

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Ainul Yaqin, M. Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, <11 Agustus 2025>

Yang Menyatakan,



Romansa Yudharistya

HALAMAN PERSEMBAHAN

Saya mengucapkan puji dan syukur kepada Allah Subhanahu wa Ta'ala atas segala nikmat-Nya yang telah menyertai saya dalam menyelesaikan tugas akhir yang berjudul “ANALISIS SENTIMEN REVIEW PENGGUNA TERHADAP APLIKASI RUANG GURU PADA PLAY STORE MENGGUNAKAN ALGORITMA BERT DENGAN METODE NLP”. Dengan rasa syukur dan kerendahan hati, saya mempersembahkan tugas akhir ini kepada:

1. Kedua orang tua tercinta, yang selalu mendoakan dan mendukung saya.
2. Bapak Ainul Yaqin, M. Kom sebagai pembimbing dan pengajar yang sangat baik. Terima kasih atas ilmu, waktu, dan dorongan yang telah diberikan selama pengerjaan skripsi.
3. Seluruh teman-teman IF 02 Informatika yang selalu memberikan dukungan dan masukan.

KATA PENGANTAR

Segala puji dan syukur kita panjatkan kepada Allah SWT atas berkat dan rahmat-Nya. Berkat karunia-Nya, penulis berhasil menyelesaikan skripsi yang berjudul “Analisis Sentimen Review Pengguna Terhadap Aplikasi Ruang Guru pada Play Store Menggunakan Algoritma BERT dengan Metode NLP.

Penulis menyadari bahwa banyak pihak yang turut membantu dan mendukung dalam penulisan skripsi ini. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

- Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
- Andika Agus Slameto, S. Kom. selaku dosen wali.
- Bapak Ainul Yaqin, M. Kom. selaku dosen pembimbing
- Seluruh teman-teman yang telah mendukung dan membantu penelitian ini.

Akhir kata, penulis berharap dengan penelitian ini dapat memberikan manfaat besar bagi para pembaca dan peneliti. Semoga penelitian ini dapat terus dikembangkan dan memberikan manfaat bagi seluruh masyarakat.

Yogyakarta, 11 Agustus 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12

2.2.1	NLP	12
2.2.2	Analisis Sentimen	12
2.2.3	<i>Deep Learning</i>	13
2.2.4	Google Play	13
2.2.5	Algoritma BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	13
2.2.5.1	<i>Framework</i>	15
2.2.5.2	Cara Kerja BERT	15
2.2.6	Evaluasi Model	16
2.2.7	<i>Confusion Matrix</i>	16
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian	19
3.2	Alur Penelitian	19
3.2.1	<i>Dataset</i>	20
3.2.2	<i>Data Preprocessing</i>	21
3.2.2.1	<i>Case Folding</i>	22
3.2.2.2	<i>Data Cleaning</i>	22
3.2.2.3	Normalisasi	23
3.2.2.4	<i>Stopword Removal</i>	24
3.2.2.5	<i>Tokenizing</i>	25
3.2.2.6	<i>Stemming</i>	26
3.2.2.7	<i>Labeling</i>	27
3.2.3	<i>Balancing</i>	28
3.2.4	<i>EDA (Exploratory Data Analyst)</i>	29
3.2.4.1	Persentase	29

3.2.4.2	Visualisasi <i>WordCloud</i>	30
3.2.5	<i>Data Preparation</i>	30
3.2.5.1	<i>Tokenizing dan Encoding BERT</i>	30
3.2.5.2	<i>Data Spliting</i>	30
3.2.5.3	<i>Creating Batch Data using DataLoader</i>	31
3.2.6	<i>Modeling BERT</i>	31
3.2.6.1	<i>Pretrained Model</i>	32
3.2.6.2	<i>Optimizer dan Scheduler</i>	32
3.2.7	<i>Prediction and Evaluation</i>	32
3.3	Alat dan Bahan	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data	34
4.2	<i>Processing Dataset</i>	36
4.2.1	<i>Case Folding</i>	36
4.2.2	<i>Data Cleaning</i>	37
4.2.3	Tokenisasi	38
4.2.4	Normalisasi	39
4.2.5	<i>Stopwords</i>	40
4.2.6	<i>Stemming</i>	41
4.2.7	<i>Labeling</i>	42
4.3	<i>Balancing</i>	43
4.4	EDA	44
4.4.1	Visualisasi <i>Persentase</i>	45
4.4.2	Visualisasi <i>WordCloud</i>	45
4.4.2.1	Visualisasi Positif	45

4.4.2.2	Visualisasi Negatif.....	46
4.4.2.3	Visualisasi Netral	46
4.5	<i>Data Preparation</i>	47
4.5.1	<i>Tokenizing dan Encoding BERT</i>	47
4.5.2	Data Splitting	48
4.5.3	<i>Data Loader</i>	49
4.6	<i>Modeling BERT</i>	49
4.6.1	<i>Pretrained BERT</i>	49
4.6.2	<i>Optimizer dan Scheduler</i>	50
4.6.3	<i>Training BERT</i>	51
4.7	Prediksi dan Evaluasi	52
BAB V PENUTUP		54
5.1	Kesimpulan	54
5.2	Saran	54
REFERENSI		55

DAFTAR TABEL

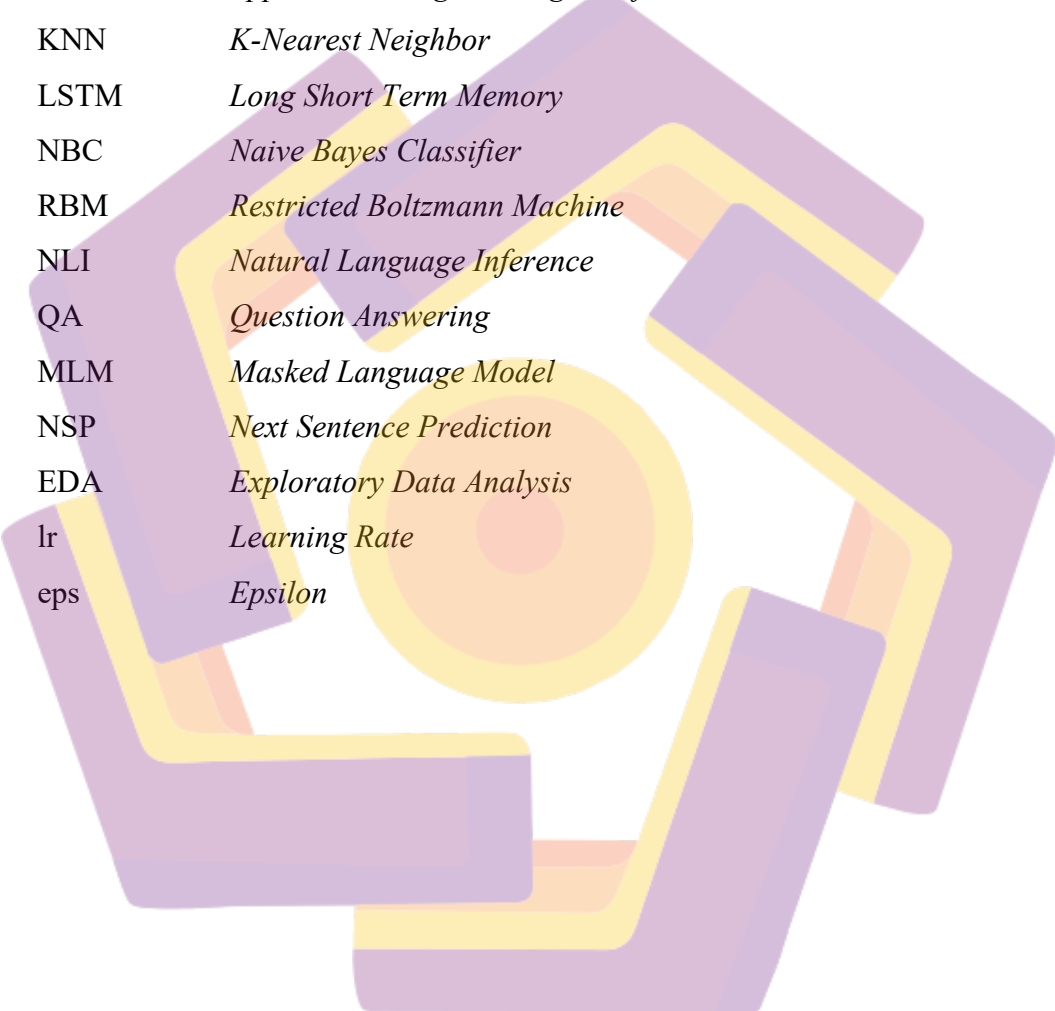
Tabel 2.1. Keaslian Penelitian	8
Tabel 2.2. <i>Confusion Matrix</i>	17
Tabel 3.1. Hasil memuat <i>dataset</i>	20
Tabel 4.1. Hasil <i>Crawling</i>	34
Tabel 4.2. Hasil <i>Case Folding</i>	36
Tabel 4.3. Hasil <i>Data Cleanning</i>	37
Tabel 4.4. Hasil Tokenisasi	38
Tabel 4.5. Hasil Normalisasi	39
Tabel 4.6. Hasil <i>Stopwords</i>	41
Tabel 4.7. Hasil <i>Stemming</i>	42
Tabel 4.8. Hasil <i>Labeling</i>	43
Tabel 4.9. Hasil <i>Data Splitting</i>	48
Tabel 4.10. Hasil <i>Confusion Matrix</i>	53



DAFTAR GAMBAR

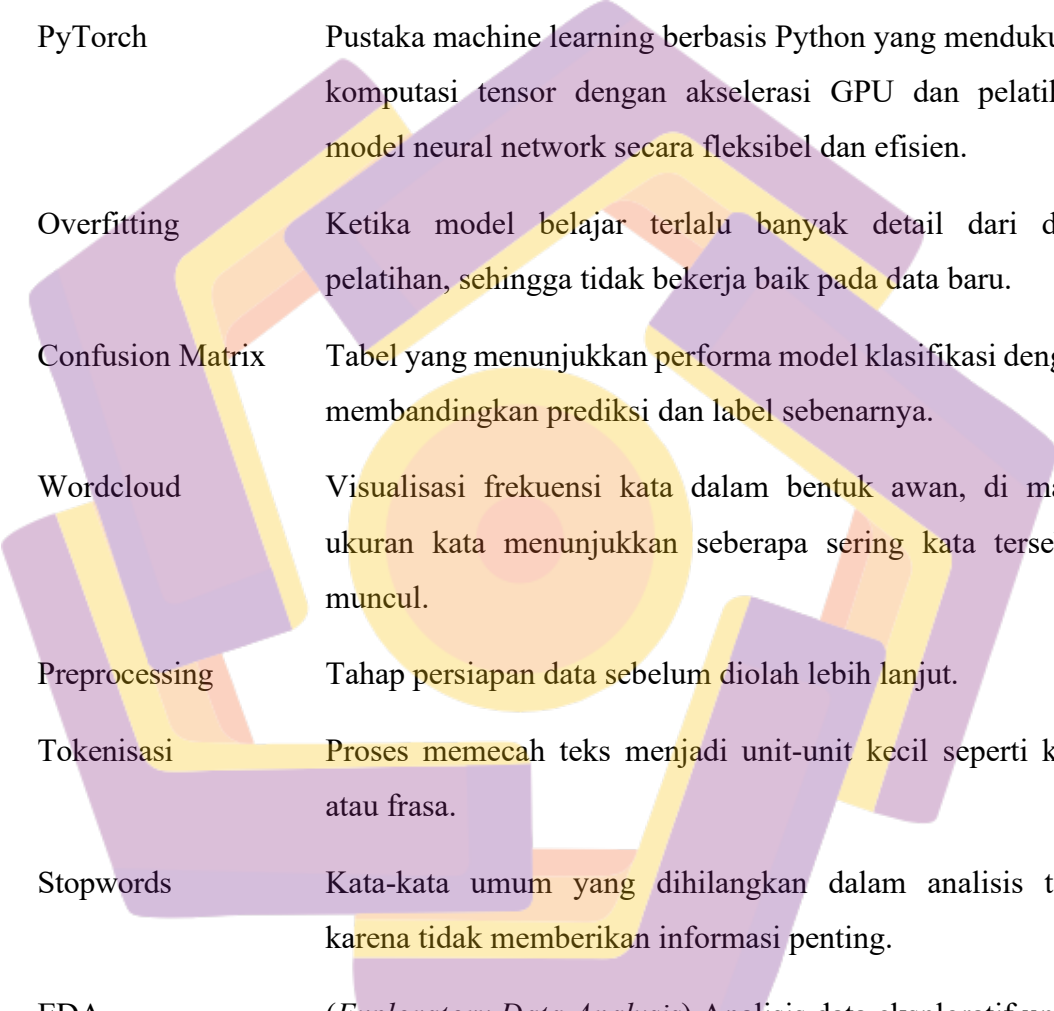
Gambar 2.1. BERT <i>Architecture Diagram</i>	14
Gambar 3.1. Alur Penelitian	19
Gambar 3.2. Alur <i>Case Folding</i>	22
Gambar 3.3. Alur <i>Data Cleaning</i>	23
Gambar 3.4. Alur Normalisasi	24
Gambar 3.5. Alur <i>Stopword Removal</i>	25
Gambar 3.6. Alur <i>Tokenizing</i>	26
Gambar 3.7. Alur <i>Stemming</i>	27
Gambar 3.8. Alur <i>Labeling</i>	28
Gambar 3.9. Alur <i>Balancing</i>	29
Gambar 4.1. Hasil <i>Balancing</i>	44
Gambar 4.2. Hasil Distribusi Sentimen	44
Gambar 4.3. Hasil Visualisasi Persentase	45
Gambar 4.4. Hasil Visualisasi Positif	46
Gambar 4.5. Hasil Visualisasi Negatif	46
Gambar 4.6. Hasil Visualisasi Netral	47
Gambar 4.7. Hasil Pengecekan Panjang Teks	48
Gambar 4.8. Hasil <i>Tokenizing</i> dan <i>Encoding BERT</i>	48
Gambar 4.9. Hasil <i>Pretrained</i> dan Parameter	50
Gambar 4.10. Visualisasi <i>Training</i> , <i>Validation Acc</i> dan <i>Loss</i>	51
Gambar 4.11. Hasil <i>Confusion Matrix</i>	52

DAFTAR LAMBANG DAN SINGKATAN



NLP	<i>Natural Language Processing</i>
CoC	<i>Class of Champions</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
API	<i>Application Programming Interface</i>
KNN	<i>K-Nearest Neighbor</i>
LSTM	<i>Long Short Term Memory</i>
NBC	<i>Naive Bayes Classifier</i>
RBM	<i>Restricted Boltzmann Machine</i>
NLI	<i>Natural Language Inference</i>
QA	<i>Question Answering</i>
MLM	<i>Masked Language Model</i>
NSP	<i>Next Sentence Prediction</i>
EDA	<i>Exploratory Data Analysis</i>
lr	<i>Learning Rate</i>
eps	<i>Epsilon</i>

DAFTAR ISTILAH



Epoch	Satu putaran penuh melalui seluruh dataset selama pelatihan model.
Adam	Optimizer yang digunakan dalam pelatihan jaringan saraf.
PyTorch	Pustaka machine learning berbasis Python yang mendukung komputasi tensor dengan akselerasi GPU dan pelatihan model neural network secara fleksibel dan efisien.
Overfitting	Ketika model belajar terlalu banyak detail dari data pelatihan, sehingga tidak bekerja baik pada data baru.
Confusion Matrix	Tabel yang menunjukkan performa model klasifikasi dengan membandingkan prediksi dan label sebenarnya.
Wordcloud	Visualisasi frekuensi kata dalam bentuk awan, di mana ukuran kata menunjukkan seberapa sering kata tersebut muncul.
Preprocessing	Tahap persiapan data sebelum diolah lebih lanjut.
Tokenisasi	Proses memecah teks menjadi unit-unit kecil seperti kata atau frasa.
Stopwords	Kata-kata umum yang dihilangkan dalam analisis teks karena tidak memberikan informasi penting.
EDA	(<i>Exploratory Data Analysis</i>) Analisis data eksploratif untuk menemukan pola, anomali, atau asumsi.
Sentiment Analysis	Analisis untuk mengidentifikasi dan mengekstraksi informasi subjektif dari teks. Analisis untuk mengidentifikasi dan mengekstraksi informasi subjektif dari teks.

INTISARI

Ruangguru merupakan perusahaan teknologi terbesar di Indonesia yang berfokus pada layanan berbasis pendidikan. Ruangguru telah memiliki lebih dari 22.000.000 pengguna baik didalam App Store maupun Play Store serta mengelola lebih dari 300.000 guru yang menawarkan jasa lebih dari 100 bidang pelajaran. Ruangguru telah dipercaya untuk bermitra dengan 32 Pemerintah Provinsi dan 326 Pemerintah Kota dan Kabupaten di Indonesia. Sehubungan dengan kepopuleran aplikasi Ruangguru beragam reaksi dan tanggapan positif maupun negatif dari pengguna(masyarakat). Tanggapan ini pun dilontarkan baik didalam situs Ruangguru dan Play Store atau App Store. Maka dari itu perlu dilakukan analisis customer review pada aplikasi Ruangguru pada Play Store untuk mengukur bagaimana respon masyarakat terhadap Ruangguru. Algoritma untuk penelitian ini menggunakan BERT dengan metode NLP, dengan klasifikasi sentimen positif, netral, dan negatif. Dengan dataset awal sebanyak 5001 data dan setelah di balancing menjadi 8493 data, rasio pembagian 80:10:10 untuk latih, validasi, dan uji, serta *hyperparameter* batch size 32, *optimizer* AdamW dengan *learning rate* $2e-6$, dan 20 epoch. Hasil penelitian ini dengan model BERT menghasilkan akurasi sebesar 91%. Model BERT dalam mengklasifikasikan menunjukkan *precision*, *recall*, dan *F1-score* rata-rata makro sebesar 0.91. Kinerja untuk kelas negatif dan netral kuat, dengan *precision* dan *recall* masing-masing di atas 88%, sementara kelas positif memiliki *precision tinggi* (94%) tetapi *recall* lebih rendah (85%), menunjukkan beberapa sampel positif tidak terdeteksi.

Kata kunci: Ruangguru, BERT, Klasifikasi Sentimen, Customer Review, NLP

ABSTRACT

Ruangguru is the largest technology company in Indonesia focusing on education-based services. Ruangguru has more than 22,000,000 users on both the App Store and Play Store, and manages over 300,000 teachers offering services in more than 100 subject areas. Ruangguru has been trusted to partner with 32 Provincial Governments and 326 City and District Governments in Indonesia. Due to the popularity of the Ruangguru application, there are various positive and negative reactions and feedback from users (the public). These responses are expressed both on the Ruangguru site and on the Play Store or App Store. Therefore, it is necessary to conduct customer review analysis of the Ruangguru application on the Play Store to measure how the public responds to Ruangguru. The algorithm used for this research is BERT with the NLP method, classifying sentiments into positive, neutral, and negative. With an initial dataset of 5,001 data points, balanced to 8,493 data points, an 80:10:10 split was applied for training, validation, and testing, with hyperparameters batch size 32, AdamW optimizer with a learning rate of $2e-6$, and 20 epochs. The research results using the BERT model achieved an accuracy of 91%. The BERT model in classification showed a macro-average precision, recall, and F1-score of 0.91. Performance for the negative and neutral classes was strong, with precision and recall both above 88%, while the positive class achieved high precision (94%) but lower recall (85%), indicating that some positive samples were not detected.

Keyword: *Ruangguru, BERT, Sentiment Classification, Customer Review, NLP*