

**ANALISIS SENTIMEN KELUHAN MASYARAKAT TERHADAP
JARINGAN INTERNET MENGGUNAKAN METODE XLNET
DAN ROBERTA DI KABUPATEN BANGGAI KEPULAUAN**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi *SI Informatika*



disusun oleh

NUR FADHLUR RAHMAN

21.11.4349

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN KELUHAN MASYARAKAT TERHADAP
JARINGAN INTERNET MENGGUNAKAN METODE XLNET
DAN ROBERTA DI KABUPATEN BANGGAI KEPULAUAN**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI Informatika*



disusun oleh

NUR FADHLUR RAHMAN

21.11.4349

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KELUHAN MASYARAKAT TERHADAP
JARINGAN INTERNET MENGGUNAKAN METODE XLNET
DAN ROBERTA DI KABUPATEN BANGGAI KEPULAUAN**

yang disusun dan diajukan oleh

NUR FADHLUR RAHMAN

21.11.4349

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 5 Desember 2025

Dosen Pembimbing,



Akhmad Dahlan, M.Kom

NIK. 190302174

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN KELUHAN MASYARAKAT TERHADAP
JARINGAN INTERNET MENGGUNAKAN METODE XLNET
DAN ROBERTA DI KABUPATEN BANGGAI KEPULAUAN

yang disusun dan diajukan oleh

NUR FADHLUR RAHMAN

21.11.4349

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Hartatik, S.T., M.Cs.

NIK. 190302232

Afrig Aminuddin, S.Kom., M.Eng., Ph.D.

NIK. 190302351

Yoga Pristyanto, S.Kom., M.Eng.

NIK. 190302412

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Desember

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nur Fadhlur Rahman
NIM : 21.11.4349

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS SENTIMEN KELUHAN MASYARAKAT TERHADAP JARINGAN INTERNET MENGGUNAKAN METODE XLNET DAN ROBERTA DI KABUPATEN BANGGAI KEPULAUAN

Dosen Pembimbing : Akhman Dahlan, 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, 19 Desember 2025

Yang Menyatakan,



The image shows a handwritten signature in black ink over a yellow revenue stamp. The stamp features the Garuda Pancasila emblem, the text 'SEPAJUTU RIBU RUPIAH', '10000', 'MEYERAI TEMPEL', and the serial number '504ANX142157344'.

Nur Fadhlur Rahman

HALAMAN PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan Penuh rasa Syukur Kepada Allah SWT, Skripsi ini saya persembahkan dengan penuh rasa syukur dan cinta kepada:

Ayah, Ibu dan Keluarga Besar

Untuk cinta, doa, dan dukungan tanpa henti yang telah diberikan sepanjang perjalanan hidup saya.

Teman-teman dan Sahabat

Untuk semangat dan kebersamaan yang tak ternilai, yang selalu menemani di setiap langkah.

Dosen pembimbing

Untuk bimbingan, arahan, dan ilmu yang telah diberikan dengan penuh kesabaran dan dedikasi dalam penyusunan skripsi ini.

Almamater tercinta Universitas Amikom Yogyakarta

Untuk ilmu, kesempatan, dan pengalaman yang telah diberikan selama menempuh pendidikan.

Dinas Komunikasi dan Informatika (Kominfo)

Untuk dukungan dan kesempatan yang diberikan selama proses penelitian ini. Tanpa bantuan dari Dinas Kominfo, penelitian ini tidak akan dapat terlaksana dengan baik.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Yang Maha Esa atas limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Sentimen Keluhan Masyarakat Terhadap Jaringan Internet Menggunakan Metode XLNet dan RoBERTa di Kabupaten Banggai Kepulauan". Skripsi ini disusun guna memenuhi salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika di Universitas Amikom Yogyakarta.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Akhmad Dahlan, M.Kom., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji yang telah memberikan masukan dan saran yang sangat berharga untuk perbaikan skripsi ini.
3. Orang tua dan keluarga besar saya yang selalu memberikan dukungan moril dan materil serta doa yang tiada henti.
4. Universitas Amikom Yogyakarta yang telah memberikan kesempatan dan dukungan dalam penelitian ini.
5. Dinas Komunikasi dan Informatika (Kominfo) Kabupaten Banggai Kepulauan yang telah memberikan izin serta dukungan penuh selama proses penelitian ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak untuk perbaikan di masa mendatang.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang berkepentingan.

Yogyakarta, 5 Desember 2025

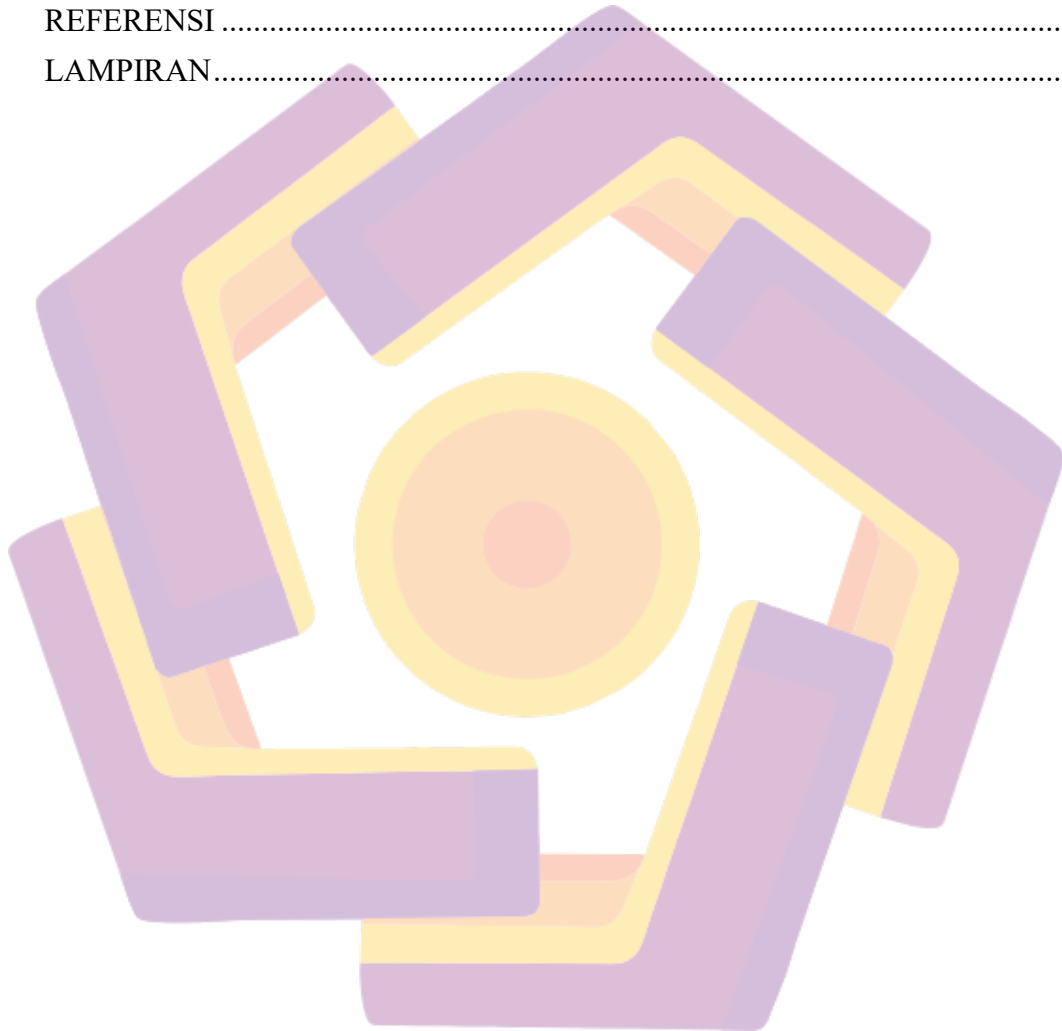
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xv
INTISARI	xviii
<i>ABSTRACT</i>	xix
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
1.5.1 Manfaat Segi Teoritis Dan Praktis.....	3
1.5.2 Manfaat Bagi Penelitian Selanjutnya.....	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	10
2.2.1 Machine Learning	10
2.2.2 Bahasa Alami (NLP).....	10
2.2.3 Sentiment Analysis	11
2.2.4 Text Mining	11
2.2.5 Text Preprocessing.....	12

2.2.6	Handling Unbalanced Data	13
2.2.7	XLNet (eXtreme MultiLingual Network).....	14
2.2.8	RoBERTa.....	15
2.2.9	Packages.....	16
2.2.9.1	Pandas	17
2.2.9.2	Numpy.....	17
2.2.9.3	Matplotlib.....	18
2.2.9.4	Seaborn	20
2.2.9.5	NLTK.....	20
2.2.9.6	Transformers (Hugging Face).....	21
2.2.9.7	Sklearn	22
2.2.9.8	Wordcloud.....	21
2.2.10	Confusion Matrix	23
BAB III METODE PENELITIAN		25
3.1	Objek Penelitian.....	26
3.2	Alur Penelitian	26
3.3	Alat dan Bahan.....	32
3.3.1	Alat.....	32
3.3.2	Bahan Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		35
4.1	Pengumpulan Data	35
4.2	Preprocessing Data	35
4.2.1	Cleaning Data.....	36
4.2.2	Lowercase	36
4.2.3	Tokenisasi	37
4.2.4	Normalisasi	37
4.3	Labeling Data.....	38
4.4	Handling Unbalanced Data	39
4.5	Splitting Data	41
4.6	Training Model Xlnet dan Roberta	41
4.6.1	Perhitungan Class Weights	42
4.6.2	Training Model XLNet	42
4.6.3	Training Model RoBERTa.....	43
4.7	Evaluasi Model	44

4.7.1 Evaluasi Model XLNet	45
4.7.2 Evaluasi Model RoBERTa.....	46
4.8 Perbandingan Hasil dan Kesimpulan	48
BAB V PENUTUP	50
5.1 Kesimpulan	50
5.2 Saran	51
REFERENSI	52
LAMPIRAN.....	58

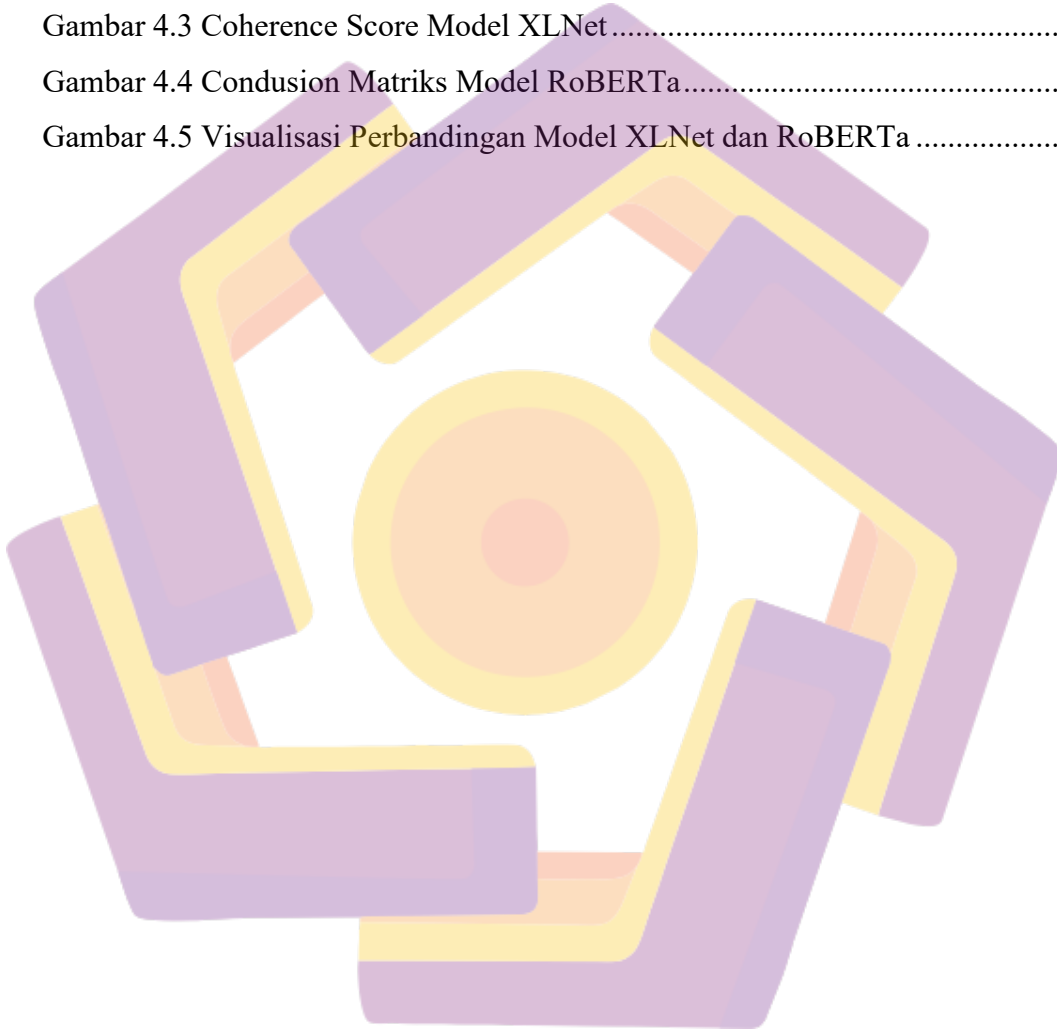


DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 2.2 Confusion Matrix	24
Tabel 3.1 Data penelitian	34
Tabel 4.1 Visualisasi Data Komentar	35
Tabel 4.2 Contoh Proses Cleaning Data	36
Tabel 4.3 Contoh Proses Lowercase	36
Tabel 4.4 Lanjutan Contoh Proses Lowercase	37
Tabel 4.5 Contoh Proses Tokenisasi	37
Tabel 4.6 Contoh Proses Normalisasi	38
Tabel 4.7 Hasil Pelabelan Otomatis	36
Tabel 4.8 Hasil Distribusi Data Awal	40
Tabel 4.9 Hasil Proses Splitting Data	41
Tabel 4.10 Hasil Proses Training Model XLNet	43
Tabel 4.11 Hasil Proses Training Model RoBERTa	44
Tabel 4.12 Matriks Evaluasi Model XLNet	45
Tabel 4.14 Matriks Evaluasi Model RoBERTa	47

DAFTAR GAMBAR

Gambar 2.1 Contoh Wordcloud	22
Gambar 3.1 Alur Penelitian	26
Gambar 4.1 Hasil Distribusi Data Setelah Balancing	40
Gambar 4.2 Hasil Perhitungan Class Weights	42
Gambar 4.3 Coherence Score Model XLNet	46
Gambar 4.4 Condusion Matriks Model RoBERTa	48
Gambar 4.5 Visualisasi Perbandingan Model XLNet dan RoBERTa	49



DAFTAR LAMPIRAN

Lampiran 1 Source Code	58
------------------------------	----

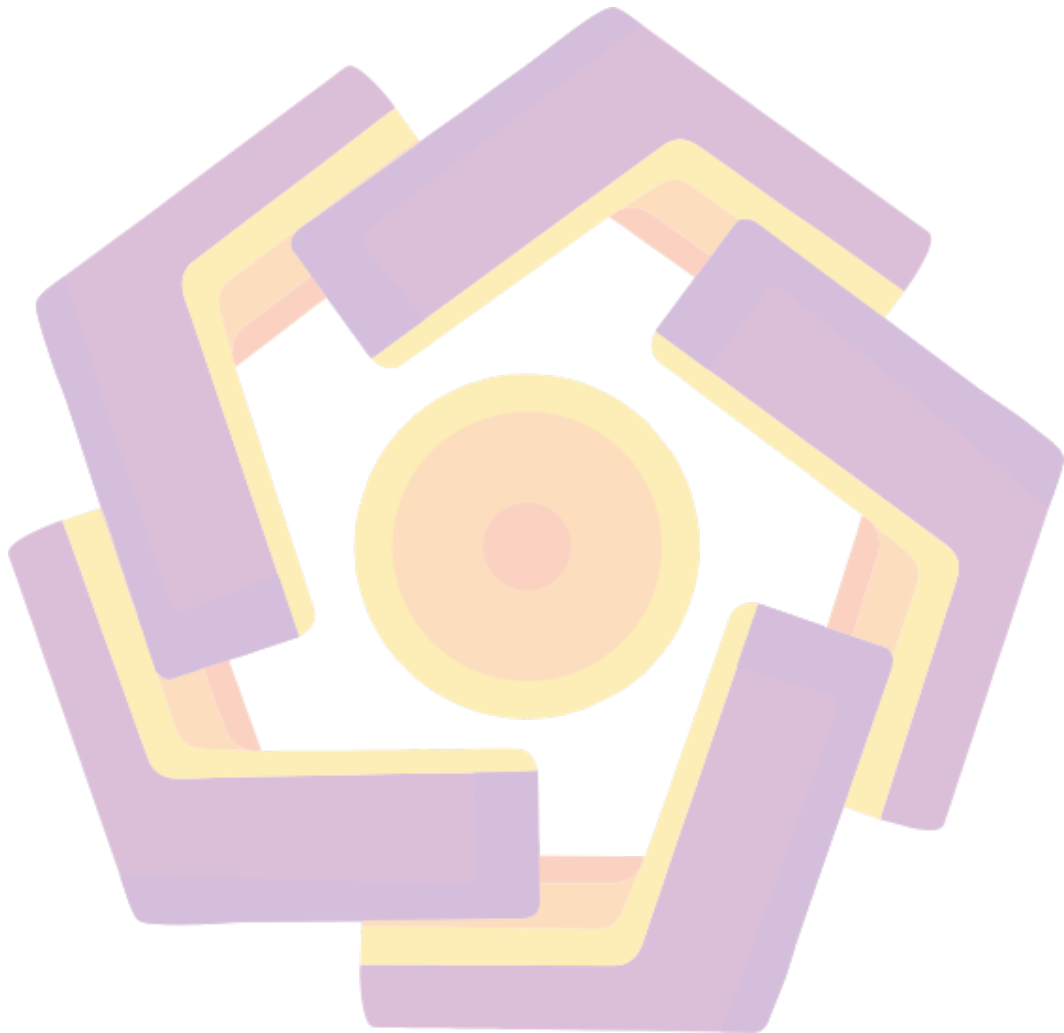


DAFTAR LAMBANG DAN SINGKATAN

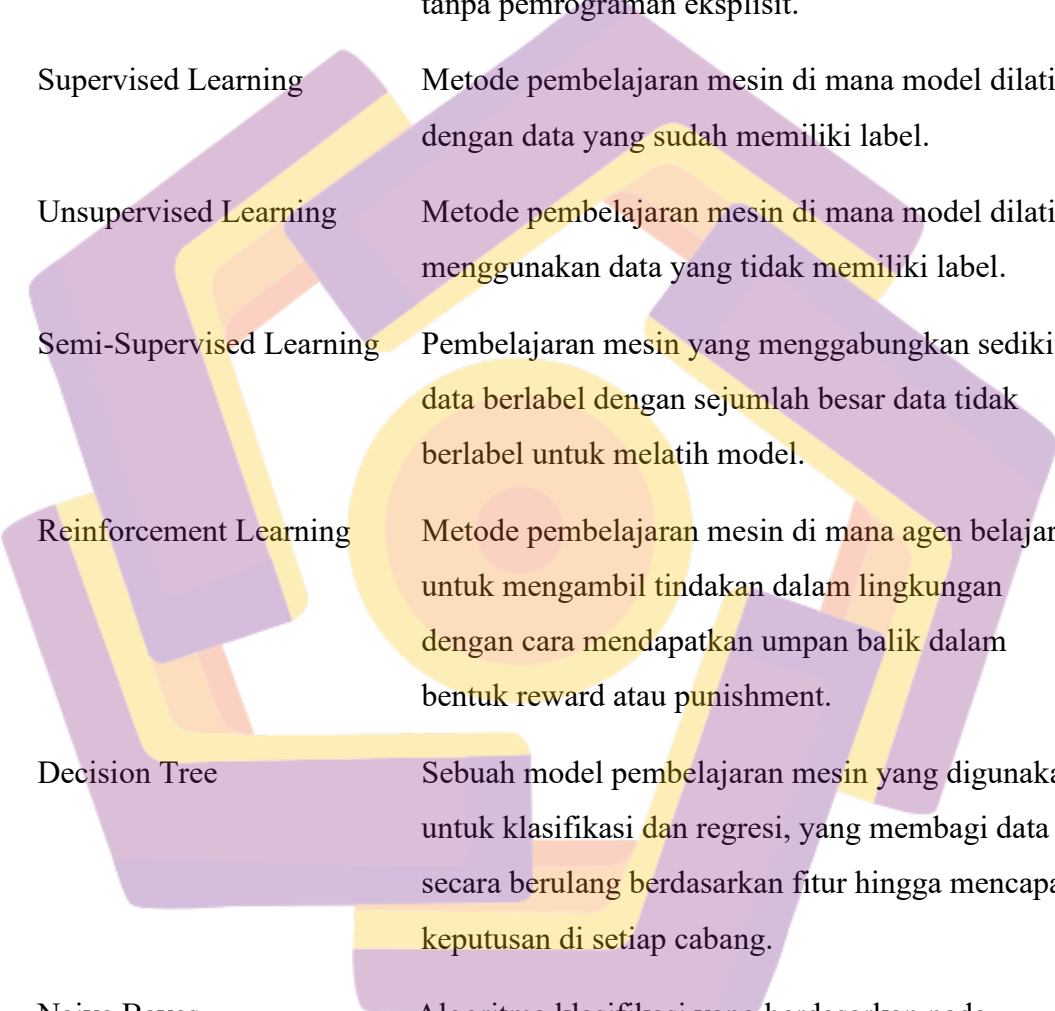


KNN	K-Nearest Neighbor
SVM	Support Vector Machine
NLP	Natural Language Processing
SMOTE	Synthetic Minority Over-sampling Technique
XLNet	eXtreme Language Network
RoBERTa	Robustly optimized BERT approach
NLTK	Natural Language Toolkit
BERT	Bidirectional Encoder Representations from Transformers
NSP	Next Sentence Prediction
MLM	Masked Language Modeling
PCA	Principal Component Analysis
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
t-SNE	t-Distributed Stochastic Neighbor Embedding
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ZT	Himpunan semua kemungkinan permutasi dari urutan panjang indeks T

$p\theta$	Merupakan fungsi yang digunakan dalam pemodelan bahasa permutasi untuk memprediksi token ke- t berdasarkan konteks token sebelumnya
z_t	Elemen ke- t

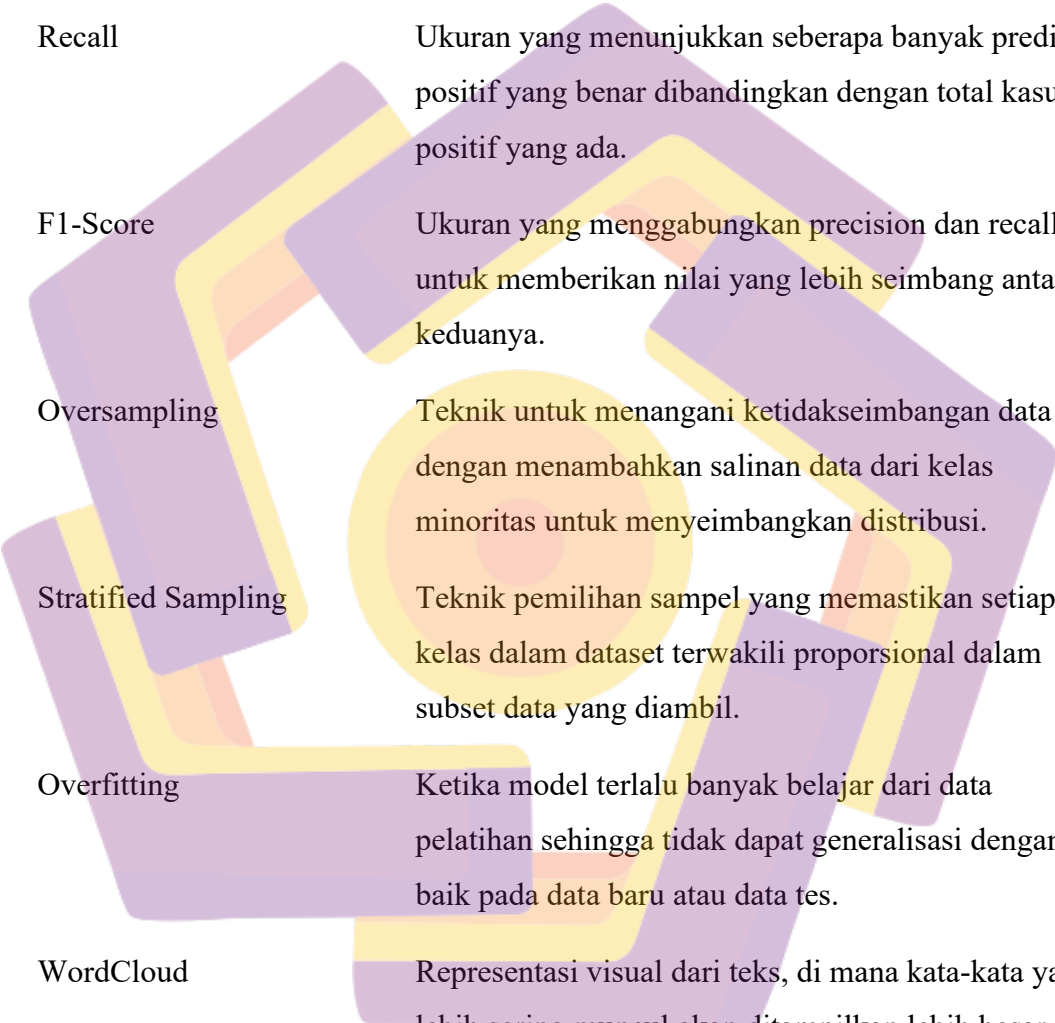


DAFTAR ISTILAH



Machine Learning	Sebuah cabang dari kecerdasan buatan (AI) yang mempelajari algoritma dan model statistik yang memungkinkan komputer untuk belajar dari data tanpa pemrograman eksplisit.
Supervised Learning	Metode pembelajaran mesin di mana model dilatih dengan data yang sudah memiliki label.
Unsupervised Learning	Metode pembelajaran mesin di mana model dilatih menggunakan data yang tidak memiliki label.
Semi-Supervised Learning	Pembelajaran mesin yang menggabungkan sedikit data berlabel dengan sejumlah besar data tidak berlabel untuk melatih model.
Reinforcement Learning	Metode pembelajaran mesin di mana agen belajar untuk mengambil tindakan dalam lingkungan dengan cara mendapatkan umpan balik dalam bentuk reward atau punishment.
Decision Tree	Sebuah model pembelajaran mesin yang digunakan untuk klasifikasi dan regresi, yang membagi data secara berulang berdasarkan fitur hingga mencapai keputusan di setiap cabang.
Naive Bayes	Algoritma klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi bahwa fitur-fitur bersifat independen.
Regresi	Teknik statistik yang digunakan untuk memprediksi nilai numerik berdasarkan hubungan antara variabel independen dan dependen.

Case Folding/Lowercase	Proses mengubah semua teks menjadi huruf kecil.
Cleaning Text	Proses pembersihan teks dengan menghapus karakter seperti angka, simbol, dan spasi ekstra.
Tokenization	Proses memecah teks menjadi unit-unit kecil (tokens) seperti kata atau frasa.
Stopword Removal	Menghapus kata-kata umum yang tidak memberikan informasi penting dalam analisis teks, seperti "dan", "atau", "yang".
Lemmatization	Proses mengubah kata ke bentuk dasar.
Normalization	Proses mengubah teks ke bentuk yang baku.
Unbalanced Data	Kondisi di mana distribusi kelas dalam dataset tidak seimbang.
Pandas	Perpustakaan Python yang digunakan untuk manipulasi dan analisis data.
NumPy	Perpustakaan Python untuk komputasi numerik yang mendukung array multidimensi.
Matplotlib	Perpustakaan Python untuk visualisasi data yang memungkinkan pembuatan grafik dan diagram.
Seaborn	Perpustakaan Python yang dibangun di atas Matplotlib untuk visualisasi data statistik
Scikit-learn	Perpustakaan Python yang menyediakan alat sederhana dan efisien untuk analisis data dan pemodelan statistik.
Confusion Matrix	Matriks yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menunjukkan jumlah prediksi yang benar dan salah berdasarkan kategori.



Accuracy	Ukuran yang menunjukkan seberapa sering model memberikan prediksi yang benar.
Precision	Ukuran yang menunjukkan seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif.
Recall	Ukuran yang menunjukkan seberapa banyak prediksi positif yang benar dibandingkan dengan total kasus positif yang ada.
F1-Score	Ukuran yang menggabungkan precision dan recall untuk memberikan nilai yang lebih seimbang antara keduanya.
Oversampling	Teknik untuk menangani ketidakseimbangan data dengan menambahkan salinan data dari kelas minoritas untuk menyeimbangkan distribusi.
Stratified Sampling	Teknik pemilihan sampel yang memastikan setiap kelas dalam dataset terwakili proporsional dalam subset data yang diambil.
Overfitting	Ketika model terlalu banyak belajar dari data pelatihan sehingga tidak dapat generalisasi dengan baik pada data baru atau data tes.
WordCloud	Representasi visual dari teks, di mana kata-kata yang lebih sering muncul akan ditampilkan lebih besar dan lebih menonjol, sementara kata-kata yang jarang muncul akan tampil lebih kecil.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen keluhan masyarakat terhadap kualitas jaringan internet di Kabupaten Banggai Kepulauan dengan menggunakan teknik analisis sentimen berbasis pembelajaran mesin. Data yang digunakan dalam penelitian ini terdiri dari 646 sampel komentar yang dikumpulkan melalui media sosial WhatsApp dan kuesioner Google Form.

Proses *preprocessing* data dilakukan untuk mempersiapkan data sebelum digunakan oleh model, yang mencakup *pembersihan* data dari karakter-karakter yang tidak *relevan*, normalisasi kata gaul ke dalam bentuk kata baku, serta tokenisasi untuk memecah teks menjadi unit yang lebih kecil. Selanjutnya, data yang telah diproses dilabeli menggunakan dua model pembelajaran mesin, yaitu XLNet dan RoBERTa, untuk mengklasifikasikan sentimen komentar ke dalam tiga kategori: positif, netral, dan negatif. Selain itu, untuk mengatasi ketidakseimbangan data, penelitian ini juga menerapkan teknik *random oversampling* untuk memastikan distribusi yang lebih seimbang antar kelas sentimen.

Hasil penelitian menunjukkan bahwa model RoBERTa memberikan hasil yang lebih baik dibandingkan XLNet dalam hal akurasi dan F1-score. RoBERTa mencapai akurasi 87,77% dengan F1-score 87,75%, sementara XLNet mencapai akurasi 83,45% dan F1-score 83,37%. Dalam hal efisiensi, waktu pelatihan RoBERTa juga lebih singkat, yaitu sekitar 44,3 menit, dibandingkan XLNet yang memerlukan sekitar 76,1 menit. Analisis sentimen menunjukkan bahwa komentar negatif mendominasi, mencerminkan keluhan masyarakat terhadap kualitas jaringan internet di Kabupaten Banggai Kepulauan yang banyak terkait dengan masalah seperti lambatnya jaringan dan hilangnya koneksi. Hasil penelitian ini diharapkan dapat menjadi wawasan bagi pemerintah daerah untuk memperbaiki kebijakan dan pengembangan infrastruktur jaringan di wilayah tersebut.

Kata kunci: analisis sentimen, XLNet, RoBERTa, preprocessing, oversampling.

ABSTRACT

This study aims to analyze the sentiment of public complaints regarding the quality of internet network in the Banggai Kepulauan Regency using machine learning-based sentiment analysis techniques. The data used in this research consists of 646 sample comments collected from WhatsApp social media and Google Form surveys.

The data preprocessing process was carried out to prepare the data before being used by the model, which includes cleaning irrelevant characters, normalizing slang words into standard terms, and tokenization to break down the text into smaller units. The processed data was then labeled using two machine learning models, XLNet and RoBERTa, which were used to classify the sentiment of the comments into three categories: positive, neutral, and negative. Furthermore, to address the issue of data imbalance, this study also applies the Random Oversampling technique to ensure a more balanced distribution of sentiment classes.

The results show that the RoBERTa model outperforms XLNet in terms of accuracy and F1-score. RoBERTa achieved an accuracy of 87.77% with an F1-score of 87.75%, while XLNet achieved an accuracy of 83.45% and an F1-score of 83.37%. In terms of efficiency, the training time of RoBERTa is also shorter, around 44.3 minutes, compared to XLNet, which requires approximately 76.1 minutes. Sentiment analysis indicates that negative comments dominate, reflecting the public's complaints about the internet network quality in the Banggai Kepulauan Regency, which are mainly related to issues such as slow network speeds and connection losses. The findings of this study are expected to provide insights for the local government to improve policies and infrastructure development in the region.

Keywords: *Sentiment analysis, XLNet, RoBERTa, preprocessing, oversampling*