

INTISARI

PEMODELAN TOPIK UNTUK Mencari Isu Tentang Tokoh Politik di Portal Berita Indonesia Menggunakan LATENT DIRICHLET ALLOCATION DAN NAMED ENTITY RECOGNITION

Oleh
Aziz Muslim
21/486154/PPA/06238

Sebagai pengumpul informasi, Badan Intelijen Negara (BIN) bertugas untuk memetakan isu-isu yang sedang berkembang di masyarakat. Tujuannya untuk melakukan pencegahan atas aksi atau tindakan yang dapat merugikan negara maupun rakyat. Seiring berjalannya waktu, volume dan kecepatan pertumbuhan portal berita begitu masif, menimbulkan permasalahan, pasalnya, proses pemetaan saat ini masih dilakukan secara manual yang menyita waktu dan membutuhkan banyak tenaga ahli untuk melakukannya. Untuk itulah, perlu dikembangkan sebuah sistem pemetaan isu dari media berita online secara otomatis untuk memantau perkembangan isu berita saat ini. Hasil yang diperoleh, dapat digunakan oleh pembuat kebijakan untuk salah satu rujukan informasi dasar dalam membuat kebijakan.

Dalam penelitian ini, metode baru dikembangkan untuk memetakan berita dan isu terkait tokoh politik dan kata kunci tertentu dari berita berbahasa Indonesia. Berita dikumpulkan melalui teknik *web scraping* dari portal berita *online* selama 2022-2023. Data yang diperoleh kemudian diekstraksi dengan teknik *Named Entity Recognition* (NER) yang menggabungkan *Bidirectional LSTM* (BiLSTM) dan *Convolutional Neural Network* (CNN) untuk klasifikasi entitas dalam berita. Selanjutnya, pemodelan topik dilakukan menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi isu terkait tokoh politik dan kata kunci tertentu.

Hasil dari penelitian ini menunjukkan model NER BiLSTM-CNN memiliki akurasi sebesar 98,59% dari dataset yang dibangun sejumlah 4.548 kalimat. Selain itu berhasil melakukan ekstraksi NER pada 48.180 artikel Bahasa Indonesia. Berhasil melakukan ujicoba pemodelan topik dengan metode LDA pada korpus data besar dengan jumlah 2.015.645 token. Dengan metode ini, diharapkan BIN dapat lebih efisien dan efektif dalam memantau dan memetakan isu tokoh politik dari portal berita *online*.

Kata Kunci: NER, Pemodelan Topik, *Bidirectional Long Short-term Memory* (BiLSTM), *Convolutional Neural Network* (CNN), *Latent Dirichlet Allocation* (LDA).

ABSTRACT

TOPIC MODELING TO DISCOVER ISSUE REGARDING POLITICAL FIGURES IN INDONESIAN NEWS PORTALS USING LATENT DIRICHLET ALLOCATION AND NAMED ENTITY RECOGNITION

By
Aziz Muslim
21/486154/PPA/06238

As an information collector, the State Intelligence Agency (BIN) is tasked with mapping issues that are currently developing in society. The aim is to prevent actions that can harm the state and the people. As time goes by, the volume and speed of growth of news portals is massive, causing problems, because the current mapping process is still done manually which is time-consuming and requires a lot of experts to do it. For this reason, it is necessary to develop an automatic issue mapping system from online news media to monitor the development of current issues. The results obtained can be used by policy makers as a reference for basic information in making policies.

In this research, a new method was developed to map news and issues related to political figures and specific keywords from Indonesian language news articles. News was collected using web scraping techniques from online news portals during 2022-2023. The gathered data was then extracted using the Named Entity Recognition (NER) technique, which combines Bidirectional LSTM (BiLSTM) and Convolutional Neural Network (CNN) for entity classification within the news. Subsequently, topic modeling was conducted using Latent Dirichlet Allocation (LDA) to identify issues associated with political figures and specific keywords.

The results of this study indicate that the BiLSTM-CNN NER model has an accuracy of 98.59% from a dataset consisting of 4,548 sentences. Moreover, it successfully extracted NER from 48,180 Indonesian language articles. It successfully tested topic modeling using the LDA method on a large corpus containing 2,015,645 tokens. With this method, it is hoped that [BIN] can be more efficient and effective in monitoring and mapping issues related to political figures from online news portals.

Keyword: NER, Pemodelan Topik, *Bidirectional Long Short-term Memory* (BiLSTM), *Convolutional Neural Network* (CNN), *Latent Dirichlet Allocation* (LDA).