

# DAFTAR ISI

|                                                 |             |
|-------------------------------------------------|-------------|
| <b>LEMBAR PERSETUJUAN</b>                       | <b>i</b>    |
| <b>LEMBAR PENGESAHAN</b>                        | <b>ii</b>   |
| <b>PERNYATAAN LEMBAR KEASLIAN SKRIPSI</b>       | <b>iii</b>  |
| <b>PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI</b> | <b>iv</b>   |
| <b>ABSTRAK</b>                                  | <b>v</b>    |
| <b>ABSTRACT</b>                                 | <b>vi</b>   |
| <b>KATA PENGANTAR</b>                           | <b>vii</b>  |
| <b>DAFTAR ISI</b>                               | <b>ix</b>   |
| <b>DAFTAR LAMPIRAN</b>                          | <b>xii</b>  |
| <b>DAFTAR GAMBAR</b>                            | <b>xiii</b> |
| <b>DAFTAR TABEL</b>                             | <b>xiv</b>  |
| <b>DAFTAR SIMBOL</b>                            | <b>xv</b>   |
| <b>BAB I PENDAHULUAN</b>                        | <b>1</b>    |
| 1.1 Latar Belakang                              | 1           |
| 1.2 Rumusan Masalah                             | 4           |
| 1.3 Batasan Masalah                             | 4           |
| 1.4 Tujuan Penelitian                           | 5           |
| 1.5 Metode Penelitian                           | 6           |
| 1.6 Sistematika Penulisan                       | 7           |
| <b>BAB II LANDASAN TEORI</b>                    | <b>8</b>    |
| 2.1 <i>Natural Language Processing</i>          | 8           |
| 2.2 Model Bahasa                                | 9           |
| 2.2.1 Formulasi Probabilitas Model Bahasa       | 9           |

|                |                                                                      |           |
|----------------|----------------------------------------------------------------------|-----------|
| 2.2.2          | Model N-Gram dan Asumsi Markov . . . . .                             | 11        |
| 2.2.3          | Keterbatasan Model Bahasa Berbasis Frekuensi . . . . .               | 13        |
| 2.3            | Neural Probabilistic Language Model . . . . .                        | 15        |
| 2.4            | <i>Likelihood</i> dan <i>Maximum Likelihood Estimation</i> . . . . . | 17        |
| 2.4.1          | <i>Negative Log-Likelihood</i> dan <i>Cross-Entropy</i> . . . . .    | 18        |
| 2.5            | Representasi Kata dan <i>Word Embedding</i> . . . . .                | 19        |
| 2.6            | Konteks Lokal . . . . .                                              | 20        |
| 2.7            | Arsitektur Dasar NPLM . . . . .                                      | 21        |
| 2.8            | Komponen Arsitektur NPLM Modern . . . . .                            | 23        |
| 2.8.1          | <i>Global Context</i> . . . . .                                      | 24        |
| 2.8.2          | <i>Average Pooling</i> . . . . .                                     | 24        |
| 2.8.3          | <i>Kernel Aggregator</i> . . . . .                                   | 25        |
| 2.8.4          | <i>Feed-Forward Network</i> . . . . .                                | 27        |
| 2.8.5          | <i>Residual Connection</i> . . . . .                                 | 28        |
| 2.8.6          | <i>Layer Normalization</i> . . . . .                                 | 28        |
| 2.8.7          | <i>Dropout</i> . . . . .                                             | 28        |
| 2.8.8          | <i>Adaptive Softmax</i> . . . . .                                    | 29        |
| 2.8.9          | Optimisasi Adam . . . . .                                            | 30        |
| 2.9            | Evaluasi Model Bahasa . . . . .                                      | 32        |
| 2.9.1          | <i>Perplexity</i> . . . . .                                          | 32        |
| 2.10           | <i>Overfitting</i> dan <i>Underfitting</i> . . . . .                 | 32        |
| 2.11           | Representasi Semantik dalam Ruang Vektor . . . . .                   | 35        |
| 2.11.1         | Hipotesis Distribusional . . . . .                                   | 35        |
| 2.11.2         | Cosine Similarity . . . . .                                          | 36        |
| 2.11.3         | <i>Nearest Neighbor</i> . . . . .                                    | 38        |
| <b>BAB III</b> | <b>METODOLOGI PENELITIAN . . . . .</b>                               | <b>39</b> |
| 3.1            | Alur Penelitian . . . . .                                            | 39        |
| 3.2            | Sumber dan Karakteristik Dataset . . . . .                           | 40        |
| 3.2.1          | Struktur Dataset . . . . .                                           | 40        |
| 3.2.2          | Statistik Deskriptif Dataset . . . . .                               | 41        |
| 3.3            | Prapemrosesan Data dan Pembentukan Kosakata . . . . .                | 42        |
| 3.3.1          | Tahapan Prapemrosesan Teks . . . . .                                 | 42        |
| 3.3.2          | Pembentukan Kosakata . . . . .                                       | 42        |
| 3.4            | Pembagian Data dan Pembentukan Sampel Konteks-Target . . . . .       | 43        |
| 3.4.1          | Pembentukan Context Window . . . . .                                 | 43        |
| 3.5            | Arsitektur Model NPLM yang Digunakan . . . . .                       | 44        |
| 3.6            | Konfigurasi Pelatihan . . . . .                                      | 45        |

|               |                                                                  |           |
|---------------|------------------------------------------------------------------|-----------|
| 3.6.1         | Analisis Sensitivitas OFAT . . . . .                             | 45        |
| 3.6.2         | Ruang Pencarian Hiperparameter . . . . .                         | 46        |
| 3.6.3         | Hiperparameter yang Ditetapkan . . . . .                         | 46        |
| 3.6.4         | Validasi Interaksi Antar-Hiperparameter . . . . .                | 46        |
| 3.7           | Studi Ablasi Arsitektur . . . . .                                | 47        |
| 3.8           | Prosedur Pelatihan Model . . . . .                               | 47        |
| 3.9           | Evaluasi Model . . . . .                                         | 48        |
| 3.10          | Analisis Distribusi Probabilitas . . . . .                       | 48        |
| 3.11          | Analisis Representasi Semantik . . . . .                         | 48        |
| 3.12          | Alat dan Lingkungan Penelitian . . . . .                         | 49        |
| <b>BAB IV</b> | <b>HASIL DAN PEMBAHASAN . . . . .</b>                            | <b>50</b> |
| 4.1           | Deskripsi Data dan Hasil Prapemrosesan . . . . .                 | 50        |
| 4.1.1         | Karakteristik Korpus . . . . .                                   | 50        |
| 4.1.2         | Hasil Prapemrosesan dan Pembentukan Kosakata . . . . .           | 51        |
| 4.2           | Hasil Pelatihan dan Validasi Model . . . . .                     | 52        |
| 4.2.1         | Analisis Sensitivitas Hiperparameter . . . . .                   | 52        |
| 4.2.2         | Validasi Interaksi Antar-Hiperparameter . . . . .                | 54        |
| 4.2.3         | Evaluasi dan Pemilihan Arsitektur Model . . . . .                | 56        |
| 4.3           | Analisis Distribusi Probabilitas dan Entropi . . . . .           | 57        |
| 4.4           | Representasi Semantik ( <i>Embedding</i> ) . . . . .             | 62        |
| 4.4.1         | Analisis <i>Cosine Similarity</i> . . . . .                      | 62        |
| 4.4.2         | Analisis Tetangga Terdekat ( <i>Nearest Neighbor</i> ) . . . . . | 63        |
| 4.4.3         | Visualisasi Ruang <i>Embedding</i> . . . . .                     | 64        |
| <b>BAB V</b>  | <b>KESIMPULAN DAN SARAN . . . . .</b>                            | <b>67</b> |
| 5.1           | Kesimpulan . . . . .                                             | 67        |
| 5.2           | Saran . . . . .                                                  | 68        |
|               | <b>DAFTAR PUSTAKA . . . . .</b>                                  | <b>69</b> |
|               | <b>LAMPIRAN . . . . .</b>                                        | <b>73</b> |