TESIS

ANALISIS STRATEGI PEMBOBOTAN PADA LOSS FUNCTON UNTUK KLASIFIKASI KEBUTUHAN PERANGKAT LUNAK DALAM AGILE USER STORY MENGGUNAKAN BERT

Oleh:
Muhammad Fahshon Hakim
23206051019

PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UIN SUNAN KALIJAGA

YOGYAKARTA
2026

HALAMAN PENGESAHAN

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2411/Un.02/DST/PP.00.9/10/2026

Tugas Akhir dengan judul : ANALISIS STRATEGI PEMBOBOTAN PADA LOSS FUNCTON UNTUK KLASIFIKASI KEBUTUHAN PERANGKAT LUNAK DALAM AGILE USER STORY MENGGUNAKAN BERT

yang dipersiapkan dan disusun oleh:

Nama : MUHAMMAD FAHSHON HAKIM, S.Kom.
Nomor Induk Mahasiswa : 23206051019
Telah diujikan pada : Jumat, 28 Agustus 2026
Nilai ujian Tugas Akhir : A/B

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 6abff6e7902ea8

Penguji I

Dr. Ir. Bambang Sugiantoro, M.T., IPU., ASEAN Eng.
SIGNED

Valid ID: 6ab622af50300

Penguji II

Dr. Ir. Sumarsono, S.T., M.Kom.
SIGNED

Valid ID: 6abcac841c195

Yogyakarta, 28 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6ac3209fc9e06

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

SURAT PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama	: Muhammad Fahshon Hakim
NIM	: 23206051019
Jenjang	: Magister
Program Studi	: Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 26 Agustus 2026
Saya yang menyatakan,



Muhammad Fahshon Hakim
NIM: 23206051019

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu'alaikum wr. wb.

Disampaikan dengan hormat, setelah melakukan bimbingan, arahan dan koreksi terhadap penulisan tesis yang berjudul:

ANALISIS STRATEGI PEMBOBOTAN PADA LOSS FUNCTON UNTUK KLASIFIKASI KEBUTUHAN PERANGKAT LUNAK DALAM AGILE USER STORY MENGGUNAKAN BERT

Yang ditulis oleh:

Nama	: Muhammad Fahshon Hakim
NIM	: 23206051019
Jenjang	: Magister
Program Studi	: Informatika

Saya berpendapat bahwa tesis tersebut sudah bisa diajukan kepada Program Studi Magister Informatika UIN Sunan Kalijaga Yogyakarta untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Wassalamu'alaikum wr. wb.

Yogyakarta, 26 Agustus 2026
Pembimbing,

Ir. Agung Fatwanto, S.Si, M.Kom, Ph.D
NIP. 19770103 200501 1 003

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

ABSTRAK

Ketidakseimbangan kelas pada klasifikasi kebutuhan perangkat lunak dalam *Agile User Story* memerlukan evaluasi yang memperhatikan performa kelas mayoritas dan minoritas. Penelitian ini menganalisis strategi pembobotan pada Cross-Entropy dan Focal Loss menggunakan BERT-base-uncased. Dataset terdiri atas 10.006 *user story* berbahasa Inggris dengan delapan kelas, yaitu satu kelas *functional requirements* (FR) dan tujuh kategori *non-functional requirements* (NFR). Sepuluh konfigurasi dibentuk dari dua fungsi loss dan lima kondisi: tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999. Static merupakan pengali loss seragam 0,25, sedangkan inverse-frequency dan class-balanced menggunakan bobot berdasarkan frekuensi kelas pada data pelatihan. Parameter dasar mengikuti konfigurasi terbaik penelitian acuan dan dipertahankan sama. Macro F1 digunakan sebagai metrik utama, dilengkapi Accuracy, Weighted F1, Matthews Correlation Coefficient (MCC), metrik per kelas, dan *confusion matrix*. Cross-Entropy dengan class-balanced beta 0,999 menghasilkan skor agregat tertinggi pada run yang diamati: Macro F1 0,8907, Accuracy 0,9341, Weighted F1 0,9344, dan MCC 0,8840. Focal Loss menghasilkan Macro F1 lebih tinggi daripada Cross-Entropy pada kondisi tanpa pembobotan dan inverse-frequency, sedangkan Cross-Entropy lebih tinggi pada static dan kedua kondisi class-balanced. Inverse-frequency mengurangi kesalahan NFR menjadi FR, tetapi meningkatkan kesalahan FR menjadi NFR. Hasil menunjukkan bahwa hubungan antara fungsi loss dan performa bergantung pada strategi pembobotan, dengan respons kelas yang tidak seragam.

Kata kunci: Klasifikasi Kebutuhan Perangkat Lunak; BERT; Cross-Entropy; Focal Loss; Strategi Pembobotan.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

ٱلَّذِى خَلَقَ ٱلْمَوْتَ وَٱلْحَيَوٰةَ لِيَبْلُوَكُمْ أَيُّكُمْ أَحْسَنُ عَمَلًا ۚ وَهُوَ ٱلْعَزِيزُ ٱلْغَفُورُ

"Yang menciptakan mati dan hidup, untuk menguji kamu, siapa di antara kamu yang lebih baik amalnya. Dan Dia Maha Perkasa, Maha Pengampun"

Al-Mulk (67:2)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Tesis ini penulis persembahkan kepada:

Almamater Tercinta Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahi Rabbil ‘Alamin. Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan pertolongan-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul “Analisis Strategi Pembobotan pada Loss Functon untuk Klasifikasi Kebutuhan Perangkat Lunak dalam Agile User Story Menggunakan BERT”. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman. Semoga kita termasuk umat yang memperoleh syafaat beliau di hari akhir.

Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Selain sebagai bagian dari proses akademik, penelitian ini diharapkan dapat memberikan tambahan pengetahuan dan wawasan serta menjadi referensi yang bermanfaat bagi para pembaca.

Penyusunan dan penyelesaian tesis ini tidak terlepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis menyampaikan penghargaan kepada:

1. Allah SWT atas segala rahmat, pertolongan, dan kemudahan yang diberikan kepada penulis selama menempuh pendidikan hingga menyelesaikan tesis ini.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

2. Kedua orang tua dan seluruh keluarga yang senantiasa memberikan doa, dukungan, motivasi, serta bantuan baik secara moral maupun material selama penulis menempuh pendidikan.
3. Prof. Noorhaidi Hasan, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
5. Dr. Ir. Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Magister Informatika, serta Ir. Muhammad Didik Rohmad Wahyudi, S.T., M.T., selaku Sekretaris Program Studi Magister Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
6. Ir. Agung Fatwanto, S.Si., M.Kom., Ph.D., selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan arahan, bimbingan, masukan, dan koreksi selama proses penelitian dan penyusunan tesis ini.
7. Seluruh dosen dan staf Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta, atas ilmu, bantuan, dan pengalaman yang diberikan selama penulis menempuh pendidikan.
8. Teman-teman seperjuangan di Program Studi Magister Informatika yang telah menjadi bagian dari proses pembelajaran serta memberikan dukungan selama masa perkuliahan dan penyusunan tesis.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

9. Seluruh pihak yang telah memberikan bantuan dan dukungan selama penulis menempuh pendidikan magister dan menyelesaikan tesis ini, yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa tesis ini masih memiliki berbagai keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan sebagai bahan evaluasi dan perbaikan di masa mendatang. Penulis berharap penelitian ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi penelitian selanjutnya yang berkaitan dengan bidang yang dibahas dalam tesis ini.

Akhir kata, penulis mengucapkan terima kasih kepada seluruh pihak yang telah menjadi bagian dari perjalanan penyelesaian tesis ini. Semoga segala kebaikan dan bantuan yang telah diberikan mendapatkan balasan terbaik dari Allah SWT.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN SAMPUL ..... **i**
HALAMAN PENGESAHAN ..... **ii**
SURAT PERNYATAAN KEASLIAN ..... **iii**
NOTA DINAS PEMBIMBING ..... **iv**
ABSTRAK ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **xi**
DAFTAR TABEL ..... **xiv**
DAFTAR GAMBAR ..... **xv**
DAFTAR LAMPIRAN ..... **xvi**
DAFTAR SINGKATAN ..... **xvii**
BAB I PENDAHULUAN **1**
A. Latar Belakang 1
B. Rumusan Masalah 5
C. Batasan Masalah 6
D. Tujuan Penelitian 7
E. Manfaat Penelitian 7
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI **9**
A. Kajian Pustaka 9
B. Landasan Teori 20
1. Persyaratan Fungsional dan Nonfungsional 20
2. Agile User Story 21
3. Klasifikasi Teks Menggunakan BERT 22
4. Ketidakseimbangan Kelas 24
5. Cross-Entropy 26

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

6. Focal Loss 27
7. Strategi Pembobotan 28
8. Evaluasi Klasifikasi 30
9. Optimisasi dan Regularisasi Pelatihan 32
BAB III METODE PENELITIAN 35
A. Data Penelitian 35
1. Sumber dan Karakteristik Data 35
2. Distribusi dan Pemetaan Label 37
3. Pembagian Data 39
B. Alat dan Lingkungan 41
1. Lingkungan Komputasi 41
2. Model dan Bahan Penelitian 43
C. Tahapan Penelitian 44
1. Alur Penelitian 44
2. Persiapan dan Representasi Teks 45
3. Desain Eksperimen 46
4. Loss Function dan Pembobotan 48
5. Konfigurasi Pelatihan 51
6. Pemilihan Checkpoint Terbaik 53
7. Prosedur Evaluasi dan Analisis 53
BAB IV HASIL DAN PEMBAHASAN 56
A. Gambaran Hasil Evaluasi 56
B. Perbandingan Loss Function 59
C. Pengaruh Strategi Pembobotan 62
D. Analisis Performa Per Kelas 65
E. Analisis Confusion Matrix 70
F. Pembahasan Terpadu 76
1. Performa Cross-Entropy dan Focal Loss 76
2. Performa Kelas Mayoritas dan Minoritas 78

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

G. Keterbatasan Penelitian 80
BAB V PENUTUP 83
A. Kesimpulan 83
B. Saran 85
DAFTAR PUSTAKA 87
LAMPIRAN 92
CV/CP 102

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR TABEL

Tabel 2. 1	Penelitian terdahulu yang relevan	16
Tabel 3. 1	Definisi dan pemetaan label penelitian	38
Tabel 3. 2	Distribusi kelas pada keseluruhan dataset	39
Tabel 3. 3	Distribusi kelas pada training dan evaluation	40
Tabel 3. 4	Komponen perangkat lunak penelitian	42
Tabel 3. 5	Konfigurasi eksperimen	47
Tabel 3. 6	Konfigurasi pelatihan model	52
Tabel 4. 1	Hasil evaluasi	57
Tabel 4. 2	Perbandingan loss function	61
Tabel 4. 3	Pengaruh strategi pembobotan	64
Tabel 4. 4	Skor F1 per kelas pada seluruh konfigurasi	67
Tabel 4. 5	Arah kesalahan klasifikasi pada seluruh konfigurasi	71

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 3. 1	Alur pelaksanaan penelitian	45
Gambar 4. 1	Confusion matrix konfigurasi CE	72
Gambar 4. 2	Confusion matrix konfigurasi CE_CB_999	73
Gambar 4. 3	Confusion matrix konfigurasi FL_IF	74

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR LAMPIRAN

Lampiran 1	Classification Report Sepuluh Konfigurasi	92
Lampiran 2	Confusion Matrix Sepuluh Konfigurasi	97

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR SINGKATAN

ASCII	:	American Standard Code for Information Interchange
AUC	:	Area Under the Curve
BERT	:	Bidirectional Encoder Representations from Transformers
BNB	:	Binarized Naive Bayes
BoW	:	Bag of Words
CB	:	Class-balanced, pembobotan berdasarkan effective number of samples
CE	:	Cross-Entropy
CHI2	:	Chi-Squared, metode seleksi fitur berdasarkan statistik chi-kuadrat
[CLS]	:	Token khusus BERT yang representasi akhirnya digunakan untuk klasifikasi urutan
CM	:	Compatibility, label kelas kebutuhan kompatibilitas
CNN	:	Convolutional Neural Network
CPU	:	Central Processing Unit
CSV	:	Comma-Separated Values
F1	:	F1-score, rata-rata harmonik precision dan recall
FL	:	Focal Loss
FN	:	False Negative
FP	:	False Positive
FR	:	Functional Requirements, label kelas kebutuhan fungsional
GPU	:	Graphics Processing Unit
HPO	:	Hyperparameter Optimization
ID	:	Identifier, pengenal numerik kelas
IF	:	Inverse-frequency, pembobotan berdasarkan kebalikan frekuensi kelas
kNN	:	k-Nearest Neighbors
LLM	:	Large Language Model
LR	:	Logistic Regression pada kajian penelitian terdahulu
MCC	:	Matthews Correlation Coefficient
MN	:	Maintainability, label kelas kebutuhan kemudahan pemeliharaan
MNB	:	Multinomial Naive Bayes

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NFR	:	Non-Functional Requirements, kebutuhan nonfungsional
NLP	:	Natural Language Processing
PE	:	Performance, label kelas kebutuhan kinerja
PO	:	Portability, label kelas kebutuhan portabilitas
RE	:	Requirements Engineering
RL	:	Reliability, label kelas kebutuhan keandalan
S	:	Static, penanda pengali loss seragam 0,25 dalam kode konfigurasi
SE	:	Security, label kelas kebutuhan keamanan
[SEP]	:	Token khusus BERT sebagai pemisah atau penanda akhir urutan
SL	:	Supervised Learning
SSL	:	Semi-Supervised Learning
SVM	:	Support Vector Machine
TF-IDF	:	Term Frequency–Inverse Document Frequency
TN	:	True Negative
TP	:	True Positive
US	:	Usability, label kelas kebutuhan kemudahan penggunaan
USL	:	Unsupervised Learning
WCE	:	Weighted Cross-Entropy

Kode Konfigurasi Eksperimen

CE	:	Cross-Entropy tanpa pembobotan
CE_S	:	Cross-Entropy dengan pengali loss seragam 0,25
CE_IF	:	Cross-Entropy dengan inverse-frequency weighting
CE_CB_999	:	Cross-Entropy dengan class-balanced weighting, beta 0,999
CE_CB_9999	:	Cross-Entropy dengan class-balanced weighting, beta 0,9999
FL	:	Focal Loss tanpa pembobotan
FL_S	:	Focal Loss dengan pengali loss seragam 0,25
FL_IF	:	Focal Loss dengan inverse-frequency weighting
FL_CB_999	:	Focal Loss dengan class-balanced weighting, beta 0,999
FL_CB_9999	:	Focal Loss dengan class-balanced weighting, beta 0,9999

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB I
PENDAHULUAN

A. Latar Belakang

Persyaratan perangkat lunak tidak hanya menjelaskan fungsi yang harus disediakan oleh suatu sistem, tetapi juga atribut kualitas yang perlu dipenuhi ketika fungsi tersebut dijalankan. Atribut kualitas dinyatakan sebagai *non-functional requirements* (NFR), antara lain Compatibility, Maintainability, Performance, Portability, Reliability, Security, dan Usability. Dalam pengembangan Agile, kebutuhan lazim dituliskan secara ringkas dalam bentuk *user story*. Bentuk yang ringkas membantu komunikasi kebutuhan, tetapi informasi atribut kualitas tidak selalu dinyatakan melalui istilah yang eksplisit. (Gilson, Galster and Georis, 2019) menunjukkan bahwa atribut kualitas dalam *user story* dapat disampaikan secara implisit dan bahwa tingkat keberhasilan pengenalannya berbeda antar kategori. Kondisi tersebut menjadikan identifikasi NFR penting sekaligus menantang karena suatu teks perlu dipahami berdasarkan konteksnya, bukan hanya melalui pencocokan kata kunci.

Pertambahan jumlah *user story* dalam kumpulan kebutuhan membuat penelaahan manual terhadap setiap teks sulit dipertahankan sebagai satu-satunya cara pengelompokan. Klasifikasi otomatis dapat membantu memisahkan *functional requirements* (FR) dan kategori NFR secara konsisten pada kumpulan data yang besar. Namun, klasifikasi tersebut bukan sekadar persoalan membedakan FR dari NFR. Pada tugas multikelas satu-label, model harus menetapkan satu dari beberapa kelas yang tersedia sehingga kesalahan dapat terjadi antara FR dan suatu

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

kategori NFR maupun antarkategori NFR. Kajian (Binkhonain and Zhao, 2019) juga memperlihatkan bahwa penelitian klasifikasi kebutuhan menggunakan sumber data, jumlah kategori, dan prosedur evaluasi yang beragam. Oleh karena itu, performa suatu pendekatan perlu dibaca sesuai dengan objek klasifikasi dan rancangan eksperimennya.

Pendekatan klasifikasi kebutuhan berkembang dari algoritma pembelajaran mesin dengan fitur yang direkayasa secara manual menuju jaringan saraf dan model bahasa *pre-trained*. (Kurtanović and Maalej, 2017), (Abad *et al.*, 2017), serta (Canedo and Mendes, 2020) menunjukkan penggunaan representasi leksikal atau tekstual dengan algoritma klasik pada dokumen kebutuhan. Penelitian yang lebih dekat dengan konteks *user story* kemudian memanfaatkan representasi yang dipelajari dan *transfer learning* untuk mengenali atribut kualitas atau kebutuhan tertentu (Gilson, Galster and Georis, 2019; Casillo, Deufemia and Gravino, 2022). Perkembangan ini relevan karena model bahasa seperti BERT membentuk representasi kontekstual dan dapat disesuaikan untuk tugas klasifikasi melalui proses *fine-tuning* (Devlin *et al.*, 2019).

(Alhaizaey and Al-Mashari, 2025a) menyusun dataset berisi 10.006 *user story* dengan delapan kelas yang terdiri atas FR dan tujuh kategori NFR. Pada dataset tersebut, FR mencakup 64,53% data, sedangkan setiap kategori NFR memiliki proporsi yang jauh lebih kecil. Menggunakan dataset yang sama, (Alhaizaey and Al-Mashari, 2025b) melakukan *fine-tuning* BERT, DistilBERT, RoBERTa, dan XLNet dengan Focal Loss serta *Bayesian hyperparameter optimization*. Penelitian tersebut menunjukkan kelayakan model bahasa *pre-trained*

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

untuk klasifikasi delapan kelas. Meskipun demikian, fokus utamanya adalah perbandingan beberapa model dan pencarian konfigurasi masing-masing model, bukan perbandingan terarah antara Cross-Entropy dan Focal Loss dalam beberapa kondisi pembobotan pada satu model yang dipertahankan tetap.

Distribusi kelas yang tidak seimbang perlu diperhatikan karena ukuran performa agregat dapat lebih banyak dipengaruhi oleh kelas berfrekuensi tinggi. Accuracy yang tinggi, misalnya, tidak dengan sendirinya menunjukkan bahwa kelas minoritas telah dikenali dengan baik (He and Garcia, 2009). Pada dataset penelitian ini, dominasi FR dapat menutupi perbedaan performa tujuh kategori NFR apabila evaluasi hanya bertumpu pada satu metrik agregat yang dipengaruhi jumlah contoh. Sebaliknya, kelas-kelas minoritas juga tidak dapat diasumsikan memiliki karakteristik dan respons yang seragam. Penilaian karena itu perlu menggabungkan metrik yang memberi kontribusi setara kepada setiap kelas dengan metrik agregat lain, hasil per kelas, dan pola kesalahan klasifikasi.

Salah satu aspek pelatihan yang berkaitan dengan persoalan tersebut adalah cara menghitung kontribusi kesalahan model. Cross-Entropy memberikan dasar umum untuk klasifikasi multikelas, sedangkan Focal Loss mengurangi kontribusi contoh yang telah diprediksi dengan baik agar perhatian relatif lebih besar diberikan kepada contoh yang sulit (Lin *et al.*, 2017). Mekanisme Focal Loss berbeda dari pembobotan berdasarkan frekuensi kelas: tingkat kesulitan ditentukan oleh probabilitas prediksi, sedangkan *inverse-frequency weighting* dan *class-balanced weighting* mengubah kontribusi relatif kelas berdasarkan jumlah contohnya. *Class-balanced weighting*

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

menggunakan konsep *effective number of samples*, dan respons model terhadap parameter pembobotannya dapat berbeda menurut karakteristik data (Cui *et al.*, 2019). Perbedaan mekanisme tersebut memberi alasan untuk menelaah loss function dan pembobotan secara terpisah maupun dalam kombinasinya, tanpa mengasumsikan sejak awal bahwa satu perlakuan akan selalu lebih baik.

Penelitian (Alhaizaey and Al-Mashari, 2025b) menggunakan Focal Loss dengan alpha 0,25 dan gamma 2. Pada penerapan alpha sebagai satu skalar yang sama untuk seluruh contoh, seperti kondisi static dalam penelitian ini, alpha 0,25 hanya mengalikan setiap nilai *loss* dengan 0,25. Nilai *loss* per contoh dengan demikian menjadi seperempat dari nilai sebelum pengalian alpha, tetapi perbandingan relatif kontribusi antarcontoh dan antarkelas tetap sama. Penskalaan tersebut juga menskalakan gradien dan dapat memengaruhi dinamika optimisasi, sehingga tidak berarti bahwa hasil akhir pelatihan pasti identik. Namun, alpha seragam tidak secara khusus memberi perhatian lebih besar kepada kelas yang jumlah datanya lebih sedikit. Karena tugas penelitian mencakup delapan kelas dengan distribusi yang tidak seimbang, diperlukan pengujian pembobotan berbasis frekuensi setiap kelas melalui *inverse-frequency weighting* dan *class-balanced weighting*. Kedua strategi tersebut memungkinkan kontribusi relatif antarkelas berubah, berbeda dari alpha 0,25 yang diterapkan seragam.

Berdasarkan ruang kajian tersebut, penelitian ini memusatkan perhatian pada perbandingan empiris sepuluh konfigurasi yang dibentuk dari Cross-Entropy dan Focal Loss dengan lima kondisi pembobotan, yaitu tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999. Istilah static pada penelitian

ini merujuk pada pengali *loss* seragam sebesar 0,25 untuk seluruh contoh, bukan pembobotan relatif antarkelas. BERT-base-uncased digunakan sebagai model tetap. Konfigurasi dasar direplikasi semirip mungkin dari parameter terbaik penelitian (Alhaizaey and Al-Mashari, 2025b) melalui artikel, kode, dan konfigurasi model, kemudian tidak dioptimisasi ulang agar pengaruh konfigurasi loss dan pembobotan dapat diamati ketika faktor pelatihan lainnya dikendalikan sama.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah penelitian ini adalah sebagai berikut.

1. Bagaimana performa Cross-Entropy pada kondisi tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999 dalam klasifikasi delapan kelas *Agile User Story* menggunakan BERT-base-uncased?
2. Bagaimana performa Focal Loss pada kondisi tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999 dalam klasifikasi delapan kelas *Agile User Story* menggunakan BERT-base-uncased?
3. Bagaimana perbedaan performa Cross-Entropy dan Focal Loss pada kondisi pembobotan yang sama berdasarkan Macro F1 sebagai metrik utama serta Accuracy, Weighted F1, dan Matthews Correlation Coefficient sebagai metrik pelengkap?
4. Bagaimana perubahan precision, recall, dan F1 kelas mayoritas FR serta tujuh kelas minoritas NFR pada sepuluh konfigurasi tersebut, dan bagaimana pola kesalahan klasifikasinya berdasarkan *confusion matrix*?

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

C. Batasan Masalah

Untuk menjaga pembahasan tetap sesuai dengan rancangan eksperimen, penelitian ini dibatasi sebagai berikut.

1. Penelitian menggunakan dataset sekunder berisi 10.006 *user story* berbahasa Inggris yang dijelaskan oleh (Alhaizaey and Al-Mashari, 2025a) serta digunakan dalam penelitian klasifikasi (Alhaizaey and Al-Mashari, 2025b). Tugas penelitian berupa klasifikasi multikelas satu-label terhadap FR dan tujuh kategori NFR.
2. Model yang digunakan dibatasi pada BERT-base-uncased. Eksperimen membandingkan Cross-Entropy dan Focal Loss dalam lima kondisi pembobotan, yaitu tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999.
3. Konfigurasi dasar direplikasi semirip mungkin dari parameter terbaik penelitian (Alhaizaey and Al-Mashari, 2025b), digunakan sama pada seluruh eksperimen, dan tidak dioptimisasi ulang untuk setiap kombinasi loss dan pembobotan.
4. Data dibagi secara acak dan berstrata menjadi 80% *training set* dan 20% *evaluation set* menggunakan seed 42. Setiap konfigurasi dijalankan satu kali, sedangkan *evaluation set* yang sama digunakan untuk pemilihan *checkpoint* dan pelaporan hasil tanpa *test set* independen.
5. Macro F1 digunakan sebagai metrik utama dengan metrik agregat lain, hasil per kelas, dan *confusion matrix* sebagai pelengkap. Kesimpulan dibatasi pada dataset, pembagian data, model, dan

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

konfigurasi yang diuji serta tidak dimaksudkan sebagai bukti keunggulan universal atau signifikansi statistik.

D. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. mendeskripsikan dan membandingkan performa Cross-Entropy pada kondisi tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999 dalam klasifikasi delapan kelas *Agile User Story* menggunakan BERT-base-uncased;
2. mendeskripsikan dan membandingkan performa Focal Loss pada kondisi tanpa pembobotan, static, *inverse-frequency*, serta *class-balanced weighting* dengan beta 0,999 dan 0,9999 dalam klasifikasi delapan kelas *Agile User Story* menggunakan BERT-base-uncased;
3. menganalisis perbedaan performa Cross-Entropy dan Focal Loss pada kondisi pembobotan yang sama berdasarkan Macro F1, Accuracy, Weighted F1, dan Matthews Correlation Coefficient; dan
4. menganalisis perubahan precision, recall, dan F1 kelas mayoritas FR serta tujuh kelas minoritas NFR beserta pola kesalahan klasifikasinya berdasarkan *confusion matrix* pada sepuluh konfigurasi yang diuji.

E. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah memberikan bukti empiris mengenai perbedaan performa Cross-Entropy dan Focal Loss dalam lima kondisi pembobotan pada klasifikasi multikelas satu-label terhadap FR dan tujuh kategori NFR. Bukti tersebut memperluas pembahasan studi acuan melalui perbandingan yang secara khusus

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

memusatkan perhatian pada loss function dan pembobotan dalam satu pipeline BERT yang dikendalikan tetap. Kemudain harapanya dapat memberikan gambaran mengenai perbedaan fungsi pengali *static weighting*, *inverse-frequency weighting*, dan *class-balanced weighting* dalam rancangan klasifikasi yang tidak seimbang. Hasil agregat dan analisis per kelas dapat menjadi bahan pertimbangan dalam merancang eksperimen lanjutan yang menilai pertukaran performa antara kelas mayoritas dan kelas minoritas, dengan tetap memperhatikan keterbatasan dataset dan protokol evaluasi penelitian ini.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB V
PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian terhadap sepuluh konfigurasi BERT-base-uncased, kelompok Cross-Entropy menghasilkan Macro F1 sebesar 0,8705 pada konfigurasi tanpa pembobotan, 0,8822 pada static, 0,8692 pada *inverse-frequency*, 0,8907 pada *class-balanced weighting* beta 0,999, dan 0,8801 pada *class-balanced weighting* beta 0,9999. CE_CB_999 menghasilkan nilai tertinggi dalam kelompok ini sekaligus pada seluruh konfigurasi yang diamati, dengan Macro F1 0,8907, Accuracy 0,9341, Weighted F1 0,9344, dan MCC 0,8840. Dibandingkan CE tanpa pembobotan, CE_CB_999 meningkatkan Macro F1 sebesar 0,0202 skor, sedangkan CE_IF menjadi satu-satunya kondisi dalam kelompok Cross-Entropy yang menghasilkan Macro F1 lebih rendah daripada CE.

Pada kelompok Focal Loss, Macro F1 yang dihasilkan adalah 0,8743 pada konfigurasi tanpa pembobotan, 0,8819 pada static, 0,8801 pada *inverse-frequency*, 0,8852 pada *class-balanced weighting* beta 0,999, dan 0,8779 pada *class-balanced weighting* beta 0,9999. Keempat kondisi pembobotan tersebut menghasilkan Macro F1 lebih tinggi daripada FL tanpa pembobotan, dengan nilai tertinggi pada FL_CB_999. Namun, FL_S menghasilkan Accuracy dan Weighted F1 tertinggi dalam kelompok Focal Loss, sedangkan MCC FL_S dan FL_CB_999 hampir sama. Dengan demikian, konfigurasi dengan Macro F1 tertinggi tidak selalu menempati urutan tertinggi pada metrik pelengkap.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Perbandingan pada kondisi pembobotan yang sama menunjukkan bahwa Focal Loss menghasilkan Macro F1 lebih tinggi daripada Cross-Entropy pada kondisi tanpa pembobotan dan *inverse-frequency*, masing-masing dengan selisih 0,0037 dan 0,0110 skor. Cross-Entropy menghasilkan Macro F1 lebih tinggi pada kondisi static serta *class-balanced weighting* beta 0,999 dan 0,9999, masing-masing dengan selisih 0,0003, 0,0055, dan 0,0022 skor. Oleh karena itu, hasil yang diamati tidak menunjukkan bahwa salah satu *loss function* selalu lebih tinggi pada seluruh kondisi pembobotan. Kondisi static pada kedua *loss function* merupakan pengali *loss* seragam sebesar 0,25 yang tidak mengubah kontribusi relatif antarkelas, sehingga hasilnya tidak dapat ditafsirkan sebagai bukti efektivitas pembobotan kelas seperti pada *inverse-frequency* dan *class-balanced weighting*. Macro F1 yang digunakan sebagai metrik utama dalam tesis ini juga tidak disamakan dengan Weighted F1 yang dilaporkan oleh artikel acuan.

Kelas mayoritas FR mempertahankan F1 tertinggi di antara delapan kelas pada seluruh konfigurasi, dengan rentang 0,9570–0,9675, sedangkan tujuh kelas NFR menunjukkan respons yang berbeda-beda terhadap konfigurasi. CE_CB_999 meningkatkan F1 seluruh kelas NFR dibandingkan CE dan menghasilkan F1 tertinggi pada PO, SE, dan US, tetapi konfigurasi lain menghasilkan F1 tertinggi pada CM, MN, PE, dan RL. Pada *inverse-frequency weighting*, jumlah kesalahan NFR→FR berkurang, tetapi kesalahan FR→NFR meningkat; perubahan ini tampak pada menurunnya recall FR serta meningkatnya recall tetapi menurunnya precision US, sehingga F1 US pada CE_IF dan FL_IF justru menurun. Pertukaran FR–US menjadi pola kesalahan paling menonjol, disusul FR–SE. CE_CB_999 mencatat total kesalahan paling

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

sedikit, yaitu 132 dari 2.002 contoh, serta kesalahan antarkelas NFR paling sedikit, yaitu 44 contoh. Temuan tersebut menunjukkan bahwa perubahan performa kelas mayoritas dan minoritas melibatkan pertukaran precision–recall yang tidak seragam dan tidak seluruhnya tercermin oleh satu metrik agregat.

Seluruh kesimpulan tersebut bersifat deskriptif untuk hasil yang diamati. Setiap konfigurasi dijalankan satu kali dengan seed 42 menggunakan parameter dasar yang sama, sedangkan *evaluation set* yang sama digunakan untuk pemilihan *checkpoint* sekaligus pelaporan hasil. Oleh karena itu, perbedaan skor tidak menunjukkan signifikansi statistik, kestabilan lintas seed, ataupun keunggulan universal suatu konfigurasi.

B. Saran

Berdasarkan temuan dan keterbatasan penelitian, penelitian selanjutnya disarankan memisahkan data untuk pemilihan *checkpoint* dan pengujian akhir, misalnya melalui *validation set* dan *test set* independen. Pengujian pada pembagian data lain atau melalui prosedur validasi yang sesuai juga diperlukan untuk menilai apakah pola performa antarkonfigurasi tetap muncul pada data yang tidak digunakan dalam proses pemilihan model.

Setiap konfigurasi juga perlu dijalankan menggunakan beberapa seed. Hasil pengulangan dapat dilaporkan dalam bentuk ukuran pemusatan dan variasi serta, jika sesuai dengan rancangan analisis, diuji menggunakan metode statistik yang tepat. Langkah ini diperlukan untuk menilai kestabilan hasil dan membedakan pola yang konsisten dari variasi antar-run.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Cakupan pengujian dapat diperluas ke dataset *Agile User Story* lain, distribusi kelas yang berbeda, dan model bahasa selain BERT-base-uncased. Perluasan tersebut dapat digunakan untuk menilai apakah pola hubungan antara *loss function*, pembobotan, dan performa kelas mayoritas–minoritas juga ditemukan pada kondisi yang berbeda.

HPO dapat dilakukan secara terpisah untuk setiap kombinasi *loss function* dan kondisi pembobotan dengan ruang pencarian, anggaran komputasi, metrik pemilihan, dan prosedur evaluasi yang setara. Penelitian lanjutan juga dapat mengeksplorasi nilai beta *class-balanced weighting* yang lebih luas serta parameter Focal Loss. Rancangan ini perlu dibedakan dari penelitian sekarang yang menggunakan parameter dasar yang sama untuk seluruh konfigurasi agar pengaruh perubahan *loss function* dan pembobotan dapat dibandingkan dalam kondisi yang dikendalikan.

Selain itu, keluaran prediksi individual perlu disimpan bersama teks, label aktual, dan label prediksi. Data tersebut memungkinkan analisis kesalahan pada tingkat contoh untuk menelusuri karakteristik linguistik atau konteks semantik pada pertukaran FR–US, FR–SE, dan kesalahan antarkelas NFR, yang belum dapat dijelaskan hanya melalui metrik dan *confusion matrix* agregat.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR PUSTAKA

Abad, Z.S.H. *et al.* (2017) 'What Works Better? A Study of Classifying Requirements', *2017 IEEE 25th International Requirements Engineering Conference (RE)* [Preprint]. IEEE. Available at: https://doi.org/10.1109/RE.2017.36.

Alhaizaey, A. and Al-Mashari, M. (2025a) 'An Aggregated Dataset of Agile User Stories and Use Case Taxonomy for AI-Driven Research', *International Journal of Advanced Computer Science and Applications*, 16(9), pp. 250–259.

Alhaizaey, A. and Al-Mashari, M. (2025b) 'Automated Classification and Identification of Non-Functional Requirements in Agile-Based Requirements Using Pre-Trained Language Models', *IEEE Access*, 13, pp. 87401–87417. Available at: https://doi.org/10.1109/ACCESS.2025.3570359.

Alhaizaey, A. and Al-Mashari, M. (2026) 'Agile_Req_NFRs'. Available at: https://github.com/a-re22/Agile_Req_NFRs.

Amna, A.R. and Poels, G. (2022) 'Systematic Literature Mapping of User Story Research', *IEEE Access*, 10, pp. 51723–51746. Available at: https://doi.org/10.1109/ACCESS.2022.3173745.

Binkhonain, M. and Zhao, L. (2019) 'A Review of Machine Learning Algorithms for Identification and Classification of Non-Functional Requirements', *Expert Systems with Applications: X*, 1, p. 100001. Available at: https://doi.org/10.1016/j.eswax.2019.100001.

Buda, M., Maki, A. and Mazurowski, M.A. (2018) 'A Systematic Study of the Class Imbalance Problem in Convolutional Neural

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Networks', *Neural Networks*, 106, pp. 249–259. Available at: https://doi.org/10.1016/j.neunet.2018.07.011.

Canedo, E.D. and Mendes, B.C. (2020) 'Software Requirements Classification Using Machine Learning Algorithms', *Entropy*, 22(9), p. 1057. Available at: https://doi.org/10.3390/e22091057.

Casillo, F., Deufemia, V. and Gravino, C. (2022) 'Detecting Privacy Requirements from User Stories with NLP Transfer Learning Models', *Information and Software Technology*, 146, p. 106853. Available at: https://doi.org/10.1016/j.infsof.2022.106853.

Cawley, G.C. and Talbot, N.L.C. (2010) 'On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation', *Journal of Machine Learning Research*, 11(70), pp. 2079–2107. Available at: https://jmlr.org/papers/v11/cawley10a.html.

Cui, Y. *et al.* (2019) 'Class-Balanced Loss Based on Effective Number of Samples', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9268–9277.

Devlin, J. *et al.* (2019) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186.

Gilson, F., Galster, M. and Georis, F. (2019) 'Extracting Quality Attributes from User Stories for Early Architecture Decision Making', *2019 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pp. 129–136. Available at: https://doi.org/10.1109/ICSA-C.2019.00031.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

He, H. and Garcia, E.A. (2009) 'Learning from Imbalanced Data', *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp. 1263–1284. Available at: https://doi.org/10.1109/TKDE.2008.239.

Jurman, G., Riccadonna, S. and Furlanello, C. (2012) 'A Comparison of MCC and CEN Error Measures in Multi-Class Prediction', *PLoS ONE*, 7(8), p. e41882. Available at: https://doi.org/10.1371/journal.pone.0041882.

Kingma, D.P. and Ba, J. (2015) 'Adam: A Method for Stochastic Optimization', *3rd International Conference on Learning Representations (ICLR)* [Preprint]. Available at: https://arxiv.org/abs/1412.6980.

Kurtanović, Z. and Maalej, W. (2017) 'Automatically Classifying Functional and Non-Functional Requirements Using Supervised Machine Learning', *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pp. 490–495. Available at: https://doi.org/10.1109/RE.2017.82.

Lin, T.-Y. *et al.* (2017) 'Focal Loss for Dense Object Detection', *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988.

Loshchilov, I. and Hutter, F. (2019) 'Decoupled Weight Decay Regularization', *International Conference on Learning Representations (ICLR)* [Preprint]. Available at: https://arxiv.org/abs/1711.05101.

Micikevicius, P. *et al.* (2018) 'Mixed Precision Training', *International Conference on Learning Representations (ICLR)* [Preprint]. Available at: https://arxiv.org/abs/1710.03740.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Mosbach, M., Andriushchenko, M. and Klakow, D. (2021) 'On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines', *International Conference on Learning Representations (ICLR)* [Preprint]. Available at: https://arxiv.org/abs/2006.04884.

Opitz, J. and Burst, S. (2019) *Macro F1 and Macro F1*, *arXiv*. Available at: https://doi.org/10.48550/arXiv.1911.03347.

Paszke, A. *et al.* (2019) 'PyTorch: An Imperative Style, High-Performance Deep Learning Library', *Advances in Neural Information Processing Systems*. Available at: https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12(85), pp. 2825–2830. Available at: https://jmlr.org/papers/v12/pedregosa11a.html.

Raharjana, I.K., Siahaan, D. and Fatichah, C. (2021) 'User Stories and Natural Language Processing: A Systematic Literature Review', *IEEE Access*, 9, pp. 53811–53826. Available at: https://doi.org/10.1109/ACCESS.2021.3070606.

Rahman, K. *et al.* (2024) 'A Deep Learning Framework for Non-Functional Requirement Classification', *Scientific Reports*, 14, p. 3216. Available at: https://doi.org/10.1038/s41598-024-52802-0.

scikit-learn developers (2026) 'f1_score'. Available at: https://scikit-learn.org/1.6/modules/generated/sklearn.metrics.f1_score.html.

Srivastava, N. *et al.* (2014) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *Journal of Machine Learning*

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Research, 15(56), pp. 1929–1958. Available at: https://jmlr.org/papers/v15/srivastava14a.html.

Vaswani, A. *et al.* (2017) 'Attention Is All You Need', *Advances in Neural Information Processing Systems*. Available at: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Wolf, T. *et al.* (2020) 'Transformers: State-of-the-Art Natural Language Processing', *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, pp. 38–45. Available at: https://doi.org/10.18653/v1/2020.emnlp-demos.6.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA