# Ektraksi Entitas Organisasi Dari Artikel Berita Politik Menggunakan Metode *Bidirectional Long Short Term Memory* Dan *Conditional Random Filds*

**Ahmad Su`ud1, Husni2, Doni Abdul Fatah3**

1Afiliasi Penulis 1 (institusi asal penulis 1), Indonesia

2Afiliasi Penulis 2 (institusi asal penulis 2), Indonesia

3Afiliasi Penulis 2 (institusi asal penulis 3), Indonesia

Email (*penulis 1 saja*): pertama@afiliasi.ac.id

*Received* 30 November 201x**|** *Revised* 30 Desember 201x**|** *Accepted* 30 Januari 201x

## Abstrak

Informasi politik menjadi semakin penting dalam kehidupan masyarakat di era digital, di mana berita politik, terutama artikel-artikel di Indonesia, berfungsi sebagai dasar pemahaman terhadap isu-isu politik yang sedang berlangsung. Namun, derasnya arus informasi membuat pencarian dan analisis data tentang entitas organisasi politik secara manual menjadi sulit dan memakan waktu. Oleh karena itu, diperlukan prosedur otomatis yang efisien untuk mengekstraksi entitas dari artikel berita politik di Indonesia. *Named Entity Recognition* (NER) adalah komponen krusial dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi entitas seperti nama orang, organisasi, waktu, dan lokasi dalam teks. Penelitian ini bertujuan untuk mengembangkan metode ekstraksi entitas organisasi menggunakan algoritma NER dengan metode *Bidirectional Long Short-Term Memory* (BiLSTM) dan *Conditional Random Fields* (CRF) pada 1202 artikel berita politik Indonesia yang mencakup 5439 kalimat dan 125.954 kata. Waktu komputasi yang diperlukan untuk melatih model ini adalah 1665 detik, atau sekitar 27,75 menit. Hasil penelitian menunjukkan bahwa model BiLSTM-CRF cukup baik dalam mengekstraksi entitas organisasi dengan nilai *Precision 85.16%, Recall 68.05%, F1-Score 74.38%,* dan *Accuracy 98.71%.*

**Kata kunci**: Artikel Berita, Ekstraksi Entitas, *Named Entity Recognition* *Bidirectional LSTM*, *Conditional Random Fields*, *Natural Language Processing*, *K-Fold Cross Validation.*

## Abstract

*Political information has become increasingly important in people's lives in the digital era, where political news, especially articles in Indonesia, serves as a basis for understanding ongoing political issues. However, the rapid flow of information makes manually searching and analysing data about political organisation entities difficult and time-consuming. Therefore, an efficient automated procedure is needed to extract entities from political news articles in Indonesia. Named Entity Recognition (NER) is a crucial component in Natural Language Processing (NLP) that aims to identify entities such as names of people, organisations, times, and locations in text. This research aims to develop an organisational entity extraction method using NER algorithm with Bidirectional Long Short-Term Memory (BiLSTM) and Conditional Random Fields (CRF) methods on 1202 Indonesian political news articles that include 5439 sentences and 125,954 words. The computation time required to train this model is 1665 seconds, or about 27.75 minutes. The results show that the BiLSTM-CRF model is good enough in extracting organisational entities with Precision 85.16%, Recall 68.05%, F1-Score 74.38%, and Accuracy 98.71%.*

***Keywords****: News Articles, Entity Extraction, Bidirectional LSTM Named Entity Recognition, Conditional Random Fields, Natural Language Processing, K-Fold Cross Validation.*

## 1. PENDAHULUAN

Informasi politik semakin penting bagi kehidupan masyarakat di era komputer dan internet yang terus berkembang. Berita politik, terutama artikel berita politik di Indonesia, berfungsi sebagai dasar bagi masyarakat untuk memahami isu-isu politik yang sedang berlangsung[1]. Namun, di tengah arus informasi yang begitu besar untuk mencari dan menganalisis data tentang entitas organisasi politik secara manual menjadi sulit dan memakan waktu[2]. Solusinya, diperlukan prosedur otomatis yang efisien untuk mengekstraksi entitas organisasi dari artikel berita politik di Indonesia.

Dalam domain artikel berita politik, identifikasi entitas melibatkan pengenalan dan pemahaman terhadap berbagai entitas penting seperti nama orang atau pelaku politik, nama partai politik, serta organisasi-organisasi yang memiliki keterkaitan dalam dunia politik. Identifikasi entitas ini bukan hanya mencakup organisasi-organisasi politik yang terlibat, tetapi juga mencakup entitas-entitas yang berkaitan dalam ranah politik[3]. Tujuannya adalah untuk membuat calon pembaca mudah menemukan dan mengakses informasi tertentu melalui internet. Misalnya, mengidentifikasi partai politik atau organisasi politik tertentu juga dapat membantu pembaca memahami latar belakang, platform, dan pandangan politik yang dilaksanakan oleh organisasi atau partai politik tersebut.

*Named Entity Recognition* (NER) merupakan komponen krusial dalam pemrosesan bahasa alami (NLP) yang bertujuan mengidentifikasi entitas seperti nama orang, organisasi, waktu, dan lokasi dalam teks. Fungsi NER tidak hanya terbatas pada pengenalan entitas, melainkan juga berperan dalam tugas NLP lainnya seperti ekstraksi data dan pembuatan kueri otomatis. Dengan kemampuannya memahami dan mengekstraksi entitas dalam teks, NER menjadi pondasi penting dalam pengembangan aplikasi NLP yang lebih canggih[4].

*Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short-Term Memory* (BiLSTM) merupakan metode tradisional yang bisa digunakan dalam mengekstraksi entitas. Namun metode tersebut memiliki permsalahan yang untuk mengekstraksi entitas pada berita, memiliki biaya komputasi yang tinggi dan kesulitan dalam interpretasi model dalam mengekstraksi entitas merupakan tantangan utama dalam menggunakan metode LSTM dan BiLSTM[7].

Berdasarkan penelitian sebelumnya tentang *News text named entity Recognition based on BI-LSTM-CRF model*, hasil penelitian tersebut menghasilkan akurasi sebesar 94.53% dengan menggunakan dataset bahasa mandarin tentang berita diskriminasi. Kombinasi *BI-LSTM* dengan *CRF* berperan penting dalam meningkatkan akurasi, karena *CRF* mampu mempertimbangkan hubungan antar label secara lebih efektif, sehingga metode ini sangat baik dan akurat dalam mengekstraksi entitas-entitas penting pada berita[10].

Melihat kondisi tersebut dalam penelitian ini, bermaksud untuk mengimplementasikan metode *Bidirectional LSTM* dan *CRF* (LSTM-CRF) pada artikel berita politik di Indonesia dengan bertujuan untuk meningkatkan efisiensi dan akurasi dalam ekstraksi entitas organisasi partai politik, organisasi terkait, dan untuk memfasilitasi pemahaman yang lebih mendalam terhadap dinamika politik yang sedang berlangsung di Indonesia.

## 2. format umum

### 2.1 *Layout*

Badan dari artikel harus tersusun dalam **satu kolom**. Dokumen ini dipersiapkan dalam format yang harus digunakan oleh setiap penulis dalam artikelnya. Untuk menjaga mutu penampilan jurnal, setiap artikel yang dikirim harus sesuai dengan spesifikasi berikut ini:

1. Ukuran kertas A4 (21 cm x 29,7 cm)
2. Pada halaman ke 1, *margin* atas (jarak antara ujung atas kertas terhadap bagian atas dari baris pertama dari judul) sebesar 4,0 cm. Sedangkan pada halaman ke 2 dan seterusnya, *margin* atas sebesar 3,0 cm.
3. *Margin* kiri, bawah dan kanan untuk seluruh halaman artikel adalah sebesar 2,5 cm.
4. Jarak *Header* dari tepi kertas sebesar 1,5 cm, sedangkan untuk *Footer* 2 cm.
5. Jumlah halaman untuk setiap artikel adalah ANTARA 10-15 HALAMAN
6. File petunjuk penulisan ini telah menggunakan aturan dan format penulisan baku yang disarankan, sehingga dapat langsung dijadikan *template*.

### 2.2 Ketentuan untuk huruf dan paragraf

Font yang digunakan adalah Tahoma untuk semua *style*. Ukuran huruf yang harus digunakan dalam artikel adalah 18 point *bold* (untuk judul), 12 point (untuk nama penulis, afiliasi, dan alamat email), 11 point *italic* (untuk abstrak dan kata kunci), 11 point (untuk badan tulisan dan judul bagianmaupun subbagian), serta 10 point (untuk *headers* dan *footers*). Rumus-rumus matematika harus berupa *Equation* berukuran 11 point (menu tab *Insert Equation*), dengan ukuran *subscripts* and *superscripts* yang sedikit lebih kecil (10 point), seperti yang ditunjukan pada Persamaan (1) **dengan diawali huruf kapital**.

(1)

Judul ditulis di tengah (*centered*) bagian atas pada halaman pertama. Nama penulis (tanpa gelar) diletakkan di bawah judul, dilanjutkan dengan afiliasi dan alamat email penulis pertama, semua ditempatkan di tengah (*centered*).

Judul Abstrak harus ditulis di tengah (*centered*) setelah identitas penulis, dengan isi abstrak berukuran 11 point miring (*italic*), serta ditulis dalam dua bahasa yaitu Bahasa Indonesia dan Bahasa Inggris. Abstrak ditulis rata kiri kanan dengan margin kiri dan kanan sebesar 3,5 cm (menjorok1 cm dari margin halaman). Abstrak berbahasa Inggris ditulis setelah abstrak berbahasa Indonesia. Jumlah kata maksimum dalam setiap abstrak adalah maksimum 150 kata. Kata kunci dicantumkan setelah abstrak, berjumlah antara 5 (lima) sampai 6 (enam) buah kata kunci yang ditulis dengan ukuran huruf 11 point miring (*italic*), untuk menunjukkan subyek permasalahan artikel anda, sekaligus untuk keperluan pengindeksan.

Perlu diperhatikan bahwa tata cara penulisan paragraf yang diberlakukan adalah cara lurus, sehingga awal paragraf tidak diletakkan menjorok ke dalam. Beri jarak 1 spasi (12 point) antar paragraf. Perhatikan juga ketentuan penulisan paragraf yang baik, antara lain jumlah kalimat dalam setiap paragraf, adanya kalimat utama, satu paragraf mengandung hanya satu gagasan utama, dan ketentuan baku lainnya.

### 2.3 Penulisan Judul

**Judul Artikel:** Judul artikelharus ditulis dengan huruf kapital pada setiap awal kata, kecuali untuk kata sambung. Judul yang lebih dari dua baris disusun membentuk piramida terbalik. Pada halaman pertama dari petunjuk penulisan ini terdapat contoh penulisan yang dikehendaki. Jumlah kata pada judul yang disarankan dalam penulisan artikel adalah maksimal 12 suku kata.

**Judul Bagian:** Judul bagianharus ditulis seluruhnya dengan huruf kapital dalam jenis *bold* tanpa garis bawah, dan diletakkan di tengah (*centered*), dan diberi nomor dengan angka.

**Judul Subbagian:** Judul subbagianharus ditulis dalam jenis *bold*, *lower case* dengan huruf kapital di awal kata, dan diletakkan tanpa *indent* (tidak menjorok). Subbagian diberi nomor yang diawali oleh nomor bagian.

**Judul Sub-subbagian:** Sub-subbagianditulis dalam jenis *bold*, dengan diberi nomor berurut yang diawali oleh nomor subbagian. Judul Sub-subbagian diletakkan tanpa *indent* (tidak menjorok). Meskipun Jurnal ini mengatur format Judul Sub-subbagian, sedapat mungkin sub-subbagian ini dihindari penggunaannya. Tidak direkomendasikan adanya *heading* yang lebih rendah daripada Judul Sub-subbagian.

**Judul Gambar/Tabel:** Judul gambar atau tabel ditulis dengan font ukuran 10 point, *bold*, *lower case* dengan huruf kapital di awal kata, dan semuanya diletakkan di tengah. Gambar diberi nomor secara berurut, demikian juga dengan Tabel. Judul gambar diletakkan **di bawah** gambar, sedangkan judul tabel diletakkan **di** **atas** tabel. Judul gambar atau tabel yang lebih dari satu baris dituliskan seperti piramida terbalik.

### 2.4 Penulisan *Header* dan *Footer*

Format untuk semua *header* dan *footer* dalam template ini dapat langsung digunakan. Khusus untuk *Header* halaman pertama dan *Footer* semua halaman.

**Header halaman pertama:** *Header* pada halaman pertama terdiri dari judul jurnal, DOI Jurnal, dan nomor Jurnal yang akan dimutakhirkan oleh Redaksi. Header tersebut ditulis dengan huruf Tahoma berukuran 10 point.

**Header pada nomor halaman genap**: Header pada nomor halaman genap terdiri dari nama-nama belakang penulis artikel, dalam huruf Tahoma berukuran 10 point ditulis di tengah (*centered*). Jika artikel ditulis oleh lebih dari sama dengan 3 orang, maka hanya tuliskan nama belakang penulis pertama saja, diimbuhi dengan kata *dkk.* sesudahnya.

**Header pada nomor halaman ganjil**: Header pada nomor halaman ganjil terdiri dari judul artikel dalam *lower case* dengan huruf kapital di awal kata, huruf Tahoma berukuran 10 point, yang ditulis di tengah (*centered*).

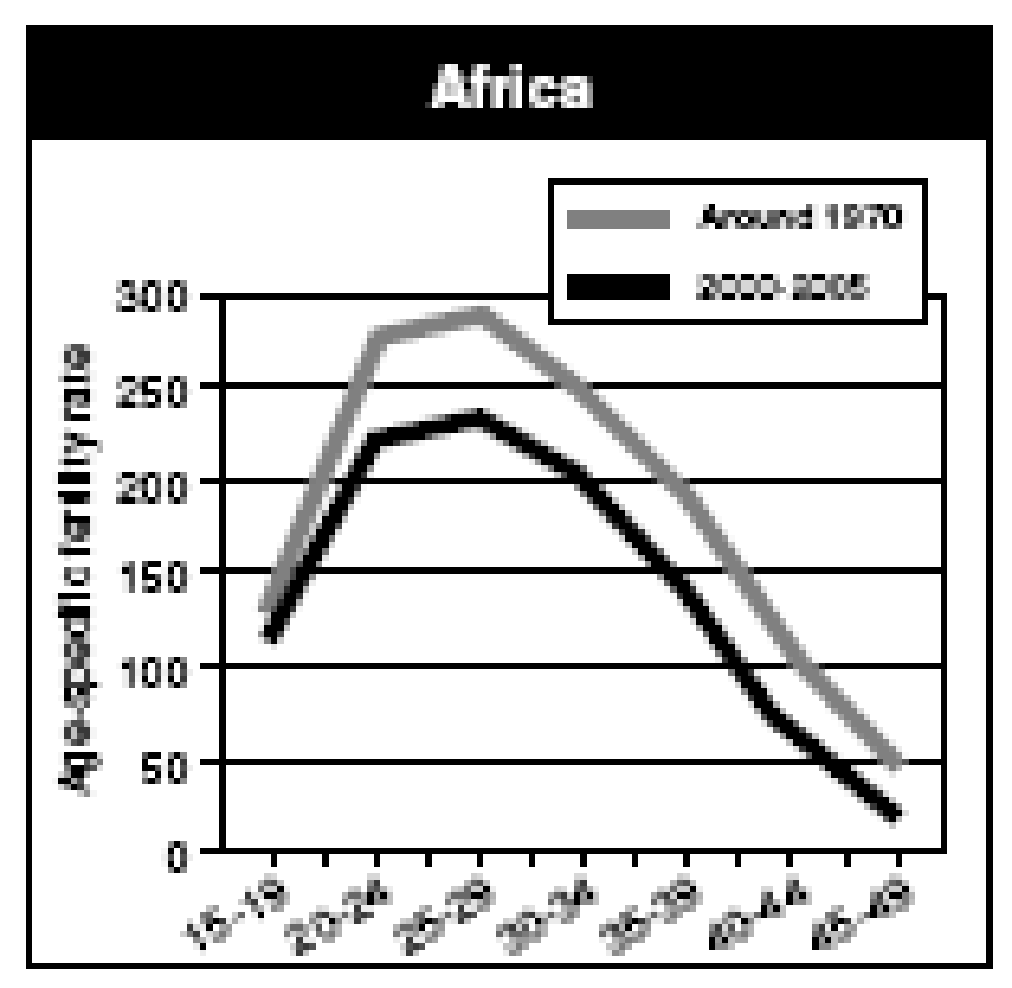
**Footer**: Seluruh halaman dalam artikel mempunyai bentuk *footer* yang sama. *Footer* terdiri dari tulisan “[Nama Jurnal]” diikuti dengan garis penghubung (*dash*) dan nomor halaman yang dimulai dengan angka 1 pada halaman pertama dalam huruf Tahoma berukuran 10 point. Nomor halaman akan dimutakhirkan oleh Redaksi pada saat dimuat.

### 2.5 Penulisan sumber dan Daftar Rujukan

Kehati-hatian dalam penulisan sumber dan Daftar Rujukan merupakan satu keharusan agar penulis dapat terhindar dari plagiarisme. Untuk itu, penulis dianjurkan untuk mengikuti secara ketat ketentuan penulisan sumber dan Daftar Rujukan dalam panduan ini. Penulis artikel bertanggungjawab sepenuhnya atas penulisan rujukan, sumber rujukan, dan Daftar Rujukan. Semua sumber yang dicantum dalam Daftar Rujukan harus dirujuk dalam badan tulisan, dan hanya pustaka yang dirujuk di dalam tulisan yang dicantumkan dalam Daftar Rujukan (*References*, bukan *Bibliography*). Sumber dituliskan dengan mengikuti tatacara (*style*) yang dikeluarkan oleh **APA**, yaitu dengan mencantumkan nama belakang penulis sumber yang dirujuk, diikuti dengan angka tahun. Daftar Rujukan juga ditulis dengan mengikuti tata cara urutan abjad dimulai dari **huruf A sampai dengan Z**. Contoh penulisan Daftar Rujukan dapat dilihat pada bagian terakhir petunjuk ini. Penulisan daftar rujukan dan cara mensitasi **diwajibkan** menggunakan *software reference* seperti mendeley, endnote, word reference, dan lainnya.

### 2.6 Gambar

Hanya gambar yang memiliki relevansi secara langsung dengan paparan yang boleh dicantumkan dalam artikel. Gambar diletakkan di tengah, di tempat yang paling relevan dengan kalimat yang merujuknya dalam artikel. Setiap gambar (foto, grafik, dan diagram) dalam artikel harus dilengkapi dengan keterangan/judul gambar dan nomor gambar berurutan, ditulis di bawah gambar pada posisi tengah dengan font ukuran 10 pt *bold*. Misalnya: “Gambar 1. Trend ASFR di Afrika”. Gambar harus relevan secara langsung dengan artikel, dan selalu dirujuk dalam artikel (disebut sebagai “Gambar 1”, **dengan diawali huruf kapital**). Penulis bertanggungjawab sepenuhnya terhadap kualitas gambar yang dicantumkan. Jurnal dapat menerima gambar berwarna, namun tidak akan melakukan koreksi apapun terhadap kualitasnya. Semua gambar sebaiknya dikompres sehingga memiliki resolusi maksimum 220 dpi, dan hapus selalu bagian yang di-*crop*. Jika gambar diambil dari sumber lain, selalu cantumkan sumber dari mana gambar tersebut diambil sebagai bagian dari judul gambar. Redaksi dapat menyesuaikan ukuran dan resolusi gambar jika dibutuhkan. Seluruh gambar tidak diperkenankan mengambil dari sumber lain, akan tetapi harus original dibuat oleh penulis.



**Gambar 1. Trend ASFR Afrika**

**2.7 Tabel**

Hanya tabel yang memiliki relevansi langsung dengan paparan yang boleh dicantumkan dalam artikel. Tabel diletakkan di tengah, di tempat yang paling relevan dengan kalimat yang merujuknya dalam artikel. Setiap tabel harus mempunyai judul dan nomor tabel berurutan, ditulis di atas setiap tabel pada posisi tengah dengan font tulisan serupa dengan gambar, seperti “Tabel 1. Perbandingan Jumlah Penduduk per Kecamatan Tahun 2010”. Tabel dirujuk dalam artikel sebagai “Tabel 1” **dengan diawali huruf kapital**. Ukuran huruf untuk isi tabel disesuaikan dengan kebutuhan, dengan memperhatikan keterbacaan. Jika sangat dibutuhkan, jenis huruf pun bisa disesuaikan dengan huruf yang lebih ramping seperti misalnya *Arial Narrow*.

**Tabel 1. Perbandingan Jumlah Penduduk Per Kecamatan Tahun 2010**

|  |  |  |  |
| --- | --- | --- | --- |
| **Kepala tabel kolom 1** | **Kepala tabel kolom 2** | **Kepala tabel kolom 3** | **Kepala tabel kolom 4** |
| Isi 11 | Isi 12 | Isi 13 | Isi 14 |
| Isi 21 | Isi 22 | Isi 23 | Isi 24 |

## 3. format isi

Paparan dalam artikel dituliskan sesuai dengan kaidah penulisan artikel ilmiah yang baik. Pada dasarnya, artikel terdiri dari bagian:

1. Pendahuluan
2. Metode
3. Hasil dan Pembahasan
4. Kesimpulan
5. Daftar Rujukan

Standar penulisan badan tulisan juga merujuk kepada penulisan artikel ilmiah yang baik. Sedapat mungkin poin pemikiran penulis dituangkan dalam bentuk paragraf, dan bukan dengan penulisan enumerasi menggunakan nomor. Penggunaan *bullet* sama sekali tidak dianjurkan. Jika tulisan dengan bullet membentuk kalimat lengkap, maka tuliskan saja sebagai kalimat dalam paragraf. Jika hanya berupa frasa, maka tuliskan sebagai bagian dari sebuah kalimat yang lengkap. Jika sangat dibutuhkan, beri nomor urut dalam tanda kurung untuk menandai, dan dipisahkan dengan tanda titik koma.

## 4. KESIMPULAN

Penulisan isi kesimpulan menggunakan huruf dan gaya paragraf yang sama dengan bagian lainnya. Untuk menghindari kesalahan penulisan artikel, disarankan untuk langsung menggunakan dokumen ini sebagai format (*template*) dengan menghapus isi petunjuk penulisan ini dan menyimpan (save as) sesuai dengan nama file yang diminta.

## UCAPAN TERIMA KASIH

Bagian ini berisi ucapan terima kasih kepada suatu instansi jika penelitian ini didanai atau mendapat dukungan oleh instansi tersebut disertai dengan nomor kontrak penelitian, atau jika ada pihak yang secara signifikan membantu langsung penelitian atau penulisan artikel ini. Jika pihak tersebut sudah tercantum sebagai penulis, maka tidak perlu disebut lagi dalam Ucapan Terima Kasih ini.

## DAFTAR rujukan

**Rujukan Buku:**

Bohmer, M. (2012). *Beginning Android ADK with Arduino.* Newyork: Apress.

Meier, R. (2012). *Professional AndroidTM 4 AQApplication Development.* Indianapolis: John Wiley & Sons, Inc.

**Rujukan Jurnal:**

Atzori, L., & Andreas. (2012). Performance Analysis of Fractal Modulation Transmission over Fast Fading Wireless Channels. *IEEE Transactions on Broadcasting, 48*(2), 103 - 110.

Darlis, A. R., Lidyawati, L., & Nataliana, D. (2016). Implementasi Visible LIght Communication (VLC) pada Sistem Komunikasi. *Elkomika, 1*(1), 13 - 25.

**Rujukan Prosiding:**

Zeng, G., & Qiu, Z. (2008). Audio Watermarking in DCT. *International COnference on Signal Processing*, (pp. 2193 - 2196).

**Rujukan Sumber *Online* :**

Macleod, D. (2010, June 25). *Post-Modernism and Urban Planning*. Retrieved from www3.sympatico.ca.

Catatan :

1. Daftar Rujukan **diwajibkan** minimal 15 yang tersebar dalam **10 tahun terakhir**.
2. Seluruh daftar rujukan diatas harus terujuk di dalam artikel dengan contoh

…..…Pada tahun 2012, penelitian yang dilakukan oleh Meier **(Meier** **,2012)** dan timnya, mencoba untuk mengirimkan data dengan kecepatan tinggi……

1. Penulisan daftar rujukan diwajibkan menggunakan aplikasi penulisan referensi, seperti halnya mendeley, endnote, word references, atau lainnya.