

ABSTRAK

NAMED ENTITY RECOGNITION (NER) BAHASA INDONESIA BERBASIS MULTI CLASS CLASSIFICATION

Oleh

VINDO RIZKIYANTO

Penelitian ini bertujuan untuk menghasilkan sistem yang dapat mengekstraksi informasi berupa data teks Bahasa Indonesia dan mendeskripsikan hasil implementasi *Named Entity Recognition* berbasis *Multi Class Classification*. Metode yang digunakan adalah *extreme programming*. Penelitian ini juga melakukan beberapa tahapan seperti studi literatur, pembuatan model, perancangan sistem, dan pengujian sistem. Jumlah kata yang diperoleh dari hasil pengumpulan data adalah 8017 kata. Dalam tahapan pengumpulan data dilakukan dengan bantuan *library BeautifulSoup* untuk proses *scraping website* dengan melakukan *import request* untuk mendapatkan *link* berita yang akan di *scraping* dan *html5lib* untuk mendapatkan konten dari *website* dengan *parsing html*. Total entitas label yang diperoleh dari hasil pelabelan data adalah 1197 entitas label. Dalam penelitian ini pelabelan data dibagi menjadi 9 kelompok. Pelatihan data dilakukan dengan menerapkan 6 skenario. Hasil skenario yang didapatkan pada pelatihan data menghasilkan *precision*, *recall*, dan *f1-score*. Dalam tahapan proses pembagian data, penelitian ini menggunakan 2 skenario pembagian data, pertama yaitu 70% *training data* dan 30% *testing data*, kemudian pada *training data* dibagi menjadi 70% *training data* dan 30% *validation data*. Pada skenario kedua data dibagi menjadi 60% *training data* dan 40% *testing data*, kemudian pada *training data* akan dibagi menjadi 60% *training data* dan 40% *validation data*. Berdasarkan hasil dari skenario yang telah dilakukan pada proses evaluasi kinerja model dapat diketahui bahwa skenario 6 dengan pembagian 60% data *training*, dan 40% data *testing* menggunakan *hyperparameter epoch* 100, *batch size* 1000, dan *learn rate* 0.001, skenario ini merupakan konfigurasi *hyperparameter* bawaan *library spaCy* dan terbukti memiliki hasil nilai *precision*, *recall*, dan *f1-score* terbaik dibandingkan dengan skenario lainnya. Sistem yang telah dibuat diuji menggunakan metode *black box testing*, dengan hasil yang didapatkan sesuai yang diharapkan berdasarkan skenario uji yang di lakukan.

Kata kunci: *Extreme Programming, Named Entity Recognition, library BeautifulSoup dan Multi Class Classification.*