



DAFTAR ISI

| | |
|-------------------------------------|-------------|
| HALAMAN PENGESAHAN | i |
| HALAMAN PERNYATAAN | ii |
| HALAMAN MOTTO | iii |
| PRAKATA | iv |
| DAFTAR ISI | v |
| DAFTAR TABEL | viii |
| DAFTAR GAMBAR | ix |
| INTISARI | xi |
| ABSTRACT | xii |
| I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Tujuan Penelitian | 4 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Metodologi Penelitian | 4 |
| 1.7 Sistematika Penulisan | 5 |
| II TINJAUAN PUSTAKA | 7 |
| III LANDASAN TEORI | 11 |
| 3.1 <i>Text Preprocessing</i> | 11 |
| 3.1.1 <i>Tokenization</i> | 11 |
| 3.1.2 <i>Stopwords</i> | 12 |
| 3.1.3 <i>Stemming</i> | 12 |
| 3.2 Representasi Kata | 13 |
| 3.3 Word2Vec | 14 |



| | | |
|-----------|---|-----------|
| 3.3.1 | <i>Continous bag-of-words (CBOW)</i> | 14 |
| 3.3.2 | <i>Skip-gram</i> | 14 |
| 3.4 | Language Model | 16 |
| 3.4.1 | <i>N-gram Model</i> | 16 |
| 3.4.2 | <i>Smoothing Model</i> | 17 |
| 3.5 | <i>Perplexity</i> | 19 |
| IV | ANALISIS DAN PERANCANGAN | 20 |
| 4.1 | Deskripsi Umum Penelitian | 20 |
| 4.2 | <i>Dataset</i> | 22 |
| 4.2.1 | <i>Author obfuscation dataset</i> | 23 |
| 4.2.2 | <i>Author verification dataset</i> | 23 |
| 4.3 | Perancangan Sistem | 24 |
| 4.3.1 | <i>Author Masking</i> | 24 |
| 4.3.2 | <i>Modulizing text</i> | 26 |
| 4.3.3 | <i>Charpos</i> | 27 |
| 4.3.4 | <i>Contractions expansion</i> | 27 |
| 4.3.5 | <i>Preprocessing</i> | 28 |
| 4.3.6 | <i>Word frequency</i> | 29 |
| 4.3.7 | <i>Part of speech (POS) tagging</i> | 29 |
| 4.3.8 | <i>Word2Vec</i> | 29 |
| 4.3.9 | <i>Normalized document</i> | 30 |
| 4.4 | <i>Truth data</i> | 31 |
| 4.5 | Pengujian <i>Author Obfuscation</i> | 32 |
| 4.5.1 | <i>Safety</i> | 32 |
| 4.5.2 | <i>Soundness</i> | 34 |
| 4.5.3 | <i>Sensibleness</i> | 36 |
| V | IMPLEMENTASI | 37 |
| 5.1 | Spesifikasi Hardware dan Software | 37 |
| 5.2 | Hirarki Modul Implementasi Kode | 37 |
| 5.3 | Modul <i>Obfuscation.py</i> | 37 |
| 5.3.1 | Meng- <i>import</i> pustaka yang dibutuhkan | 37 |
| 5.3.2 | <i>Load</i> data <i>Word2Vec</i> | 38 |
| 5.3.3 | Fungsi <i>clean_document()</i> | 39 |
| 5.3.4 | Fungsi <i>replace_contractions</i> | 39 |



| | | |
|---------------------------------|--|-----------|
| 5.3.5 | <i>Modulizing text</i> | 39 |
| 5.3.6 | Menghitung <i>Word Frequency</i> | 40 |
| 5.3.7 | Fungsi <i>Charpos</i> | 40 |
| 5.3.8 | Membuat <i>truth data</i> | 41 |
| 5.3.9 | <i>Word tokenization</i> | 41 |
| 5.3.10 | <i>POS Tagging</i> | 43 |
| 5.3.11 | Melakukan pengecekan <i>stopwords</i> | 43 |
| 5.3.12 | Fungsi <i>get_noun_verb()</i> | 43 |
| 5.3.13 | Fungsi <i>get_vector()</i> | 44 |
| 5.3.14 | Fungsi <i>replace_data()</i> | 44 |
| 5.3.15 | Fungsi <i>make_document()</i> | 45 |
| 5.3.16 | Fungsi <i>normalize_document()</i> | 46 |
| 5.3.17 | Fungsi <i>build_dataset()</i> | 46 |
| 5.4 | Modul <i>logger.py</i> | 46 |
| 5.4.1 | Meng- <i>import</i> pustaka yang diperlukan | 47 |
| 5.4.2 | Fungsi <i>setup_logging()</i> | 47 |
| VI HASIL DAN PEMBAHASAN | | 49 |
| 6.1 | Hasil Pengujian <i>Safety</i> | 49 |
| 6.1.1 | GLAD (<i>Groningen Lightweight Authorship Detection</i>) | 49 |
| 6.2 | Hasil Pengujian <i>Soundness</i> | 50 |
| 6.2.1 | Hasil Pengujian Parameter 2-gram | 51 |
| 6.2.2 | Hasil Pengujian Parameter 3-gram | 53 |
| 6.3 | Hasil Pengujian <i>Sensibleness</i> | 55 |
| 6.4 | Hasil <i>Generate</i> Dokumen | 57 |
| VII KESIMPULAN DAN SARAN | | 60 |
| 7.1 | Kesimpulan | 60 |
| 7.2 | Saran | 60 |
| DAFTAR PUSTAKA | | 61 |