



Universitas Hasanuddin

Public Complaint

Kelompok 5



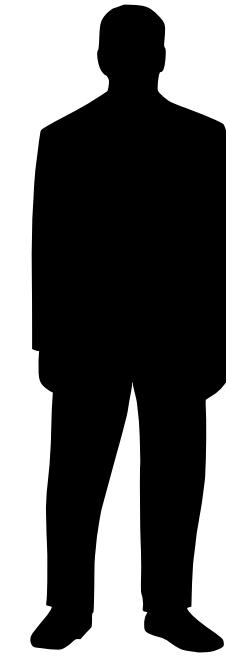


Universitas Hasanuddin

ANGGOTA KELOMPOK



AYU WIDIANTI
H071221056



ANDI ADNAN
H071221078



MUH.YUSUF FIKRY
H071221102



Topics

- Background 01
- Data Collection 02
- Exploratory Data Analysis 03
- Pre-Processing 04
- Pembuatan Model 05
- Evaluasi & Hasil Analysis 06
- Kesimpulan 07





Project Timeline

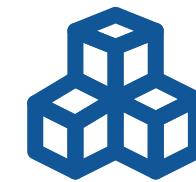
No	Fase	Jadwal						
		P9	P10	P11	P12	P13	P14	P15
1	Project Planning							
2	Data Collection							
3	Exploratory Data Analysis							
4	Pre-Processing							
5	Pembuatan Model							
6	Analysis and Documentation							
7	Presentation							



BACKGROUND



PROBLEM



Pengaduan publik sebagai bagian yang sangat penting dalam kehidupan bernegara



Tantangan Klasifikasi Pengaduan yang volumenya besar dan beragam



Input kategori manual pada form pengisian laporan pada sistem informasi pengaduan masyarakat saat ini.



GOALS

1

-• **Rilis dataset teks pengaduan publik**

2

-• **Model klasifikasi pengaduan publik dengan performa yang baik**

3

-• **Menambah pengetahuan mengenai implementasi metode deep learning**



ABOUT DATASET

Dataset ini terdiri dari 9 kategori pengaduan masyarakat yang mencakup berbagai aspek penting dalam kehidupan sehari-hari.

- Infrastruktur & Fasilitas Umum
- Keadilan Hukum dan HAM
- Kepegawaian
- Kesehatan
- Ketenagakerjaan dan Kesejahteraan Sosial
- Lingkungan Hidup
- Pendidikan
- Pertanahan dan Perizinan
- Teknologi Informasi dan Komunikasi





Data Collection, Exploratory Data Analysis, & Preprocessing

Progres Pengumpulan Data

Total Data

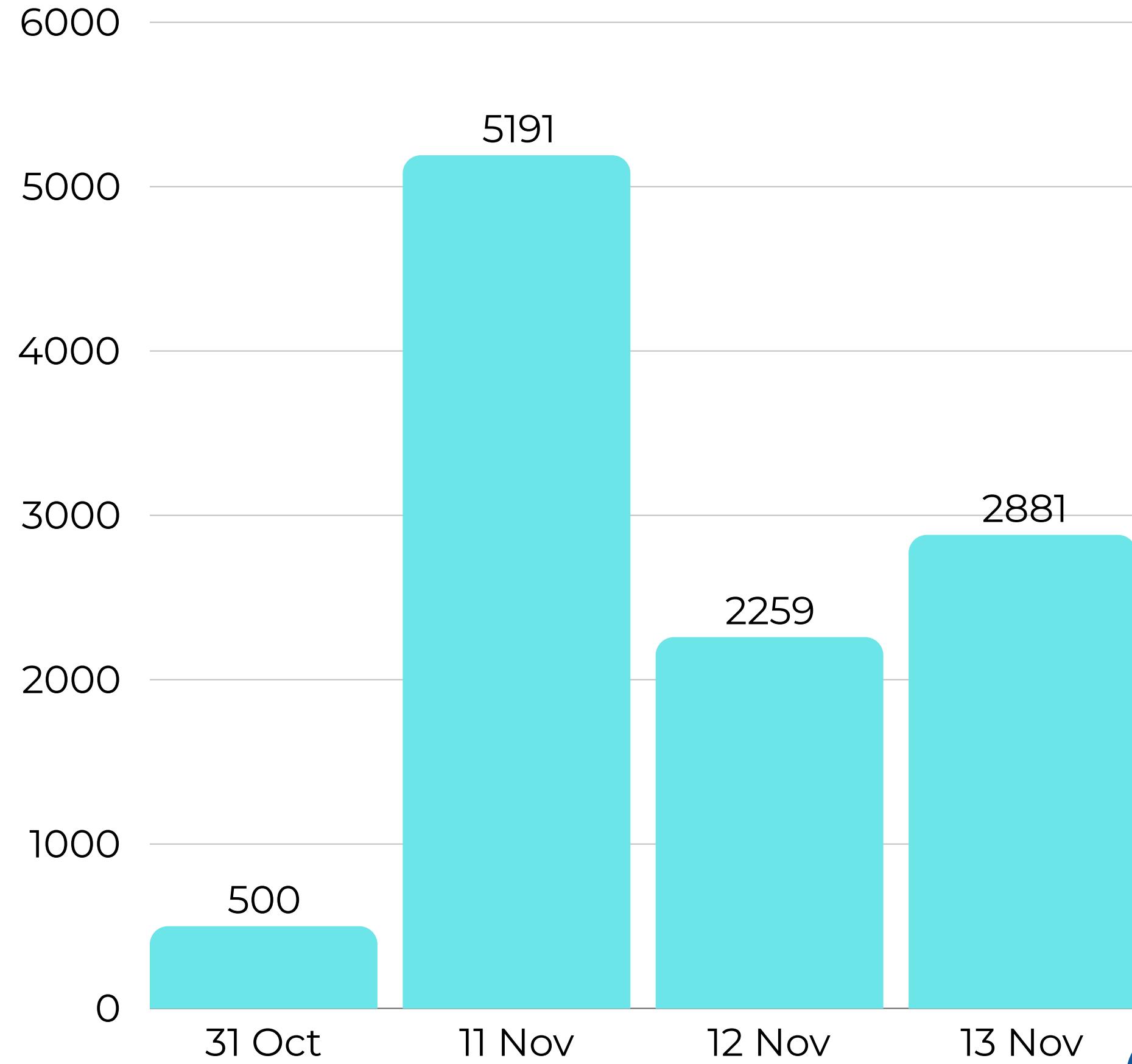
10831 data

Data Terkumpul Perpekan

- Pekan 11 : 500 data
- Pekan 12 : 0 data
- Pekan 13 : 10.331 data

Fitur Data

- **Link:** Tautan spesifik dari sumber data,
- **Title:** Judul dari pengaduan masyarakat,
- **Contents:** Isi/deskripsi/detail dari pengaduan,
- **Label:** Kategori/jenis klasifikasi dari pengaduan,
- **Timestamp:** Waktu pengambilan data dari sumber.





Pre-Processing

- 1. Lower Casing**
- 2. Cleaning Text**
- 3. Normalization**
- 4. Remove Stopwords**



Exploratory Data Analysis

1. Deteksi Missing Data & Duplikat

Ket :

Duplikat = 196

Null = 1

Jumlah data terkini : 10.634

Exploratory Data Analysis

Universitas Hasanuddin



2. Rata-rata Jumlah kata setiap kelas

Sebelum Pre-processing :

Infrastruktur dan Fasilitas Umum	83.217701
Keadilan Hukum dan HAM	476.448232
Kepegawaian	79.214421
Kesehatan	99.633144
Ketenagakerjaan dan Kesejahteraan Sosial	97.065550
Lingkungan Hidup	91.693811
Pendidikan	95.621035
Perizinan dan Pertanahan	111.457475
Teknologi Informasi dan Komunikasi	75.023613

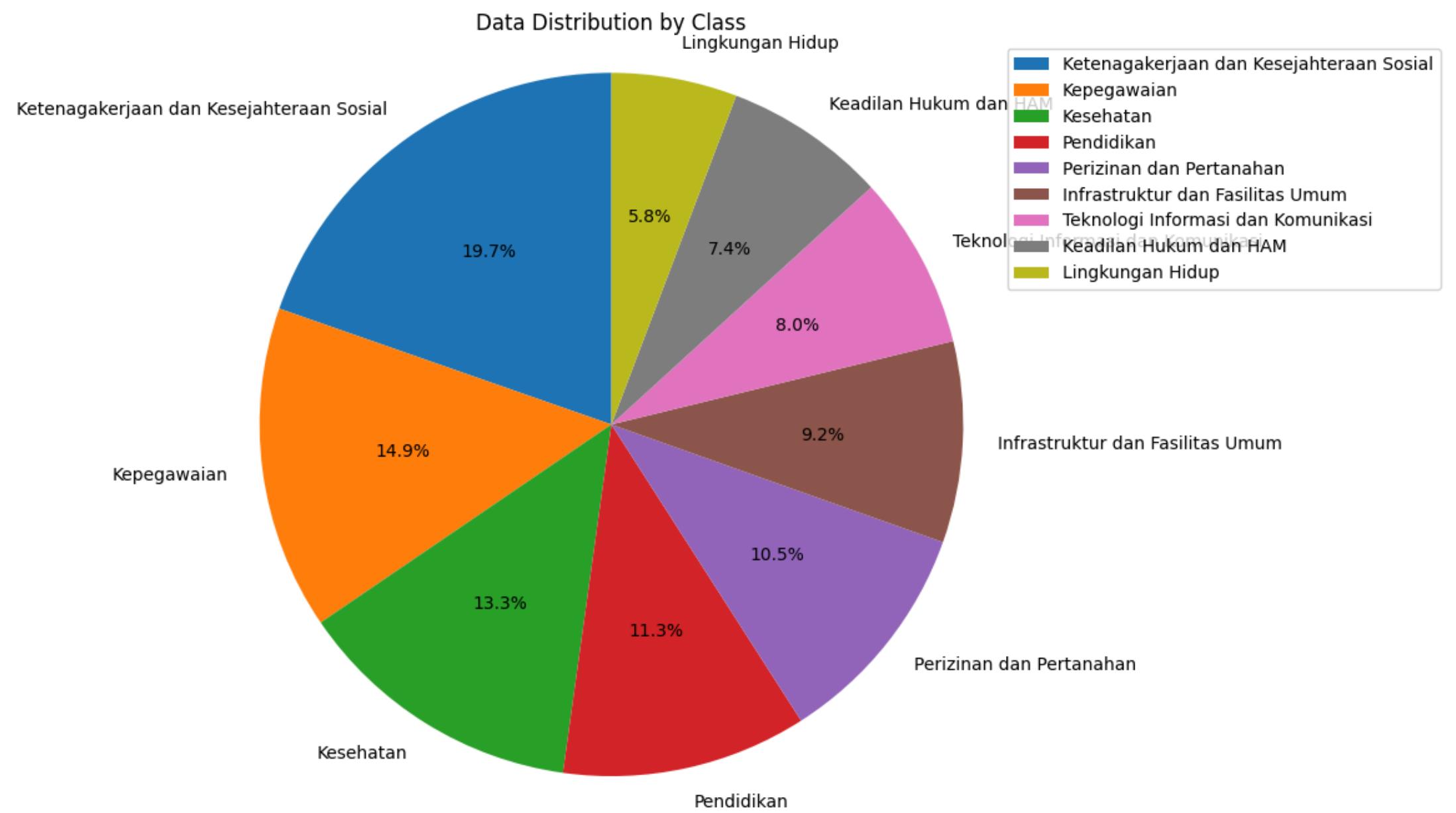
dtype: float64

Setelah Pre-processing

	labels
Infrastruktur dan Fasilitas Umum	63.525941
Keadilan Hukum dan HAM	367.198232
Kepegawaian	58.058191
Kesehatan	70.684844
Ketenagakerjaan dan Kesejahteraan Sosial	68.147368
Lingkungan Hidup	67.765472
Pendidikan	67.282972
Perizinan dan Pertanahan	78.738585
Teknologi Informasi dan Komunikasi	59.547816

Exploratory Data Analysis

3. Distribusi data setiap kelas





Exploratory Data Analysis

4. Frekuensi Data setiap kelas

Sebelum Pre-processing :



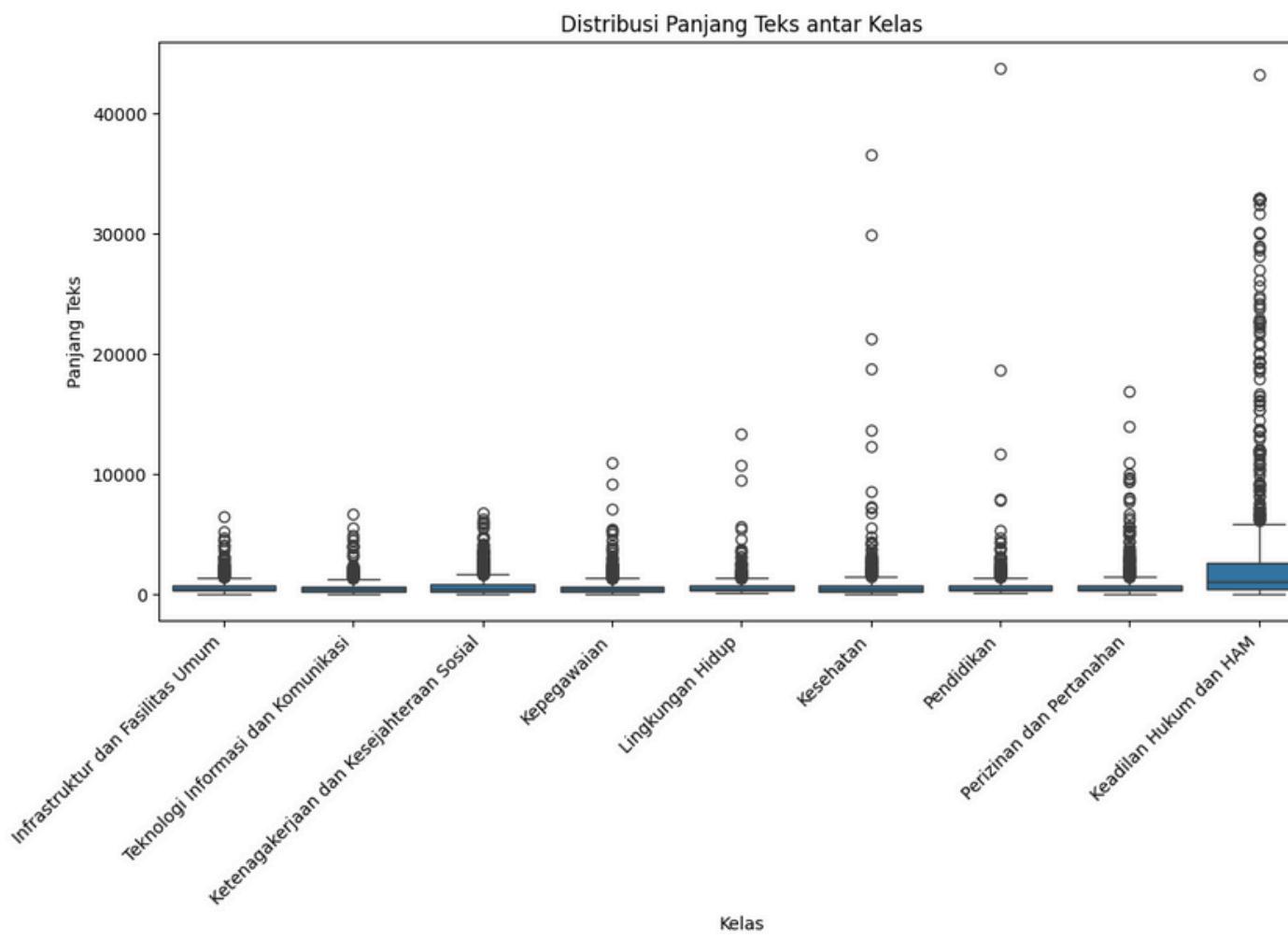
Setelah Pre-processing



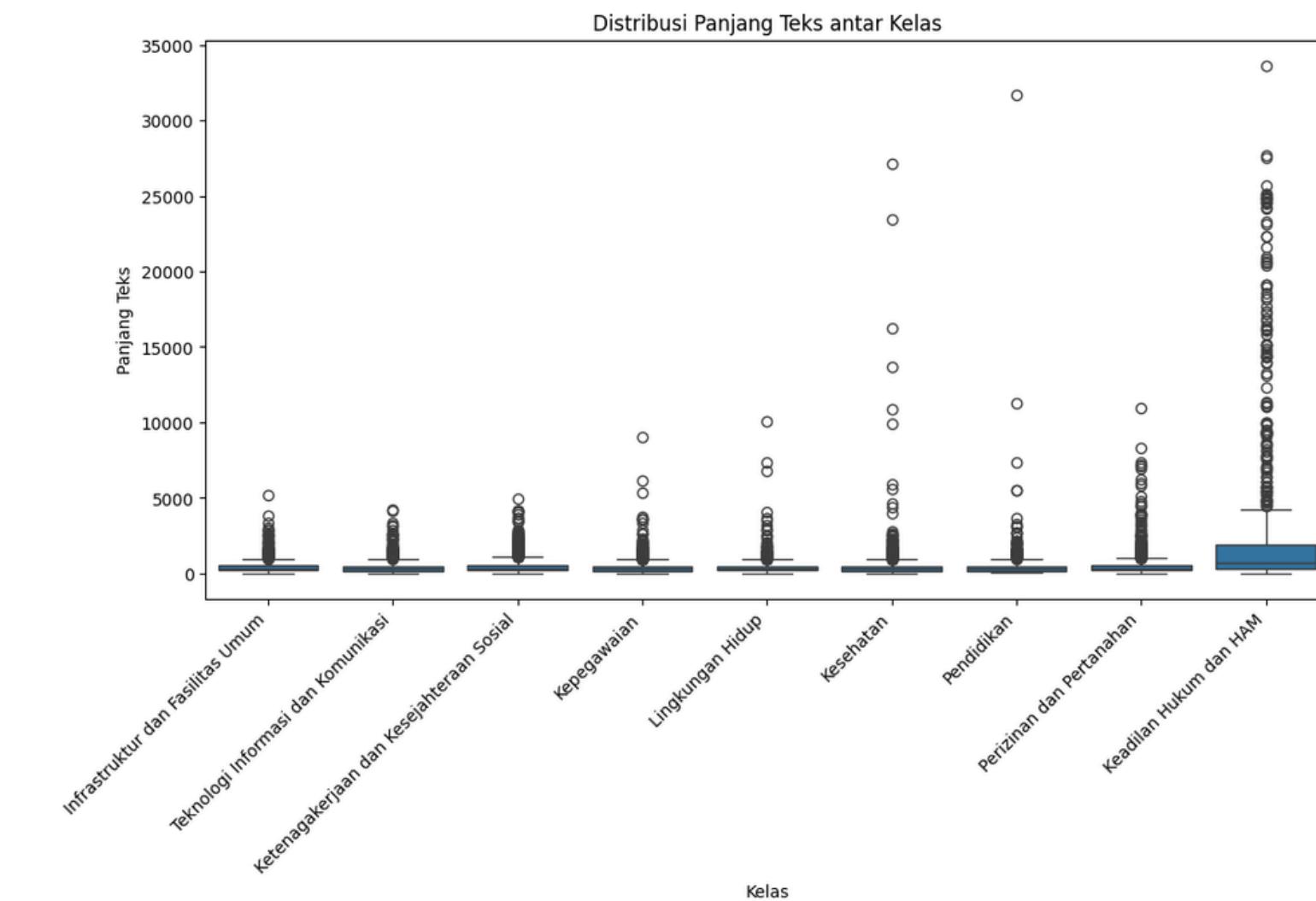
Exploratory Data Analysis

5. Panjang Karakter berdasarkan kelas

Sebelum Pre-processing :



Setelah Pre-processing :





Exploratory Data Analysis

6. Bigrams dan Trigrams

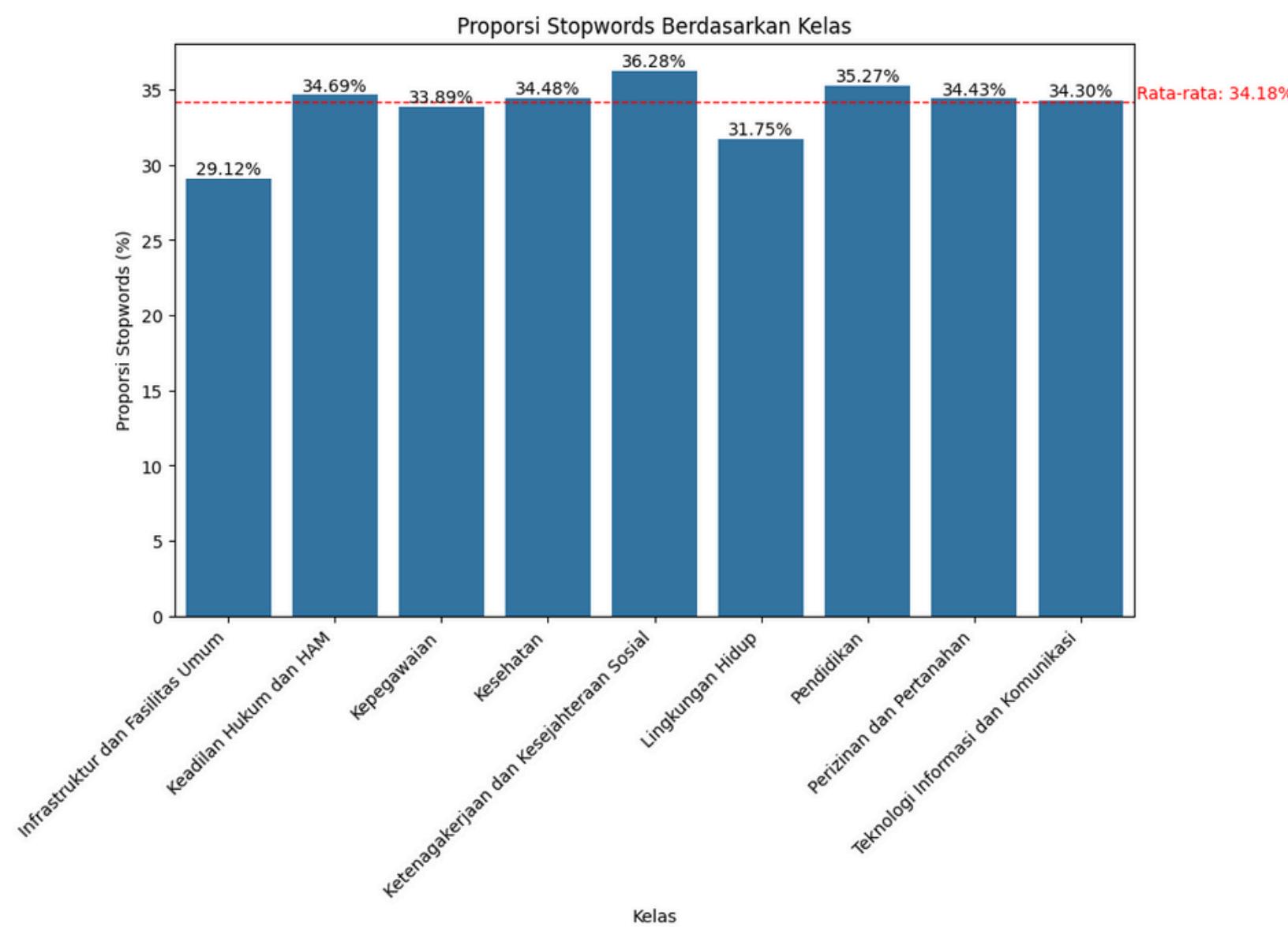
Bigrams	Jumlah	Trigrams	Jumlah
tidak', 'ada'	1871	hal', 'ini', 'dapat'	451
hal', 'ini'	1512	sampai', 'saat', 'ini'	415
saya', 'ingin'	1372	komisi', 'iii', 'dpr'	312
yang', 'tidak'	1320	saya', 'ingin', 'melaporkan'	304
saat', 'ini'	1264	tidak', 'sesuai', 'dengan'	277
terima', 'kasih'	1239	iii', 'dpr', 'ri'	258
saya', 'sudah'	1137	bapak/ibu', 'anggota', 'komisi'	223
saya', 'tidak'	965	anggota', 'komisi', 'iii'	214
tidak', 'bisa'	936	yang', 'mulia', 'bapak/ibu'	207
dan', 'tidak'	910	mulia', 'bapak/ibu', 'anggota'	205



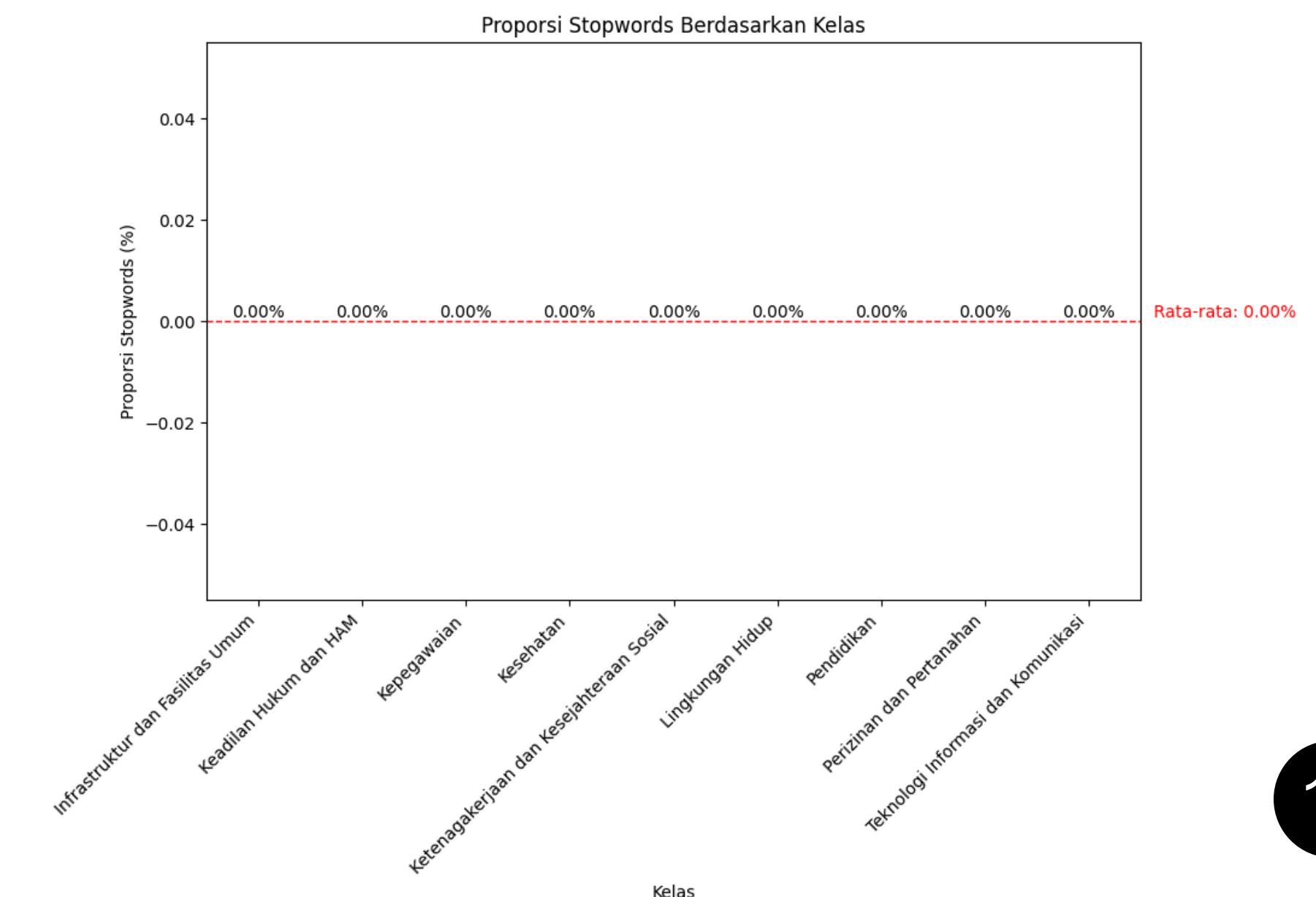
Exploratory Data Analysis

7. Proporsi Stopwords dalam Data Berdasarkan Kelas

Sebelum Pre-processing :



Setelah Pre-processing :





Pembuatan Model



Model

01

LSTM

Termasuk dalam kategori RNN.
Cocok untuk jenis data sequential
dan tugas text classification.

02

IndoBERT

Dibangun dari framework BERT
dengan penyesuaian untuk
menangani data berbahasa
indonesia.

03

XLM-RoBERTa

Dibangun dari RoBERTa sebagai
pengembangan dari BERT.
Menghilangkan tugas Next Sentence
Prediction dari BERT dan fokus pada
tugas MLM. XLM-RoBERTa dilatih pada
data multibahasa dengan volume besar.



Optimisasi

01

LSTM

- Hyperparameter Tuning
- Penambahan Dropout pada dense layer
- Pemberian EarlyStopping Callback

02

IndoBERT

- Pemberian EarlyStopping Callback
- Evaluasi per 100 steps

03

XLM-RoBERTa

- Pemberian EarlyStopping Callback

Evaluasi & Hasil Analisis

LSTM

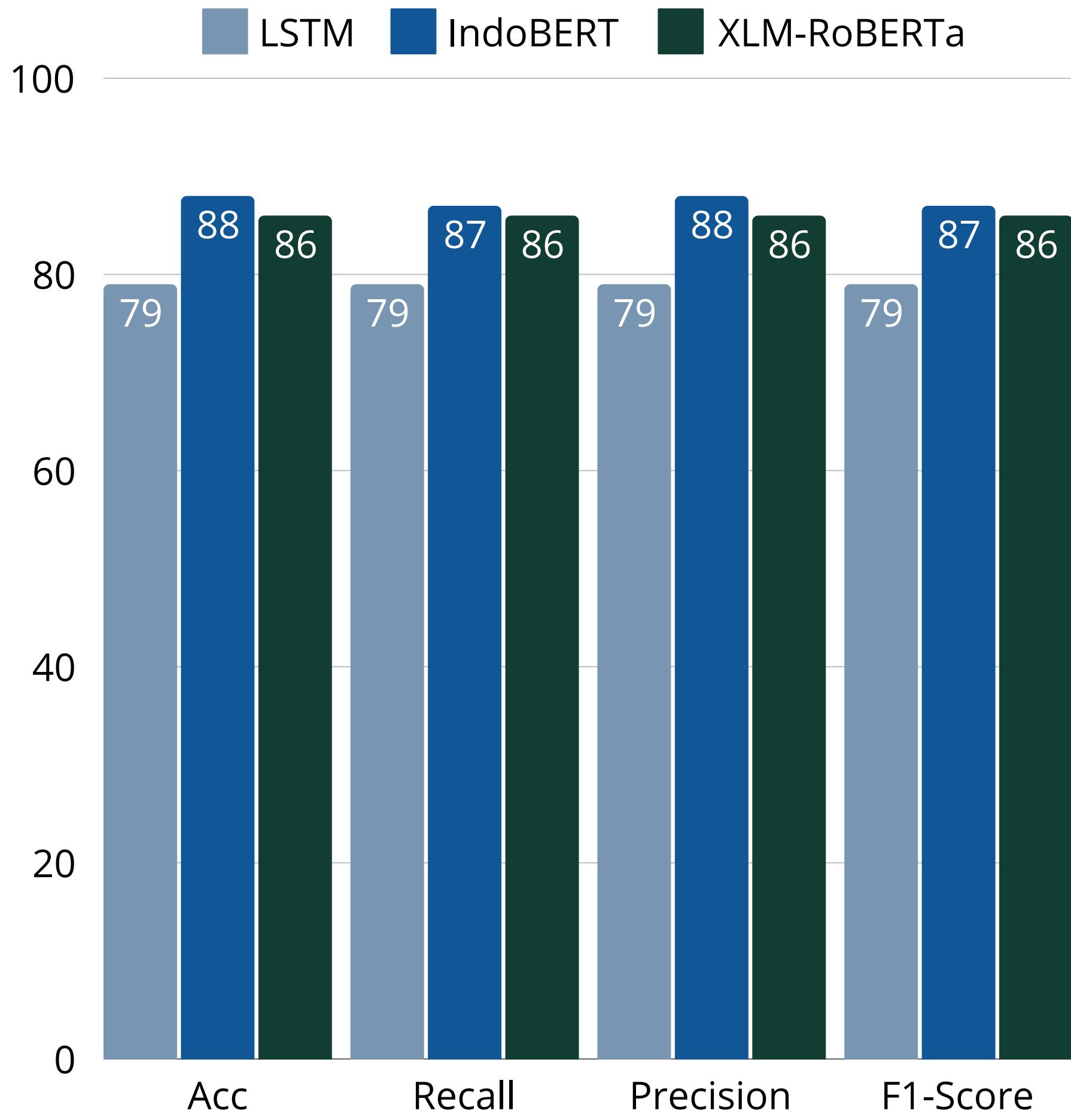
Model LSTM memiliki rata-rata nilai 79 untuk tiap metrik performa

IndoBERT

Rata-rata nilai 88 untuk tiap metrik performa

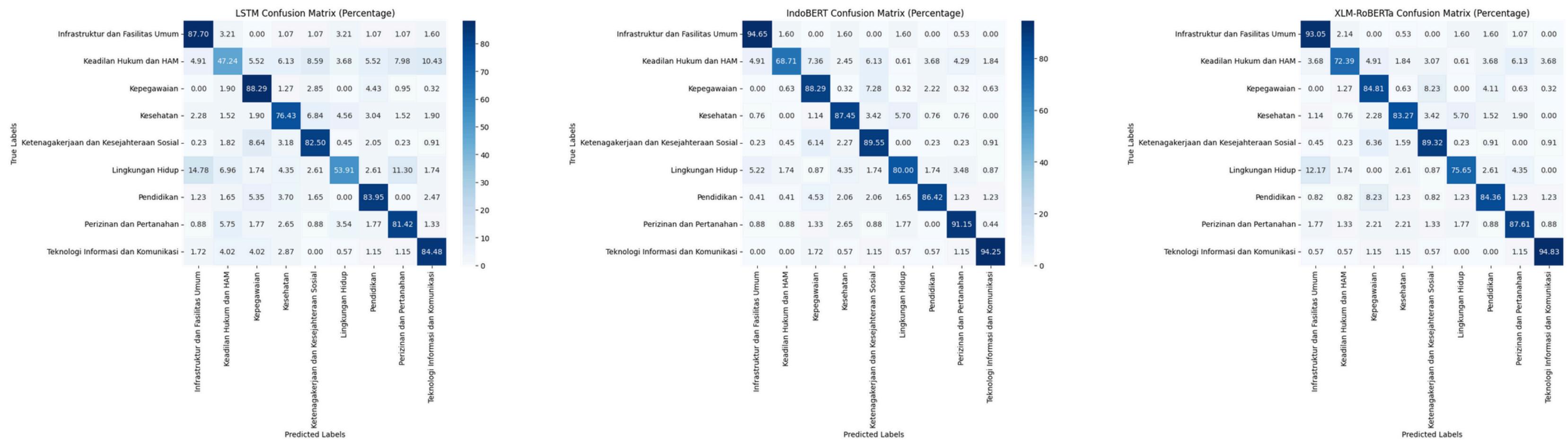
XLM-RoBERTa

Rata-rata nilai 86 untuk tiap metrik performa



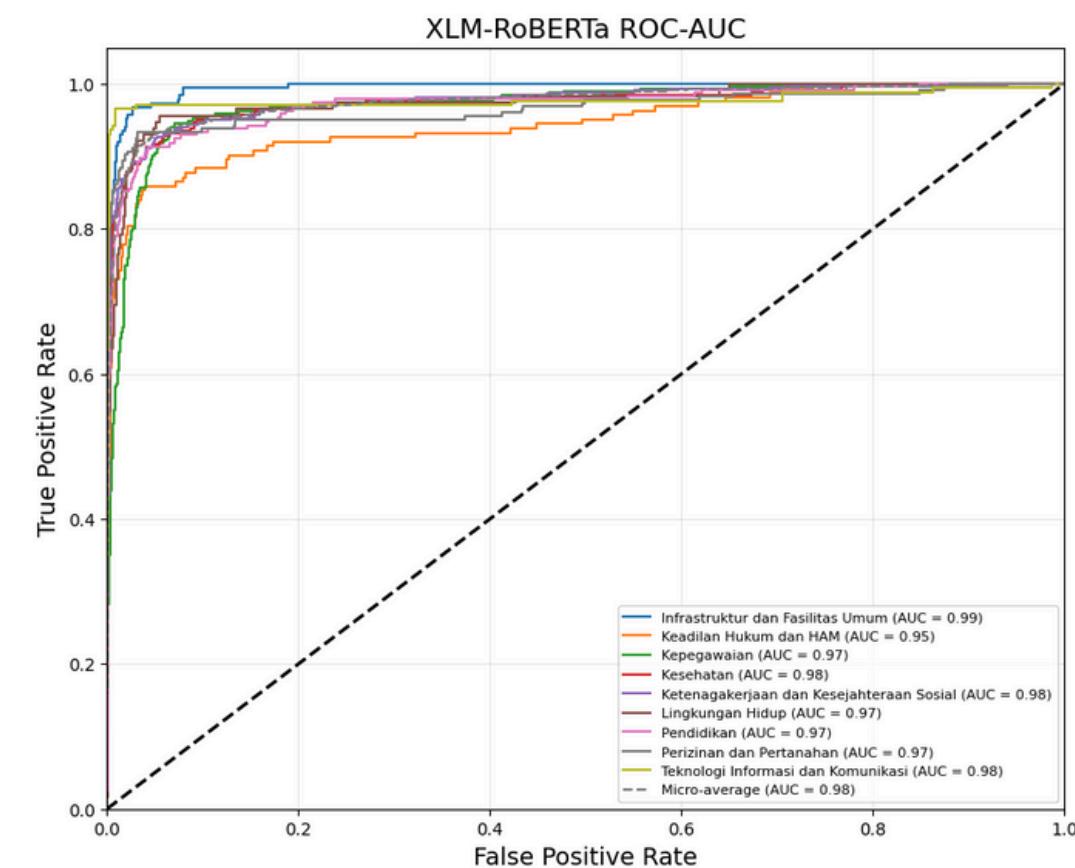
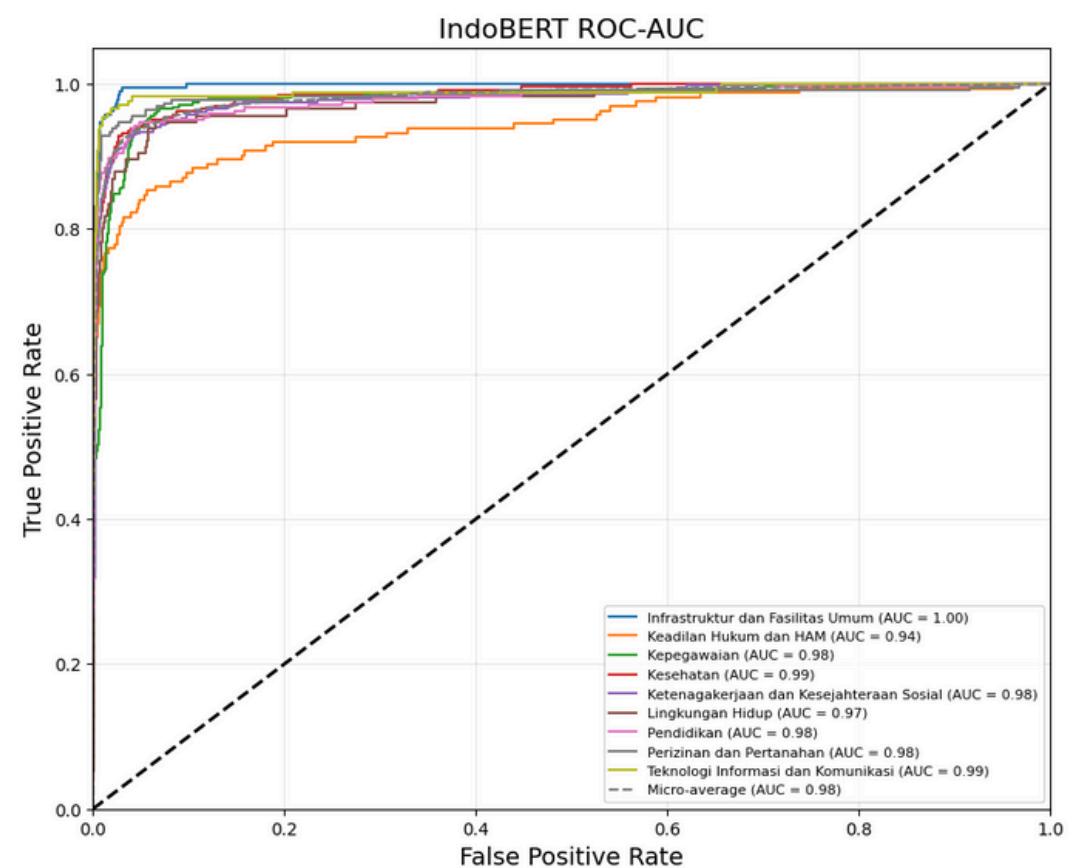
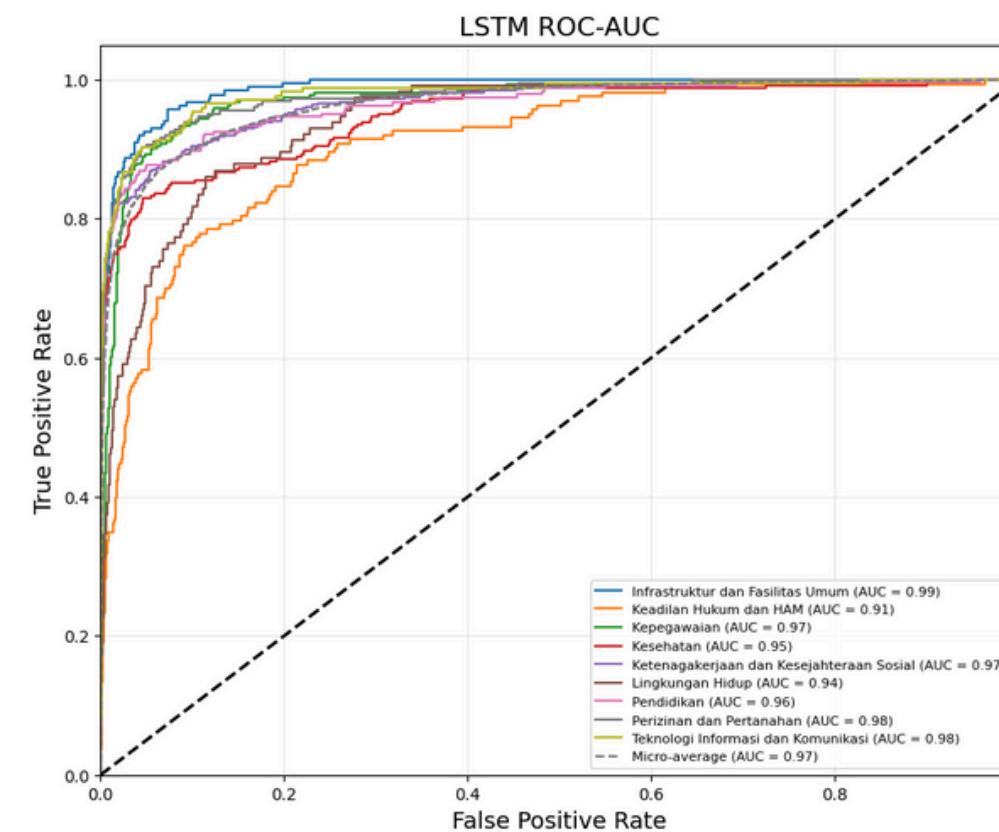


Evaluasi & Hasil Analisis



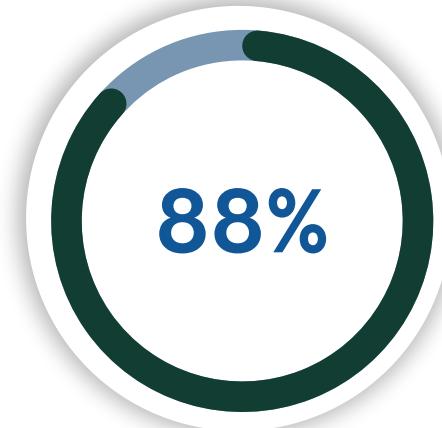


Evaluasi & Hasil Analisis



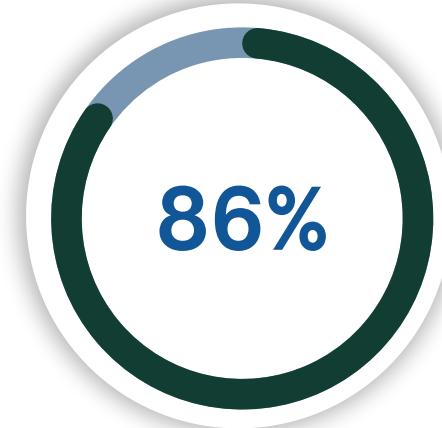


Kesimpulan



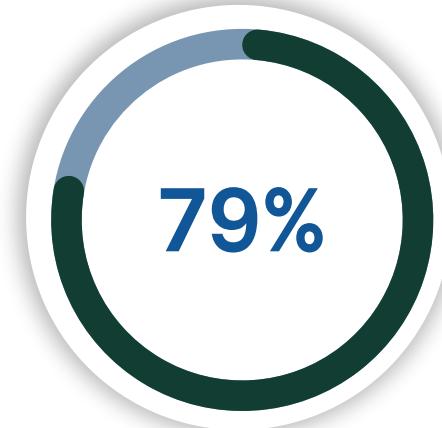
IndoBERT

IndoBERT adalah model dengan performa terbaik di antara ketiganya, menunjukkan kemampuan yang unggul untuk menangani tugas berbasis teks dalam bahasa Indonesia.



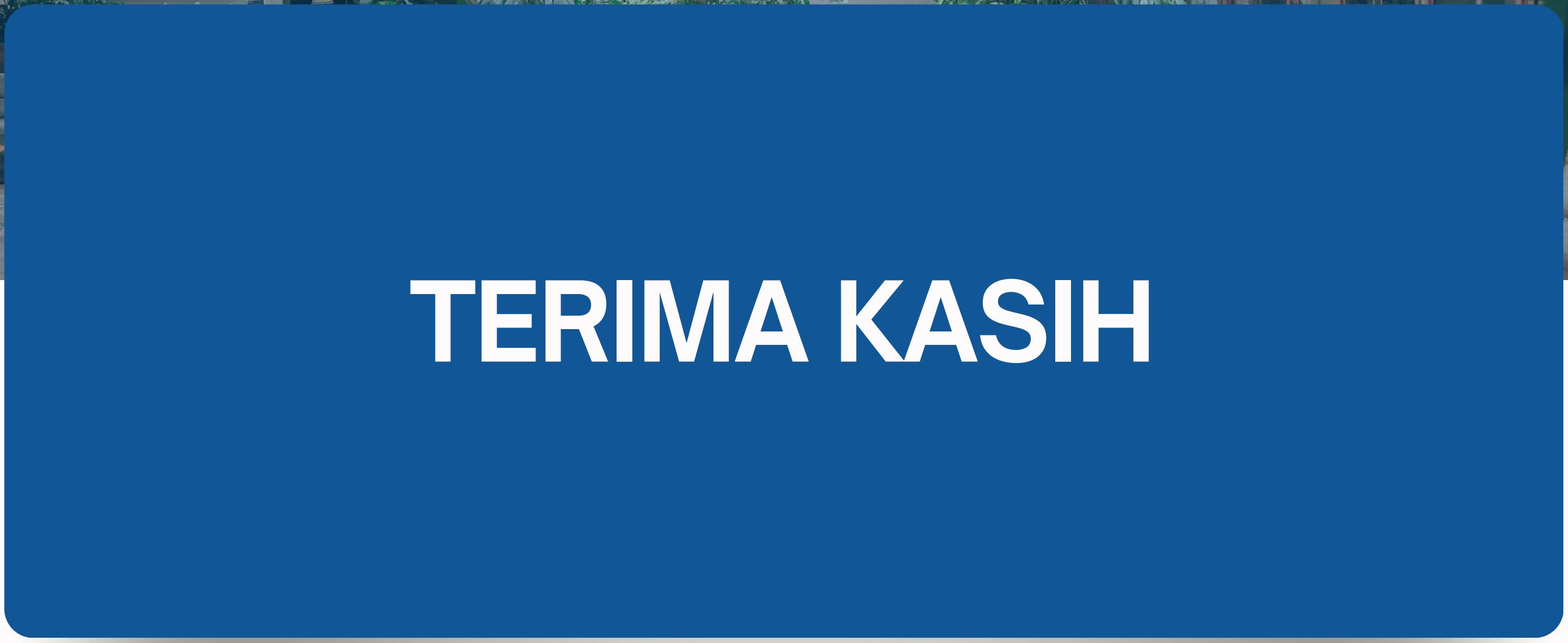
XLM-RoBERTa

Walaupun memiliki performa sedikit di bawah IndoBERT, namun XLM-RoBERTa tetap dapat menjadi alternatif yang baik terutama untuk skenario multibahasa.



LSTM

Meskipun memiliki performa terendah, namun penggunaanya mungkin tetap relevan, utamanya jika memandang dari sisi size model.



TERIMA KASIH