



DAFTAR ISI

| | |
|--|-------------|
| HALAMAN JUDUL | i |
| HALAMAN PENGESAHAN | ii |
| HALAMAN PERNYATAAN | iii |
| HALAMAN PERSEMBAHAN | iv |
| HALAMAN MOTTO | v |
| PRAKATA | vi |
| DAFTAR ISI | viii |
| DAFTAR TABEL | xi |
| DAFTAR GAMBAR | xii |
| INTISARI | xiv |
| ABSTRACT | xv |
| I PENDAHULUAN | 1 |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Tujuan Penelitian | 3 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Metodologi Penelitian | 4 |
| 1.7 Sistematika Penulisan | 5 |
| II TINJAUAN PUSTAKA | 7 |
| 2.1 Awal Pengembangan Mesin Penerjemahan | 7 |
| 2.2 Pengembangan Mesin Penerjemahan Berbasis <i>Neural</i> | 8 |
| 2.3 Mesin Penerjemah untuk Domain Bahasa di Indonesia | 9 |



| | |
|---|-----------|
| III LANDASAN TEORI | 13 |
| 3.1 Mesin Penerjemah | 13 |
| 3.2 Jaringan Saraf Tiruan | 14 |
| 3.2.1 <i>Multi Layer Perceptron</i> | 15 |
| 3.2.2 Pelatihan | 15 |
| 3.3 <i>Word Embedding</i> | 16 |
| 3.3.1 Word2Vec | 17 |
| 3.4 Transformer | 18 |
| 3.4.1 Mekanisme <i>Attention</i> | 19 |
| 3.4.2 <i>Scaled Dot-Product Attention</i> | 20 |
| 3.4.3 <i>Multi-Head Attention</i> | 20 |
| 3.4.4 <i>Position-wise Feed-Forward Network</i> | 20 |
| 3.4.5 <i>Positional Encoding</i> | 21 |
| 3.4.6 Lapisan Normalisasi | 21 |
| 3.5 BLEU | 22 |
| IV ANALISIS DAN PERANCANGAN | 23 |
| 4.1 Deskripsi Penelitian | 23 |
| 4.2 Studi Literatur dan Analisis Masalah | 23 |
| 4.3 Perancangan Akuisisi Data | 24 |
| 4.4 Perancangan Pra-pemrosesan Data | 26 |
| 4.5 Perancangan Pemisahan Data, Prapemrosesan Lanjut, dan Pembagian Data | 26 |
| 4.6 Perancangan Penerjemahan | 27 |
| 4.7 Perancangan Model dan Pengaturan Hiperparameter | 28 |
| 4.7.1 Embedding | 28 |
| 4.7.2 Mekanisme Multi-Head Attention | 30 |
| 4.7.3 Enkoder | 32 |
| 4.7.4 Dekoder | 33 |
| 4.7.5 Masked Multi-Head Self-Attention | 33 |
| 4.7.6 Cross-Attention | 34 |
| 4.7.7 Lapisan Output | 35 |
| 4.8 Perancangan Pelatihan Model | 35 |
| 4.9 Perancangan Inferensi | 38 |
| 4.10 Perancangan Evaluasi Model | 39 |



| | | |
|------------|--|-----------|
| 4.11 | Strategi Analisis | 40 |
| V | IMPLEMENTASI | 41 |
| 5.1 | Lingkungan Implementasi | 41 |
| 5.2 | Akuisisi Data | 41 |
| 5.3 | Pra-pemrosesan Data | 44 |
| 5.3.1 | Persiapan Data Setelah Scrape | 44 |
| 5.3.2 | Persiapan Data untuk Pelatihan dan Inferensi | 47 |
| 5.4 | Pembangunan Model | 50 |
| 5.5 | Pelatihan dan Validasi Model | 54 |
| 5.6 | Inferensi dan Evaluasi | 61 |
| VI | HASIL DAN PEMBAHASAN | 67 |
| 6.1 | Hasil Akuisisi Data dan Prapemrosesan Awal | 67 |
| 6.2 | Hasil Prapemrosesan Akhir dan Pembagian Batch | 70 |
| 6.3 | Analisis Hasil Proses Pelatihan | 71 |
| 6.3.1 | Evaluasi Model dengan Dimensi Kecil | 73 |
| 6.3.2 | Evaluasi Model dengan Dimensi Besar | 73 |
| 6.3.3 | Pengaruh Jumlah <i>Attention Head</i> | 75 |
| 6.4 | Evaluasi Hasil Inferensi dengan Beam Search | 75 |
| 6.5 | Hasil Penerjemahan | 76 |
| 6.6 | Perbandingan Penelitian | 82 |
| VII | PENUTUP | 84 |
| 7.1 | Kesimpulan | 84 |
| 7.2 | Saran | 85 |
| | DAFTAR PUSTAKA | 86 |
| A | PERCOBAAN TOKENISASI SUBKATA DENGAN SENTENCEPIECE | 92 |