

DAFTAR ISI

| | |
|-----------------------------------------------------------------------|-----|
| TESIS | i |
| HALAMAN PENGESAHAN..... | i |
| PERNYATAAN..... | ii |
| MOTO DAN PERSEMBAHAN | iii |
| PRAKATA..... | iv |
| DAFTAR ISI..... | vi |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL..... | xvi |
| INTISARI..... | xix |
| <i>ABSTRACT</i> | xx |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Tujuan | 4 |
| 1.5 Manfaat | 4 |
| BAB II TINJAUAN PUSTAKA..... | 6 |
| 2.1 Penelitian Terdahulu | 6 |
| 2.2 Rumusan Klaim Evaluatif..... | 8 |
| BAB III LANDASAN TEORI..... | 16 |
| 3.1 <i>Machine Learning</i> | 16 |
| 3.2 <i>Preprocessing</i> | 16 |
| 3.2.1 Tokenisasi..... | 17 |
| 3.2.2 <i>Word Stemming</i> | 18 |
| 3.2.3 Penghapusan <i>Stop Word</i> | 19 |
| 3.2.4 Pengonversian Kata ke Huruf Kecil..... | 20 |
| 3.3 Jaringan Saraf Tiruan (<i>Neural Networks</i>) | 21 |
| 3.4 <i>Natural Language Processing (NLP)</i> | 22 |
| 3.4.1 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> | 22 |
| 3.4.2 Word2Vec | 24 |
| 3.4.3 FastText..... | 31 |
| 3.4.4 Perbandingan Word Embedding | 33 |
| 3.5 Klasifikasi Model..... | 34 |
| 3.5.1 Decision Tree | 34 |
| 3.5.2 Random Forest | 38 |
| 3.6 Evaluasi Model | 41 |
| 3.6.1 Akurasi | 42 |
| 3.6.2 Recall..... | 42 |
| 3.6.3 F1-Score | 43 |
| 3.6.4 Presisi | 43 |
| 3.6.5 <i>Mean Squared Error (MSE)</i> | 44 |
| 3.6.6 <i>Mean Absolute Error (MAE)</i> | 44 |
| 3.6.7 Korefisien Determinasi (R²) | 45 |
| 3.7 Sistem Peradilan dan Putusan Pengadilan | 45 |

| | | |
|------------------------------------|-----------------------------------------------------------------------------------|-----|
| 3.7.1 | Sistem Peradilan di Indonesia | 45 |
| 3.7.2 | Jenis-Jenis Putusan Pengadilan | 46 |
| 3.7.3 | Faktor-Faktor yang Mempengaruhi Putusan Pengadilan | 46 |
| 3.7.4 | Dokumen Putusan Pengadilan..... | 46 |
| BAB IV METODOLOGI PENELITIAN | | 49 |
| 4.1 | Alur Penelitian | 49 |
| 4.2 | Pengumpulan Data | 52 |
| 4.3 | <i>Labeling Data</i> | 55 |
| 4.4 | <i>Data Cleaning (Preprocessing Data)</i> | 56 |
| 4.5 | Pelatihan Model | 57 |
| 4.5.1 | <i>Tools dan libraries</i> yang di gunakan | 57 |
| 4.5.2 | Representasi Teks dengan NLP..... | 57 |
| 4.5.3 | Pengembangan model dengan Random Forest..... | 59 |
| 4.5.4 | <i>Skenario penelitian</i> | 62 |
| 4.5.5 | Ilustrasi Input dan Output Model Prediksi | 67 |
| 4.6 | Komparasi Hasil..... | 72 |
| BAB V IMPLEMENTASI..... | | 73 |
| 5.1 | Spesifikasi <i>Hardware</i> dan <i>Software</i> | 73 |
| 5.2 | Implementasi dan Praproses Data | 73 |
| 5.2.1 | Ekstraksi Fitur dari XML ke CSV | 73 |
| 5.2.2 | Labeling Data | 74 |
| 5.2.3 | Pembersihan data..... | 75 |
| 5.3 | Implementasi pemodelan dan eksperimen | 78 |
| 5.3.1 | Implementasi berdasarkan pendekatan agregatif dan segmentasi | 79 |
| 5.3.2 | Implementasi eksperimen parameter pada metode TF-IDF, Word2Vec dan FastText | 84 |
| BAB VI HASIL DAN PEMBAHASAN | | 94 |
| 6.1 | <i>Preprocessing Data</i> | 94 |
| 6.1.1 | Perubahan Setelah Preprocessing | 94 |
| 6.1.2 | Analisis Statistik Deskriptif..... | 96 |
| 6.2 | EDA (<i>Exploratory Data Analysis</i>)..... | 98 |
| 6.2.1 | Distribusi data berdasarkan variable dependen (Y) Kategori Hukuman | 98 |
| 6.2.2 | Frekuensi kata pada dokumen | 100 |
| 6.2.3 | Frekuensi kata per kategori hukuman..... | 101 |
| 6.2.4 | Statistik Kata Unik VS Total Token..... | 103 |
| 6.2.5 | Statistik <i>rare words</i> (kata yang jarang muncul) | 105 |
| 6.2.6 | Sanity Check..... | 106 |
| 6.3 | Hasil Eksperimen Perbandingan Representasi Agregatif | 108 |
| 6.3.1 | Klasifikasi (Kategori Hukuman) | 109 |
| 6.3.2 | Regresi (Lama Hukuman) | 150 |
| 6.4 | Hasil Eksperimen Perbandingan Representasi Segmentatif | 164 |
| 6.4.1 | Klasifikasi (Kategori Hukuman) | 164 |
| 6.4.2 | Regresi (Lama Hukuman) | 194 |
| 6.5 | Perbandingan hasil akhir representasi pada agregatif dan segmenetatif. | 209 |
| 6.6 | Eksperimen Ablasi Fitur Bagian Dokumen | 213 |

| | | |
|------------------------------------|-----------------------------------------------------------|-----|
| 6.5.1 | Hasil Ablasi pada Klasifikasi (Kategori Hukuman) | 213 |
| 6.5.2 | Hasil Ablasi pada Regresi (Lama Hukuman) | 224 |
| 6.6 | Analisis Kasus dengan Performa Kurang Baik dan Baik | 226 |
| 6.6.1 | Kasus Salah Klasifikasi (Performa Buruk) | 226 |
| 6.6.2 | Kasus Klasifikasi Benar (Performa Baik) | 227 |
| 6.6.3 | Kasus Error Besar pada Regresi (Performa Buruk) | 228 |
| 6.6.4 | Kasus Error Besar pada Regresi (Performa Buruk) | 229 |
| 6.7 | Evaluasi klaim eksperimen | 230 |
| 6.6.1 | Evaluasi Klaim 1 | 230 |
| 6.6.2 | Evaluasi Klaim 2 | 231 |
| BAB VII KESIMPULAN DAN SARAN | | 233 |
| 7.1 | Kesimpulan | 233 |
| 7.2 | Saran | 233 |
| DAFTAR PUSTAKA | | 235 |