

DAFTAR ISI

| | |
|--|-----|
| HALAMAN JUDUL..... | ii |
| HALAMAN PENGESAHAN..... | iii |
| PERNYATAAN BEBAS PLAGIASI..... | iv |
| PRAKATA..... | v |
| DAFTAR ISI..... | iii |
| DAFTAR GAMBAR..... | iv |
| DAFTAR TABEL..... | v |
| INTISARI..... | vi |
| ABSTRACT..... | vii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang Masalah..... | 1 |
| 1.2 Rumusan Masalah..... | 4 |
| 1.3 Batasan Masalah..... | 4 |
| 1.4 Tujuan..... | 5 |
| 1.5 Manfaat..... | 5 |
| BAB II TINJAUAN PUSTAKA..... | 6 |
| BAB III LANDASAN TEORI..... | 13 |
| 3.1 Text Mining..... | 13 |
| 3.2 Web Scraping..... | 13 |
| 3.3 Text Preprocessing..... | 14 |
| 3.3.1 Case Folding..... | 14 |
| 3.3.2 Tokenizing..... | 14 |
| 3.3.3 Filtering (Stop Word Removal)..... | 15 |
| 3.3.4 Stemming..... | 15 |
| 3.3.5 Named Entity Recognition..... | 15 |
| 3.4 Feature Extraction..... | 16 |
| 3.5 BiLSTM-CNN..... | 16 |

| | | |
|------------------------------------|-----------------------------------|----|
| 3.5.1 | Word Embbeding Word2vec..... | 17 |
| 3.5.2 | Char Embedding (CNN) | 18 |
| 3.5.3 | BiLSTM | 18 |
| 3.5.4 | Fully Connected Layer | 20 |
| 3.6 | Topic Modeling..... | 21 |
| 3.6.1 | LDA | 22 |
| 3.7 | Evaluasi | 26 |
| 3.7.1 | F1-Score (BiLSMT-CNN) | 26 |
| 3.7.2 | Silhouette Coefficient (LDA)..... | 27 |
| BAB IV METODOLOGI PENELITIAN | | 28 |
| 4.1 | Tahapan Penelitian | 28 |
| 4.2 | Analisis dan Rancangan | 30 |
| 4.3 | Data Gathering | 32 |
| 4.4 | Data Preprocessing..... | 35 |
| 4.5 | Model NER (BiLSTM-CNN) | 40 |
| 4.6 | Model LDA | 48 |
| 4.7 | Manajemen Data | 51 |
| 4.8 | Visualisasi | 52 |
| 4.9 | Evaluasi BiLSTM-CNN..... | 53 |
| 4.10 | Evaluasi LDA..... | 53 |
| 4.11 | Alat..... | 53 |
| BAB V IMPLEMENTASI..... | | 55 |
| 5.1 | Data Gathering | 55 |
| 5.2 | Data Preprocessing..... | 61 |
| 5.3 | Proses Pelatihan Model NER | 63 |
| 5.4 | Proses Pemodelan Topik LDA..... | 64 |
| 5.5 | Evaluasi | 65 |
| BAB VI HASIL DAN PEMBAHASAN | | 68 |
| 6.1 | Hasil Data <i>Gathering</i> | 68 |

| | | |
|------------------------------------|----------------------------------|-----|
| 6.2 | Hasil Penyusunan Dataset | 68 |
| 6.3 | Hasil Data Preprocessing | 69 |
| 6.4 | Hasil Eksperimen Model | 70 |
| 6.5 | Model BiLSTM-CNN | 71 |
| 6.6 | Model LDA | 75 |
| 6.7 | Analisis Data | 79 |
| 6.7.1 | Ganjar Pranowo | 82 |
| 6.7.2 | Anies Baswedan | 86 |
| 6.7.3 | Puan Maharani | 91 |
| 6.7.4 | Erick Thohir | 94 |
| 6.7.5 | Prabowo Subianto | 98 |
| 6.7.6 | Kata kunci pajak | 102 |
| 6.7.7 | Ganjar OR PDIP | 109 |
| 6.7.8 | Ganjar AND Prabowo | 111 |
| 6.7.9 | Ganjar AND Piala Dunia U20 | 113 |
| BAB VII KESIMPULAN DAN SARAN | | 117 |
| 7.1. | Kesimpulan | 117 |
| 7.1. | Saran | 118 |
| DAFTAR PUSTAKA | | 119 |
| LAMPIRAN | | 122 |