

IMPLEMENTASI BILSTM DAN ATTENTION PADA APLIKASI SEMANTIC TEXTUAL SIMILARITY BERBASIS WEB

SKRIPSI

Oleh:

**CHRISANDY WIRANATA
NIM. 221110252
MUHAMMAD FAHIM AL - MUBAROK
NIM. 221110196
SANDY AGRE NICOLA
NIM. 221110040**



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

IMPLEMENTATION OF BILSTM AND ATTENTION FOR A WEB-BASED SEMANTIC TEXTUAL SIMILARITY APPLICATION

FINAL RESEARCH

By:

**CHRISANDY WIRANATA
ID NUMBER. 221110252
MUHAMMAD FAHIM AL - MUBAROK
ID NUMBER. 221110196
SANDY AGRE NICOLA
ID NUMBER. 221110040**



**MAJOR OF S-1 INFORMATICS ENGINEERING
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

**IMPLEMENTASI BILSTM DAN ATTENTION PADA APLIKASI
SEMANTIC TEXTUAL SIMILARITY BERBASIS WEB**

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknik Informatika

Oleh:

CHRISANDY WIRANATA
NIM. 221110252
MUHAMMAD FAHIM AL - MUBAROK
NIM. 221110196
SANDY AGRE NICOLA
NIM. 221110040

Disetujui Oleh:

Dosen Pembimbing I



Ng Poi Wong, S.Kom., M.T.I.

Dosen Pembimbing II



Sio Jurnal Pipin, S.Kom., M.Kom.

Medan, 26 Februari 2026
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-1 Teknik Informatika



Carles Jullandy, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Chrisandy Wiranata

NIM : 221110252

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi BiLSTM dan Attention pada Aplikasi Semantic Textual Similarity Berbasis Web

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Schubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 10 Januari 2026

Saya yang membuat pernyataan,



Chrisandy Wiranata

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Muhammad Fahim Al - Mubarak

NIM : 221110196

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi BiLSTM dan Attention pada Aplikasi Semantic Textual Similarity Berbasis Web

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 10 Januari 2026

Saya yang membuat pernyataan



Muhammad Fahim Al - Mubarak

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Sandy Agre Nicola

NIM : 221110040

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi BiLSTM dan Attention pada Aplikasi Semantic Textual Similarity Berbasis Web

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 10 Januari 2026

Saya yang membuat pernyataan,



Sandy Agre Nicola

IMPLEMENTASI BiLSTM DAN ATTENTION PADA APLIKASI SEMANTIC TEXTUAL SIMILARITY BERBASIS WEB

SKRIPSI

ABSTRAK

Plagiarisme parafrastik sulit dideteksi karena kesamaan makna dapat diekspresikan melalui struktur kalimat dan pilihan kata yang berbeda, sehingga pendekatan berbasis kesamaan leksikal menjadi kurang efektif. Penelitian ini mengembangkan aplikasi web untuk mengukur kesamaan semantik antarteks menggunakan model Bidirectional Long Short-Term Memory (BiLSTM) dengan mekanisme attention. Hasil pengujian menunjukkan bahwa model mencapai nilai loss sebesar 0,3237, accuracy 0,8545, dan AUC 0,9307 pada data uji, dengan F1-score sebesar 0,8837 untuk kelas non-duplikat dan 0,8059 untuk kelas duplikat. Pada pengujian tingkat dokumen, skor kemiripan yang diperoleh relatif rendah, yang diduga berkaitan dengan karakteristik data pelatihan yang didominasi oleh pasangan teks pendek dengan struktur kalimat sederhana.

Kata kunci: Kesamaan Semantik; Plagiarisme Parafrastik; BiLSTM; Mekanisme Attention.

Abstract

Paraphrastic plagiarism is difficult to detect because semantic similarity can be expressed through different sentence structures and word choices, making lexical similarity-based approaches less effective. This study develops a web-based application to measure semantic similarity between texts using a Bidirectional Long Short-Term Memory (BiLSTM) model with an attention mechanism. Experimental results show that the model achieves a loss of 0.3237, an accuracy of 0.8545, and an AUC of 0.9307 on the test data, with F1-scores of 0.8837 for the non-duplicate class and 0.8059 for the duplicate class. At the document level, the similarity scores tend to be relatively low, which is presumed to be related to the characteristics of the training data that are predominantly composed of short text pairs with simple sentence structures.

Keywords: Semantic Similarity; Paraphrastic Plagiarism; BiLSTM; Attention Mechanism.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik, yang berjudul “Implementasi BiLSTM dan Attention pada Aplikasi Semantic Textual Similarity Berbasis Web”.

Tugas Akhir ini disusun sebagai bagian dari persyaratan kurikulum pada Program Studi S-1 Teknik Informatika, Universitas Mikroskil Medan.

Penulis juga menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Ng Poi Wong, S.Kom., M.T.I., selaku Dosen Pembimbing I.
2. Bapak Sio Jurnalis Pipin, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
6. Bapak dan Ibu Dosen Universitas Mikroskil yang telah memberikan ilmunya selama menempuh pendidikan di Universitas Mikroskil Medan.
7. Keluarga penulis, khususnya orang tua, yang senantiasa memberikan dukungan moral dan materil sepanjang masa studi hingga Tugas Akhir ini terselesaikan.
8. Rekan-rekan serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan dukungan selama proses penyusunan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan di masa mendatang. Akhir kata, semoga Tugas Akhir ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan. Terima Kasih.

Medan, 10 Januari 2026

Penulis,

Chrisandy Wiranata

Muhammad Fahim Al – Mubarak

Sandy Agre Nicola

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR.....	v
DAFTAR TABEL	vii
DAFTAR LAMPIRAN.....	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	3
1.4 Manfaat.....	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR	5
2.1 <i>Artificial Intelligent (AI)</i>	5
2.1.1 <i>Machine Learning</i>	6
2.1.2 <i>Deep Learning</i>	7
2.2 <i>Artificial Neural Network (ANN)</i>	9
2.2.1 <i>Perceptron</i> dan Fungsi Aktivasi	10
2.2.2 Regularisasi, Fungsi <i>Loss</i> dan <i>Optimizer</i>	11
2.3 <i>Natural Language Processing (NLP)</i>	14
2.3.1 <i>Preprocessing</i>	14
2.3.2 Tokenisasi	15
2.3.3 <i>Word Embedding</i>	15
2.4 Pengukuran Kesamaan Teks.....	16
2.4.1 Pendekatan Leksikal	17
2.4.2 Pendekatan Semantik	18
2.4.3 Fungsi Pengukuran Kemiripan.....	18
2.5 <i>BERT</i>	19
2.6 <i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	22
2.7 <i>Attention</i>	25
2.8 Evaluasi Model.....	29
BAB III TAHAPAN PELAKSANAAN	31

3.1 Kerangka Tahapan Pelaksanaan	31
3.2 <i>Dataset</i>	32
3.3 Analisis Proses.....	33
3.4 Analisis Kebutuhan	62
3.4.1 Analisis Kebutuhan Fungsional	62
3.4.2 Analisis Kebutuhan Non Fungsional	67
3.5 Perancangan.....	68
BAB IV HASIL DAN PEMBAHASAN	74
4.1 Hasil.....	74
4.1.1 Tampilan Halaman <i>Text Similarity Analysis</i>	74
4.1.2 Tampilan Halaman <i>Document Similarity Analysis</i>	75
4.1.3 Tampilan Halaman <i>Login</i>	77
4.1.4 Tampilan Halaman Pelatihan Ulang Model.....	78
4.2 Pembahasan	84
4.2.1 Hasil Pengujian <i>Text Similarity</i>	90
4.2.2 Hasil Pengujian <i>Document Similarity</i>	94
BAB V PENUTUP.....	98
5.1 Kesimpulan.....	98
5.2 Saran	98
DAFTAR PUSTAKA.....	99
DAFTAR RIWAYAT HIDUP	110

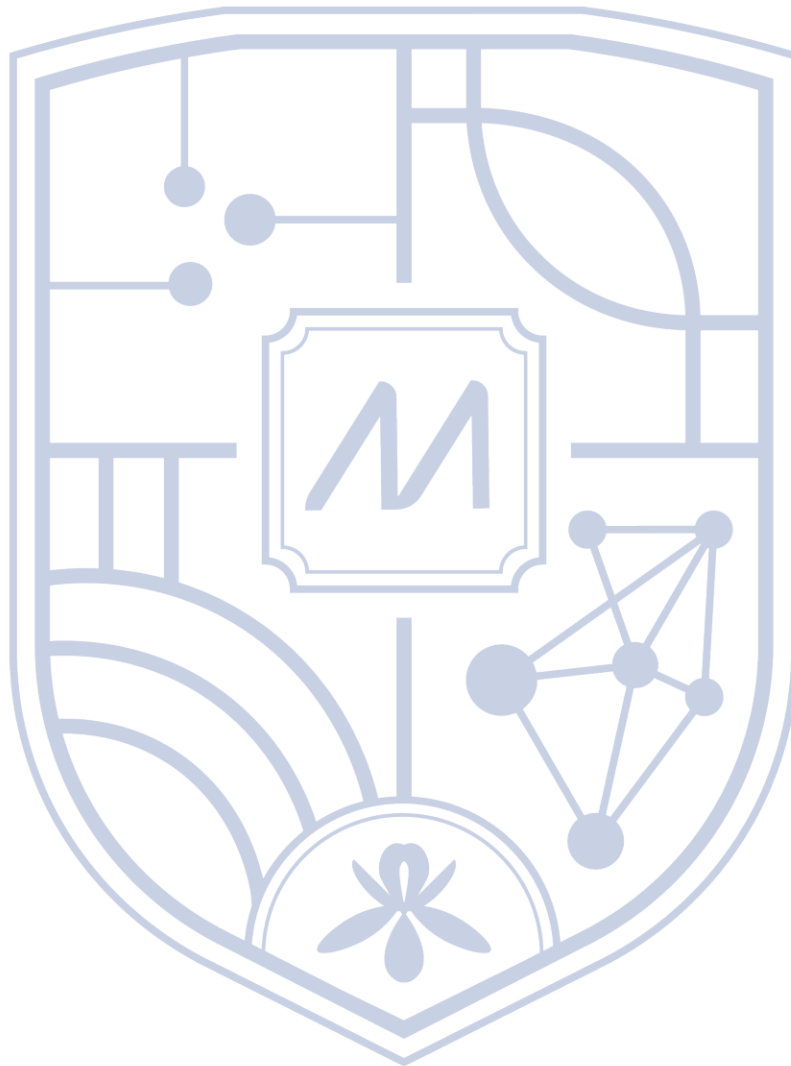
DAFTAR GAMBAR

Gambar 2.1 Komponen Kunci dalam AI [20]	6
Gambar 2.2 Hierarki Konsep Kecerdasan Buatan [23]	6
Gambar 2.3 Struktur <i>Feedforward Neural Network</i> [28]	10
Gambar 2.4 Model <i>Bi-Encoder vs. Cross-Encoder</i> [73].....	22
Gambar 2.5 Arsitektur LSTM [79]	24
Gambar 2.6 Model Arsitektur BiLSTM [81]	25
Gambar 2.7 Perbedaan Mekanisme <i>Additive Attention</i> dan <i>Self-Attention</i> [85].....	28
Gambar 2.8 Arsitektur Mekanisme <i>Additive Attention</i> [86]	28
Gambar 3.1 Contoh Pasangan Kalimat <i>Dataset QQP</i>	33
Gambar 3.2 <i>Flowchart</i> Proses Pengukuran <i>Semantic Text Similarity</i>	34
Gambar 3.3 <i>Flowchart</i> Tahap Pra-Pemrosesan Data.....	35
Gambar 3.4 <i>Flowchart</i> Tahap Tokenisasi dan <i>Embedding</i>	37
Gambar 3.5 <i>Flowchart</i> Tahap <i>Layer Normalization</i>	41
Gambar 3.6 Arsitektur <i>Training</i> Pada Model BiLSTM <i>Attention</i>	44
Gambar 3.7 <i>Use Case Diagram</i> Sistem Skoring Plagiarisme	63
Gambar 3.8 Rancangan Tampilan <i>Sidebar</i> Aplikasi.....	68
Gambar 3.9 Rancangan Tampilan Halaman <i>Text Similarity</i>	69
Gambar 3.10 Rancangan Tampilan Halaman <i>Document Similarity</i>	69
Gambar 3.11 Rancangan Tampilan Halaman Autentikasi.....	70
Gambar 3.12 Rancangan Tampilan Halaman <i>Admin Dashboard</i>	71
Gambar 4.1 Tampilan Halaman <i>Text Similarity Analysis</i>	74
Gambar 4.2 Tampilan Hasil <i>Text Similarity Analysis</i>	75
Gambar 4.3 Tampilan Halaman <i>Document Similarity Analysis</i>	76
Gambar 4.4 Tampilan Hasil <i>Document Similarity Analysis</i>	77
Gambar 4.5 Tampilan Halaman <i>Login</i>	78
Gambar 4.6 Tampilan Halaman <i>Retrain Model</i> Bagian <i>Upload Dataset</i>	79
Gambar 4.7 Tampilan Halaman <i>Retrain Model</i> Bagian <i>Dataset Summary</i>	79
Gambar 4.8 Tampilan Hasil <i>Upload Dataset</i>	79
Gambar 4.9 Tampilan Hasil <i>Dataset Summary</i>	80
Gambar 4.10 Tampilan Halaman <i>Retrain Model</i> Bagian Konfigurasi	80
Gambar 4.11 Tampilan Halaman <i>Retrain Model</i> Bagian Akhir	81
Gambar 4.12 Tampilan Proses <i>Training</i>	82

Gambar 4.13 Grafik *Accuracy* Selama 20 *Epoch* Pada *Dataset Training* dan *Validation* ..88

Gambar 4.14 Grafik *Loss* Selama 20 *Epoch* Pada *Dataset Training* dan *Validation*89

Gambar 4.15 Grafik *AUC* Selama 20 *Epoch* Pada *Dataset Training* dan *Validation*90



DAFTAR TABEL

Tabel 2.1 Perbandingan Model STS Pada <i>Dataset Quora Question Pairs</i> Berdasarkan Penelitian Sebelumnya	16
Tabel 3.1 Statistik Deskriptif <i>Dataset QQP</i>	33
Tabel 3.2 Perhitungan <i>Mean, Var, Std</i>	43
Tabel 3.3 Hasil Perhitungan Normalisasi	43
Tabel 3.4 <i>Hyperparameter</i> Yang Dicari Oleh <i>Hyperband</i>	44
Tabel 3.5 <i>Hyperparameter</i> Terbaik Berdasarkan <i>Hyperband</i>	45
Tabel 3.6 <i>Final Setting Hyperparameter</i> Model <i>BiLSTM Attention</i>	45
Tabel 3.7 Hasil <i>BiLSTM</i> Arah <i>Forward</i>	52
Tabel 3.8 <i>Weight input, Weight Hidden</i> dan Bias pada <i>LSTM Backward</i>	54
Tabel 3.9 Hasil <i>BLSTM</i> Arah <i>Backward</i>	55
Tabel 3.10 Hasil Penggabungan <i>BiLSTM</i> dari <i>Forward</i> dan <i>Backward</i>	55
Tabel 3.11 Aktivasi Nonlinier dan Skor Relevansi Token	58
Tabel 3.12 Pembentukan Vektor Konteks dari Nilai Kontribusi	59
Tabel 3.13 Peran Aktor pada <i>Use Case Diagram</i>	63
Tabel 3.14 <i>Use Case</i> Mengecek Similaritas Antar Kalimat	63
Tabel 3.15 <i>Use Case</i> Mengecek Similaritas Antar Dokumen	64
Tabel 3.16 <i>Use Case Login</i>	65
Tabel 3.17 <i>Use Case</i> Melatih Ulang Model STS	65
Tabel 3.18 <i>Use Case</i> Memilih Model	66
Tabel 3.19 Analisis Kebutuhan Non-Fungsional Sistem Usulan <i>PIECES</i>	67
Tabel 4.1 Konfigurasi Trial <i>Hyperband</i> Beserta Hasil	86
Tabel 4.2 Hasil Evaluasi <i>Full Dataset Training</i> Pada Data Pelatihan dan Data Validasi Selama <i>20 Epoch</i>	87
Tabel 4.3 <i>Confusion Matrix</i> pada Data Pengujian	91
Tabel 4.4 Hasil Evaluasi Performa pada Data Pelatihan, Validasi, dan Pengujian	91
Tabel 4.5 Nilai <i>Precision, Recall, dan F1-Score</i> untuk Setiap Label pada Data Pengujian	92
Tabel 4.6 Beberapa Contoh Pasangan Kalimat Beserta Label dan Prediksi Model	92
Tabel 4.7 Detail 10 Dokumen Yang Digunakan Untuk Perbandingan Dokumen	95
Tabel 4.8 Hasil Pengujian Terhadap 10 Abstrak dari Dokumen	96

DAFTAR LAMPIRAN

Lampiran 1 Tabel Pengujian.....	107
---------------------------------	-----

