PENGARUH MODEL MORFOLOGI DALAM MESIN PENERJEMAH STATISTIK BAHASA INDONESIA KE MINANGKABAU

SKRIPSI

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:
RIYAN HERNANDI
NIM D1041191028

FAKULTAS TEKNIK
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Riyan Hernandi

NIM : D1041191028

menyatakan bahwa dalam skripsi yang berjudul "PENGARUH MODEL MORFOLOGI DALAM MESIN PENERJEMAH STATISTIK BAHASA INDONESIA KE MINANGKABAU" tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 15 Juli 2025



Riyan Hernandi
NIM D1041191028

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS TEKNIK
Jalan Prof. Dr. H. Hadari Nawawi Pontianak 78124
Telepon (0561) 740186, WA: +6282152280907
Email : ft@untan.ac.id Website : http://teknik.untan.ac.id

HALAMAN PENGESAHAN

PENGARUH MODEL MORFOLOGI DALAM MESIN PENERJEMAH STATISTIK BAHASA INDONESIA KE MINANGKABAU

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:

Riyan Hernandi
NIM D1041191028

Telah dipertahankan di depan Penguji Skripsi pada tanggal 15 Juli 2025 dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi:

Ketua,

Prof. Dr. Herry Sujaini, S.T., M.T.
NIP 196806291997021001

Penguji Utama,

Dr. Arif Bijaksana Putra Negara, S.T., M.T.
NIP 197208081998021002

Sekretaris,

H. Hengky Anra, S.T., M.Kom.
NIP 197503251999031005

Penguji Pendamping,

Rina Septiriana, ST, M.Cs.
NIP 198709232020122001

Pontianak, 15 Juli 2025
Dekan,

Dr.-Ing Ir. Slamet Widodo, M.T., IPM
NIP 196712231992031002

HALAMAN PERSEMBAHAN

Skripsi ini kupersembahkan dengan penuh rasa syukur kepada keluargaku:

Kepada **Ayahku, Erlan (Almarhum)** dan **Ibuku, Herlenawati**, yang selalu menjadi sumber doa, dukungan, dan kasih sayang dalam setiap langkahku. Kepada **Abangku, Rizky Husain**, serta **Adikku, Rivaldi Harja Basari**, yang senantiasa memberikan semangat dan kebersamaan yang berarti.

Tak lupa, rasa terima kasih yang mendalam juga kupersembahkan kepada **para dosen** yang telah membimbing dan mengajarkan ilmu dengan penuh dedikasi.

Serta kepada **teman-teman perkuliahan**, yang telah menjadi bagian dari perjalanan akademikku dan memberikan warna dalam setiap langkah yang kulalui. Semoga ilmu yang diperoleh dapat bermanfaat, baik bagi diri sendiri maupun masyarakat luas, sehingga kita dapat berkontribusi dalam kehidupan bermasyarakat.

KATA PENGANTAR

Puji dan syukur penulis panjatkan Kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis bisa menyelesaikan penelitian dan penulisan tugas akhir dengan judul "Pengaruh Model Morfologi Dalam Mesin Penerjemah Statistik Bahasa Indonesia Ke Minangkabau" ini diselesaikan dengan baik.

Penulis dalam menyelesaikan penulisan tugas akhir ini banyak mendapat bantuan dari berbagai pihak, maka dari itu pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr.-Ing. Ir. Slamet Widodo, M.T., IPM, selaku Dekan Fakultas Teknik Universitas Tanjungpura Pontianak,
2. Bapak Dr. Yus Sholva, S.T., M.T., selaku Ketua Jurusan Fakultas Teknik Universitas Tanjungpura,
3. Bapak M. Azhar Irwansyah, S.T., M.Eng., selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Tanjungpura,
4. Bapak Prof. Dr. Herry Sujaini, S.T., M.T., dan Bapak H. Hengky Anra, S.T., M.Kom., selaku pembimbing utama dan pembimbing pendamping yang telah membimbing penulis dalam menyelesaikan penulisan tugas akhir ini.
5. Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T., dan Ibu Rina Septiriana, ST, M.Cs., selaku penguji utama dan penguji pendamping yang telah memberikan masukan serta saran untuk perbaikan tugas akhir ini.
6. Ibu Anggi Perwitasari, S.T., M.T., selaku Dosen Pembimbing Akademik,

Dalam penulisan tugas akhir ini, penulis menyadari bahwa banyak terdapat kekurangan dan masih membutuhkan kritik, saran serta masukan yang konstruktif dan pengembangan lanjutan atas penelitian ini untuk memberi nilai tambah atas manfaat penelitian.

Pontianak, 15 Juli 2025
Penulis,

Riyan Hernandi
D1041191028

ABSTRAK

Bahasa Minangkabau memiliki sistem morfologi yang berbeda dari bahasa Indonesia, terutama dalam penggunaan imbuhan dan pembentukan kata. Perbedaan ini menjadi tantangan dalam penerjemahan otomatis karena dapat menghambat mesin dalam menangkap nuansa bahasa secara akurat. Sistem penerjemahan yang ada masih mengalami keterbatasan dalam memahami struktur bahasa daerah, sehingga hasil terjemahan bisa saja kehilangan makna asli. Oleh karena itu, penelitian ini berfokus pada pengaruh morfologi bahasa Minangkabau terhadap akurasi penerjemahan otomatis sebagai satu diantara upaya untuk mendukung pelestarian bahasa daerah di era digital. Metode yang digunakan adalah *Statistical Machine Translation* (SMT) bahasa Indonesia ke Minangkabau. Penelitian ini menguji pengaruh *tagging* morfologi afiksasi pada korpus Minangkabau terhadap hasil terjemahan. Hasil eksperimen menunjukkan bahwa penerapan *tagging* hanya pada *Translation Model* (TM) tetapi *Language Model* (LM) tidak menerapkan *tagging,* menyebabkan penurunan skor BLEU dari 62,86% menjadi 48,4%. Namun, ketika tagging diterapkan pada TM dan LM dengan 5-gram, skor BLEU meningkat menjadi 62,88% (naik 0,032%). Peningkatan skor tertinggi dicapai saat tagging diterapkan pada TM dan LM dengan 4-gram, yaitu mencapai 62,95% (naik 0,095% dibandingkan model tanpa *tagging* dengan TM dan LM 5-gram).

Kata kunci: Bahasa Indonesia, Bahasa Minangkabau, BLEU, *Language Model* ,Mesin Penerjemah Statistik, Morfologi Afiksasi, *Tagging, Translation Model*

ABSTRACT

The Minangkabau language has a morphological system that differs from Indonesian, particularly in the use of affixes and word formation. These differences present challenges in machine translation, as they can hinder the system's ability to accurately capture linguistic nuances. Existing translation systems still face limitations in understanding regional language structures, which can result in a loss of original meaning in translation. Therefore, this study focuses on the influence of Minangkabau morphology on the accuracy of machine translation as one of the efforts to support the preservation of regional languages in the digital era. The method used is Statistical Machine Translation (SMT) from Indonesian to Minangkabau. This study investigates the impact of affixal morphological tagging on the Minangkabau corpus in translation output. Experimental results show that applying tagging only to the Translation Model (TM) while the Language Model (LM) does not use tagging leads to a decrease in BLEU score from 62.86% to 48.4%. However, when tagging is applied to both TM and LM with a 5-gram setting, the BLEU score increases to 62.88% (an increase of 0.032%). The highest improvement is achieved when tagging is applied to both TM and LM using 4-gram, reaching 62.95% (an increase of 0.095% compared to the model without tagging using TM and LM with 5-gram).

Keyword: Affix Morphology, BLEU, Indonesian, Minangkabau Language, Language Model, Statistical Machine Translation, Tagging , Translation Model

DAFTAR ISI

Halaman Pernyataan ..... **ii**
Halaman Pengesahan ..... **iii**
Halaman Persembahan ..... **iv**
Kata Pengantar ..... **v**
Abstrak ..... **vi**
Abstract ..... **vii**
Daftar Isi ..... **viii**
Daftar Gambar ..... **x**
Daftar Tabel ..... **xi**
Daftar Kode Program ..... **xii**
Daftar Lampiran ..... **xiii**
Daftar Notasi ..... **xiv**
Bab I Pendahuluan **1**
1.1 Latar Belakang 1
1.2 Perumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Pembatasan Masalah 3
1.5 Sistematika Penulisan 4
Bab II Tinjauan Pustaka **5**
2.1 Penelitian Terkait 5
2.2 Landasan Teori 6
2.2.1 Korpus 6
2.2.2 Moses 7
2.2.3 Mesin Penerjemah Statistik 7
2.2.4 *Language Model* 7
2.2.5 *Translation Model* 8
2.2.6 *Decoder* 9
2.2.7 Evaluasi Otomatis 9
2.2.8 Morfologi Afiksasi Bahasa Minangkabau 10
2.2.9 *Stemming* 12
Bab III Metodologi Penelitian **13**
3.1 Metodologi Penelitian 13
3.1.1 Pengumpulan Data 14
3.1.2 Praproses Korpus 14
3.1.3 Pembangunan *Statistical Machine Translation* 15
3.1.4 Implementasi *Statistical Machine Translation* Tanpa *Tag*

Imbuhan....15
3.1.5 *Stemming* dan *Tagging* Morfologi Afiksasi....17
3.1.6 Implementasi *Statistical Machine Translation* Dengan *Tag* Imbuhan....22
3.1.7 Pengujian Hasil Terjemahan....24
3.1.8 Analisis Hasil Pengujian....25
3.1.9 Penarikan Kesimpulan....26

Bab IV Hasil dan Analisis....27

4.1 Hasil Penelitian....27
4.1.1 Hasil Pengumpulan Data....27
4.1.2 Praproses Korpus....27
4.1.3 Membangun Statistical Machine Translation (SMT)....28
4.1.4 Implementasi KenLM untuk Pemodelan Bahasa (*Language Model*)....28
4.1.5 Implementasi Giza++ untuk Pemodelan Translasi (*Translation Model*)....29
4.1.6 *Stemming* dan *Tagging* Morfologi Afiksasi Korpus Bahasa Minangkabau....32
4.1.7 Implementasi *Statistical Machine Translation* (SMT) Dengan *Tag* Imbuhan....34
4.1.8 *Decoding*....38
4.1.9 Pengujian Hasil Terjemahan....41
4.1.10 Analisis Hasil Pengujian....43

Bab V Kesimpulan dan Saran....45

5.1 Kesimpulan....45
5.2 Saran....45

Daftar Pustaka....46

DAFTAR GAMBAR

Gambar 2.1 Contoh penerjemahan dengan phrase-based translation model [Tanuwijaya, 2009].....8
Gambar 3.1 Metodologi penelitian.....13
Gambar 3.2 Arsitektur Statistical Machine Translation bahasa Indonesia ke Minangkabau (Dimodifikasi dari Asparilla et al., 2018).....16
Gambar 3.3 Diagram alir dari proses tagging morfologi afiksasi korpus bahasa Minangkabau.....18
Gambar 3.4 Diagram alir dari stemming dan tagging morfologi prefiks, sufiks dan konfiks korpus bahasa Minangkabau.....19
Gambar 3.5 Diagram alir dari tagging morfologi infiks korpus bahasa Minangkabau.....20
Gambar 3.6 Diagram alir mengganti kata dengan format tag afiksasi.....22
Gambar 3.7 Skenario penelitian lima mesin translasi.....23
Gambar 3.8 Diagram alir pengembalian kata terjemahan ke format tanpa tag afiksasi.....25
Gambar 4.1 Dokumen model translasi.....30
Gambar 4.2 Diagram perbandingan skor BLEU.....42

DAFTAR TABEL

Tabel 3.1 Skenario lima mesin translasi berdasarkan penggunaan korpus target (Minangkabau) 24
Tabel 4.1 Contoh isi korpus bahasa Indonesia dan bahasa Minangkabau. 27
Tabel 4.2 Contoh dari dokumen lm_tanpatag_5g.arpa 29
Tabel 4.3 Contoh isi dokumen "ind.vcb" Mesin Translasi 1.....31
Tabel 4.4 Contoh isi dokumen "min.vcb" Mesin Translasi 1 31
Tabel 4.5 Contoh isi dokumen word alignment “ind-min.A3.final” Mesin Translasi 1 32
Tabel 4.6 Contoh isi dokumen ”min_tag.vcb” serta perbandingannya sebelum tagging.....33
Tabel 4.*7* Contoh dari dokumen lm_dengantag_3g.arpa 35
Tabel 4.8 Contoh dari dokumen lm_dengantag_4g.arpa 35
Tabel 4.9 Contoh dari dokumen lm_dengantag_5g.arpa 35
Tabel 4.10 Contoh isi dokumen word alignment “ind-min.A3.final” Mesin Translasi 5 37
Tabel 4.11 Kalimat Sumber (Bahasa Indonesia).....39
Tabel 4.12 Kalimat Terjemahan Mesin Translasi 1 40
Tabel 4.13 Contoh 20 BLEU Score Per Kalimat dari Masing-Masing Mesin Translasi 43

DAFTAR KODE PROGRAM

Kode Program 4.1 Perintah model bahasa menggunakan KenLM berformat ASCII 28
Kode Program 4.2 Perintah model bahasa menggunakan KenLM berformat binary 29
Kode Program 4.3 Perintah membangun model translasi untuk Mesin Translasi 1 (MT1) dengan menggunakan Giza++ 30
Kode Program 4.4 Perintah model bahasa 3-gram korpus target dengan tag imbuhan menggunakan KenLM berformat ASCII 35
Kode Program 4.5 Perintah model bahasa 3-gram korpus target dengan tag imbuhan menggunakan KenLM berformat binary.....36
Kode Program 4.6 Perintah membangun model translasi untuk Mesin Translasi 2 (MT2) dengan menggunakan Giza++.....36
Kode Program 4.7 Perintah membangun model translasi untuk Mesin Translasi 2 (MT3) dengan menggunakan Giza++.....37
Kode Program 4.8 Perintah decoding Mesin Translasi 1.....38

DAFTAR LAMPIRAN

Lampiran A PROGRAM PYTHON PEMBERSIHAN ENTITAS HTML A-1
Lampiran B INSTALASI MESIN PENERJEMAH STATISTik B-1
Lampiran C PROGRAM PYTHON *VOCABULARY* C-1
Lampiran D PROGRAM PYTHON *STEMMING* DAN *TAGGING* PADA PREFIKS DAN SUFIKS *VOCABULARY*..... D-1
Lampiran E PROGRAM PYTHON STEMMING DAN TAGGING INFIKS VOCABULARY.....E-1
Lampiran F HASIL PENGHAPUSAN MANUAL DARI HASIL *TAGGING* OTOMATIS.....F-1
Lampiran G PROGRAM PYTHON *TAGGING* SETIAP KATA PADA KORPUS BAHASA MINANGKABAU G-1
Lampiran H PROGRAM PYTHON PENGEMBALIAN SETIAP KATA BERFORMAT *TAGGING*..... H-1
Lampiran I PROGRAM PYTHON SACREBLEU I-1
Lampiran J CONTOH HASIL TERJEMAHAN MESIN TRANSLASI 2 .....J-1
Lampiran K CONTOH HASIL TERJEMAHAN MESIN TRANSLASI 3 ..... K-1
Lampiran L CONTOH HASIL TERJEMAHAN MESIN TRANSLASI 4 .....L-1
Lampiran M CONTOH HASIL TERJEMAHAN MESIN TRANSLASI 5 .....M-1
Lampiran N PROGRAM PYTHON SACREBLEU PER KALIMAT N-1
Lampiran O CONTOH TERJEMAHAN DENGAN NILAI SKOR BLEU YANG BERAGAM O-1

DAFTAR NOTASI

$P(w_n)$	=	Probabilitas kata ke-n
$P(w_n \mid w_{(n-1)})$	=	Probabilitas kata ke-n terhadap kata ke-(n-1)
$P(w_n \mid w_{n-2,n-1})$	=	Probabilitas kata ke-n terhadap kata ke-(n-2) dan kata ke-(n-1)
$\hat{T}$	=	Hasil terjemahan
$arg_T \max$	=	Nilai maksimal
$P(T \mid S)$	=	Probabilitas bahasa target terhadap bahasa sumber
$P(S \mid T)$	=	Probabilitas bahasa sumber terhadap bahasa target
$P(S)$	=	Probabilitas bahasa sumber
$P(T)$	=	Probabilitas bahasa target

BAB I
PENDAHULUAN

1.1 Latar Belakang

"Bahasa adalah alat komunikasi yang terorganisasi dalam bentuk satuan-satuan, seperti kata, kelompok kata, klausa, dan kalimat yang diungkapkan baik secara lisan maupun tulis" (Wiratno & Santosa, 2014:2). Melalui bahasa, manusia dapat mengekspresikan diri, menyampaikan pendapat, pemikiran, bahkan kritik. Hal tersebut sesuai dengan Halliday dalam Rabiah (2018:5) yang men gemukakan bahwa fungsi ke enam dari bahasa adalah fungsi personal, yakni penggunaan bahasa untuk mengungkapkan perasaan, emosi, kepribadian, dan sebagainya.

Indonesia memiliki sekitar 700 bahasa daerah, dengan 24 bahasa daerah berstatus terancam punah karena penutur bahasa berusia di atas 20 tahun dan orang tua tidak berbicara bahasa daerah kepada anak-anaknya, serta 5 bahasa daerah berstatus kritis karena penutur bahasa berusia dia atas 40 tahun dengan jumlah yang sangat sedikit (Anindyatri & Mufidah, 2020). Sehingga diperlukannya upaya agar jumlah bahasa daerah yang punah tidak terus bertambah. Salah satu upaya tersebut dengan cara digitalisasi bahasa daerah (Mantri, 2021:78).

Satu di antara upaya pelestarian bahasa daerah melalui digitalisasi adalah dengan memanfaatkan bidang keilmuan Pemrosesan Bahasa Alami atau yang dalam bahasa Inggris dikenal sebagai *Natural Language Processing* (NLP). NLP merupakan bidang studi dan penerapan teknologi yang bertujuan agar komputer mampu memahami dan mengolah bahasa manusia, baik dalam bentuk teks maupun lisan. Melalui penelitian di bidang ini, dikembangkan berbagai metode dan perangkat yang memungkinkan sistem komputer memahami struktur dan makna bahasa alami, sehingga dapat digunakan untuk menjalankan berbagai fungsi yang bermanfaat (Chowdhary, 2020:1). Salah satu contoh dari metode tersebut adalah Mesin Penerjemah Statistik atau yang dalam bahasa disebut *Statistical Machine Translation* (SMT), yang menurut Koehn (2009) SMT adalah pendekatan untuk penerjemah otomatis yang berdasarkan pada analisis statistik dan probabilitas dari korpus teks paralel dalam dua atau lebih bahasa.

Jarob et al. (2016:83) menyatakan bahwa proses tagging kata per kata dengan mengambil kata dasar dan imbuhannya bisa meningkatkan nilai dari akurasi mesin penerjemah statistik bahasa Indonesia - bahasa Dayak Taman sebesar 0,36% dengan pengujian otomatis oleh BLEU. Hal ini membuktikan bahwa morfologi bahasa terutama pada jenis afiksasi (imbuhan) dapat mempengaruhi nilai akurasi dari SMT.

Secara etimologis, istilah morfologi dalam bahasa Indonesia berasal dari kata *morphology* dalam bahasa Inggris. Istilah tersebut berasal dari dua morfem, yaitu *morph* yang berarti 'bentuk' dan *logy* yang berarti 'ilmu'. Morfologi dalam linguistik adalah kajian mengenai bentuk-bentuk kata dan proses pembentukan kata. Dengan kata lain, setiap bentuk bahasa yang berupa seluk beluk kata menjadi objek kajian morfologi, seperti berbagai contoh kata yang mengalami proses pembentukan, misalnya dari kata dasar *desain* menjadi *mendesain*, *terdesain*, dan lain-lain (Suparno, 2015:5).

Secara umum, afiksasi yang ada pada bahasa Minangkabau bisa dibilang cukup mirip dengan imbuhan yang ada dalam bahasa Indonesia, yaitu sama-sama memiliki jenis imbuhan prefiks (awalan), infiks (sisipan) dan sufiks (akhiran). Hanya saja, imbuhan yang ada di antara mereka berbeda, contohnya jika di Indonesia ada imbuhan awalan ”me-”, di bahasa Minangkabau imbuhannya menjadi ”ma-”. Sebagai contoh, kalau dari kata Indonesia ”memakan”, maka di bahasa Minangkabau menjadi ”mamakan” (Be et al., 1979).

Perkembangan teknologi telah membuat mesin penerjemahan otomatis semakin canggih. Meskipun *Neural Machine Translation* (NMT) dinilai memiliki performa akurasi yang tinggi. Namun, beberapa penelitian melaporkan bahwa SMT masih lebih baik dibandingkan NMT ketika sumber daya yang ada terbatas. Mereka merekomendasikan penggunaan SMT daripada NMT untuk Bahasa dengan Sumber Daya Sangat Rendah (Extremely Low-Resource Languages/ELRL). SMT direkomendasikan karena mengandalkan model statistik yang mampu menangani kumpulan data kecil secara efisien dan menghasilkan terjemahan yang andal untuk ELRL (Sujaini et al., 2023). Akan tetapi tantangan utama masih ada, terutama ketika menghadapi perbedaan struktur dan bentuk kata antara bahasa sumber dan

bahasa target. Contohnya, bahasa Minangkabau dengan sistem morfologi yang unik. Jika hal ini tidak diperhatikan, hasil terjemahan bisa saja kehilangan makna aslinya.

Oleh karena itu, tujuan utama penelitian ini adalah untuk mengetahui pengaruh morfologi afiksasi dalam bahasa Minangkabau terhadap akurasi mesin penerjemah statistik dari bahasa Indonesia ke bahasa Minangkabau. Dengan demikian, pengembangan sistem penerjemahan otomatis dari bahasa Indonesia ke bahasa Minangkabau lebih memperhatikan aspek morfologi afiksasi bahasa Minangkabau, sehingga sistem penerjemah mampu mencerminkan nuansa dan makna asli bahasa dengan lebih baik.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang sebelumnya, maka permasalahan yang didapatkan adalah apakah morfologi afiksasi yang ada pada bahasa Minangkabau berpengaruh terhadap akurasi mesin penerjemah berbasis statistik bahasa Indonesia ke bahasa Minangkabau.

1.3 Tujuan Penelitian

Tujuan dari penelitian adalah untuk mengetahui seberapa besar pengaruh morfologi afiksasi yang ada pada bahasa Minangkabau terhadap akurasi mesin penerjemah berbasis statistik bahasa Indonesia ke bahasa Minangkabau.

1.4 Pembatasan Masalah

Adapun batasan masalah pada penelitian ini adalah seperti berikut.

1. Mesin penerjemah yang digunakan dalam penelitian ini adalah mesin penerjemah berbasis statistik dengan menggunakan perangkat lunak Moses.
2. Penerjemahan dilakukan satu arah, yaitu dari bahasa Indonesia ke bahasa Minangkabau.
3. Morfologi bahasa yang dibahas pada penelitian ini adalah morfologi bahasa jenis afiksasi, yang mencakup prefiks, sufiks, infiks dan konfiks dalam bahasa Minangkabau.

1.5 Sistematika Penulisan

Berikut ini adalah sistematika dari penulisan tugas akhir ini, yang disusun dalam lima bab.

Bab I Pendahuluan, berisi latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah dan sistematika penulisan.

Bab II Tinjauan Pustaka, berisi tentang hasil-hasil penelitian yang didapat oleh peneliti terdahulu dan landasan teori yang berisi definisi korpus, moses, mesin penerjemah statistik, *language model*, *translation model*, *decoder, automatic evaluation,* morfologi afiksasi bahasa Minangkabau, dan *stemming*.

Bab III Metodologi Penelitian, adalah bab yang berisi tentang metode penelitian, bahan penelitian, alat yang digunakan, diagram alir penelitian dan arsitektur sistem.

Bab IV Hasil dan Analisis adalah bab yang berisi data hasil percobaan, pengamatan, observasi dan sebagainya yang telah dirancang pada Bab III. Setiap hasil yang didapatkan akan dianalisis untuk mengarah kepada suatu kesimpulan.

Bab V Kesimpulan dan Saran adalah bab yang berisi kesimpulan dari penelitian yang telah dilakukan serta saran atau rekomendasi untuk perbaikan, pengembangan atau kelengkapan penelitian yang telah dilakukan.