Review Named Entity Recognition Dengan Menggunakan *Machine Learning*

Dwi Swasono Rachmad

Fakultas Teknik, Jurusan Teknik Informatika, Universitas Bhayangkara Jakarta Raya
Bekasi, Jawa Barat
dwi.swasono@dsn.ubharajaya.ac.id

Abstrak

NER atau *Named Entity Recognition* yang sering dikenal sebagai salah satu komponen utama dari sistem pertanyaan jawaban. NER memiliki cara tradisional yang selanjutkan dikembangkan sebagai salah satu komponen untuk mendapatkan informasi dengan mengekstraksi kata dan terdapat teknik yang dapat difokuskan pada tahap terakhir. Pada artikel ini dapat diketahui dengan melakukan beberapa pendekatan telah digunakan oleh beberapa peneliti dalam meneliti fungsi NER sebagai ekstraksi informasi kata. *Name Entity Recognition* atau NER pada berbagai penerapan yang telah dilakukan penelitiannya. NER memiliki fungsi sebagai ekstrasi dari kata yang dapat memberikan informasi terkait kalimat atau kata-kata. Berdasarkan pada penelitian dapat diketahui terdapat beberapa masalah pada sistem penjawab pertanyaan yang masih merupakan bidang yang menarik untuk dilakukan pada bahasa Indonesia, Bahasa India khususnya Telugu, bahasa Arab, dan NER pada kelas nama, lokasi, organisasi, dan lainnya menghasilkan hasil yang baik dan akurasi tinggi. Namun NER yang tidak dilakukan pada kelas lokasi seperti tanggal, waktu, dan tempat serta tidak menggunakan data yang besar untuk ekstrasi dalam NER. Dalam hal ini, NER akan dimanfaatkan untuk machine learning yang lebih baik untuk mengenal berbagai kata atau elemen eksraksi dari suatu kata.

Kata kunci: *ANN, CFR, Machine Learning, NER, Supervised Learning*

Abstract

NER or Named Entity Recognition is often known as one of the main components of the question answering system. NER has a traditional method which is further developed as a component to get information by extracting words and there are techniques that can be focused on the final stage. In this article, it can be seen that several approaches have been used by several researchers in examining the function of NER as the extraction of word information. Name Entity Recognition or NER in various applications that have been carried out by the research. NER has a function as an extraction of words that can provide information related to a sentence or words. Based on the research, it can be seen that there are some problems in the question answering system which is still an interesting field to do in Indonesian, Indian Language especially Telugu, Arabic, and NER in class names, locations, organizations, and others to produce good results and accuracy. high. However, NERs are not performed on location classes such as date, time, and place and do not use large data for extraction in NER. In this case, NER will be utilized for machine learning which is better for recognizing various words or elements of contraction of a word.

Keywords: *ANN, CFR, Machine Learning, NER, Supervised Learning*

1. PENDAHULUAN

Dalam membangun sebuah sistem menjawab dari pertanyaan, dapat diketahui bahwa *Named Entity Recognition* atau (NER) telah banyak digunakan di berbagai bahasa negara. NER dapat

DOI 10.34128/jsi.v6i1.204 *Received: February 28th, 2020*
Accepted: April 28th, 2020

dikenal juga sebagai komponen utama dari suatu sistem bertanya jawab. NER juga dikenal sebagai komponen inti dari sistem penjawab pertanyaan. NER yang berbentuk tradisional sudah dilakukan pengembangan dan dijadikan sebagai salah satu komponen yang berfungsi sebagai sistem yang akan mengekstraksi informasi, dan teknik ini akan difokuskan kepada penggunaannya.

Dokumen adalah suatu media yang dapat memiliki informasi yang berarti, dokumen tersebut dapat berupa gambar dan teks. Dokumen teks adalah dokumen yang berisikan kumpulan dari karakter-karakter yang mejadi suatu kalimat. Jika pada dokumen bentuk teks biasanya teks tersebut memiliki informasi yang sangat penting, diantaraya berupa orang, nama, organisasi dan nama tempat. Cara untuk mendapatkan informasi dalam dokumen bentuk teks masih harus dilakukan secara konvensional, yaitu dengan cara membaca terlebih dahulu untuk semua dokumen tersebut, lalu dilanjtukan dengan menentukan kata atau kalimat yang mengandung karakteristik atau unik sesuai dengan yang ditentukan dan juga dengan melakukan hal tersebut, pastinya akan memakan waktu yang lama dalam melakukan penetuannya untuk mendapatkan informasi dalam dokumen teks tersebut. Oleh sebab itu, perlu dibuatkannya NER atau *Named Entity Recognition* yang memiliki fungsi sebagai sesuatu yang dilakukan untuk mendapatkan informasi dari suatu dokumen yang bersifat penting, yang biasanya meliputi orang, nama, nama tempat dan organisasi. Oleh karena itu, dengan diibuatkannya NER atau *Named Entity Recognition* yang bertujuan untuk mendapatkan informasi penting yang cepat dan akurat dalam melakukan proses pencairan yang mudah, efektif dan efisien.

Dalam hal ini dapat disimpulkan bahwa NER dalam penerapannya melakukan pendekatan *machine learning* yang pengakuannya secara detail dalam NER. Pendekatan tersebut telah dilakukan dengan teknis survey, dalam hal ini sangat sulit untuk dinyatakan sebagai model mana yang paling cocok untuk NER, karena pada masing-masing model tersebut memiliki kelebihan dan kekurangan pada hal-hal tertentu. Pada penelitian ini telah dapat dibuktikan bahwa pemilihan fitur dapat memodelkan peran yang sangat penting dalam kinerja *machine learning*.

Pada artikel ini dengan melakukan review terhadap 5 (lima) penelitian yang telah terpublikasi. Penelitian tersebut yang barkaitan dengan NER atau *Named Entity Recognition* dengan melakukan berbagai pendekatan *machine learning* diantaranya ialah BLSTM, CNNs, *supervised learning, semi unsupervised learning, Neural Network,* Bi-LSTM, Hibrida, dan CRF.

2. TINJAUAN PUSTAKA

2.1. NER (*Named Entity Recognition*)

NER adalah *Named Entity Recognition* yang berawal dari *named entities* atau NE salah satu kata benda yang khususnya memiliki jenis individu tertentu diantaranya orang, nama , tempat, dan organisasi. *Named Entity Recognition* saat ini sudah sering digunakan khususnya pada bidang NLP atau *Natural Language Processing*. NER awal mulanya hanya digunakan pada MUC atau *the Six Message Understanding Conference*. Pada saat sebelumnya, MUC hanya berfokus kepada ekstraksi informasi yang diantaranya ialah untuk mendapatkan informasi dari salah satu informasi yang tidak berstruktur. Oleh karena itu, untuk orang-orang ingin sekali untuk mengenali dari suatu informasi seperti nama, tempat, tanggal dan waktu [1].

2.2. Kecerdasan Buatan

Kecerdasan buatan dapat disebut juga AI yang berarti *Artificial Intelegence*, yang artinya adalah salah satu program yang dibuat pada sistem atau mesin agar dapat bekerja untuk mengenali, membantu, memahami apa yang akan dikerjakannya, prilaku kecerdasan buatan yang ditanamkan agar dapat bekerja seperti manusia kerjakan pada umumnya. Kecerdasan buatan dibuat bertujuan untuk memudahkan, meringankan, mempercepat segala aktifitas yang akan dikerjakan secara efektif dan efisien [2].

p-ISSN: 2460-173X
e-ISSN: 2598-5841

2.3. Machine Learning

Machine Learning ialah salah satu bagian dari kecerdasan kecerdasan buatan. Pada *machine learning* ini dibuat agar sistem atau mesin dibuat menjadi cerdas seperti manusia pada umumnya. Namun pada hal ini, kecerdasan dibuat dengan cara proses pembelajaran dan pelatihan terlebih dahulu, sebelum sistem tersebut melakukan pada dunia nyata. Dengan demikian, segala aktifitasnya akan mudah dikenali, dipahami, dan dikerjakan dengan baik, efektif dan efisien [2].

2.4. LSTM(*Long Short Term Memory*)

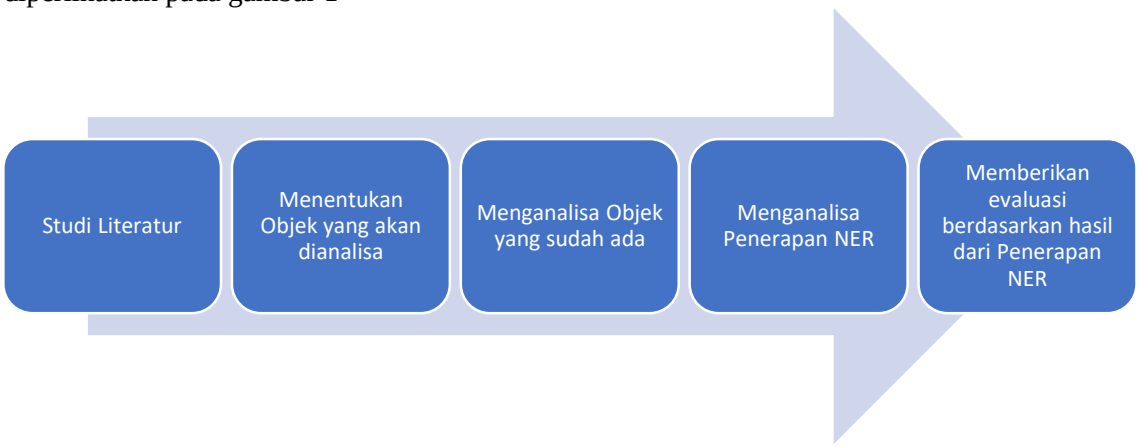
LSTM ialah bagian dari RNN yang dapat diartikan memiliki performa hasil yang baik dalam menyelesaikan berbagai tugas pada NLP. LSTM ialah merupakan bagian dari tipe RNN yang sangat popular pada saat digunakan, karena memiliki riwayat jejak rekam yang baik dalam proses prekaman long term dependencies/ atau bersifat jangka panjang. Sedangkan untuk bidirectional LSTM ialah salah satu bagian dari LSTM yang sering digunakan pada umumnya hanya saja memiliki jenis input proses yang berbeda dengan LSTM [3].

2.5. CNN (*Convolutional Neural Network*)

CNN adalah salah satu jenis dari kecerdasan buatan dalam bidang *machine learning* yang pengembangannya dari MLP(*Multilayer Perceptron*) yang disusun untuk pengolahan data yang berbentuk dua dimensi. CNN bertipe jenis *Deep Neural Network*, dikarenakan pada jaringan dalamnya memiliki tingkat dan teknik implementasinya dalam bentuk data citra. CNN membagi dalam dua metode diantaranya adalah klasifikasi *feedforward* dan *backpropagation* [4].

3. METODE PENELITIAN

Penelitian ini menitikberatkan pada NER dengan menggunakan berbagai algoritma yang berkaitan dengan penerapan NER di lingkungan sekitar. Metode yang digunakan pada penelitian ini diperlihatkan pada gambar 1



Gambar 1. Metode Penelitian NER

Pada metode penelitian ini dapat diartikan sebagai berikut:

1. Studi Literatur
   Dengan melakukan pengumpulan data dari beberapa penelitian tentang NER dan yang telah dilakukan sebelumnya tentang NER
2. Menentukan objek yang akan dianalisa
   Dengan menetapkan NER di dalam bidang tata bahasa di lingkungan pekerjaan dan sehari-hari
3. Menganalisa objek yang sudah ada
   Dalam hal ini menggunakan dan menguji penerapan NER yang sudah di implementasikan dan sudah dikategorikan lengkap sesuai dengan penerapannya

4. Menganalisa Penerapan NER
   Dalam hal ini menganalisa dari hasil yang sudah dilakukan dan di implementasikan untuk memberikan hasil yang terbaik sesuai dengan penerapannya
5. Evaluasi hasil NER
   Dalam hal ini, mengevaluasi dari berbagai penerapan NER di berbagai lingkungan tata bahasa Indonesia, bidang medis, bahasa Arab, bahasa Telugu, dan bidang sosial media. Dalam hal ini penulis melakukan penelitian dengan menggunakan beberapa algoritma, diantaranya ialah dengan artikel yang telah terpublikasi pada penerapannyadiantaranya BLSTMS, CNNs, *supervised learning, semi unsupervised learning, Neural Network*, Bi-LSTM, Hibrida, dan CRF.

4. PEMBAHASAN

Pada review penelitian *Named Entity Recognition* atau NER di berbagai penerapan dan pendekatan, dapat menghasilkan.

1. Dalam penelitian ini, didapatkan bahwa *Name Entity Recognition* machine learning dengan menggunakan BLSTM-CNN / Hybrid Bidirectional LSTM dan jaringan saraf convosional, diimplementasikan kepada bahasa Indonesia, dengan membandingkan tata Bahasa Indonesia dengan mengekstraksi informasi dari artikel ke dalam empat Bahasa kelas yang berbeda-beda, diantaranya ialah orang, organisasi, lokasi dan acara. Dengan perbandingan cara eksperimen yang dilakukan pendekatan pembelajaran yang mendalam, NER sebagai salah satu sub dari penelitian Bahasa alami dalam kecerdasan buatan, NER itu sendiri berguna dalam mengekstraksi informasi dalam bentuk teks dengan melakukan indentifikasi serta mengenali entitas dari beberapa diantaranya, orang, organisasi, lokasi, dan lain-lain. NER dijadikan sebagai salah satu proses dari NLP, alasannya ialah mampu menghasilkan dan mengimplementasikan di beberapa bidang diantaranya ialah mesin penerjemah, mesin pencari, penindeks dokumen otomatis, serta sistem tanya jawab otomatis, pencarian informasi otomatis, dan lain-lain [5].
2. Dalam penelitian yang telah dilakukan dengan teknik survey, dengan membandingkan *machine learning* model yang digunakan pada NER, penerapannya pun di berbagai bidang algoritma yang ada, dengan pembelajaran yang diawasi dengan algoritma model markov, bidang acak bersyarat, bidang entropi mesin, dan pohon keputusan. pada penerapan NER dibidang kesehatan, biomedik, dan media sosial. Dengan melakukan perbandingan antara empat algoritma dengan NER di tiga penerapannya. Maka dihasilkan sulit untuk dinyatakan model yang ideal, dikarenakan masing-masing model memiliki kekurangan dan kelebihan, bahkan dapat menimbulkan kerugian dari aturan-aturan pendekatan tersebut yang bersifat tidak portable [6].
3. Dalam penelitian NER dengan *machine learning* yang penerapannya pada bahasa Arab dengan berbagai metode diantaranya jaringan saraf tiruan, back-propagation net dengan nilai token atau corpus sebanyak 150, dapat menghasilkan nilai 92% dibandingkan dengan nilai dari pohon keputusan yang bernilai 87%. Namun dalam hal ini penelitian tersebut melanjukan dari NERA, yang sebelumnya memiliki nilai belum maksimal dalam mengekstraksi NER pada bahasa Arab dan juga baru membandingkan antara pohon keputusan dengan jaringan saraf tiruan saja, belum membandingkan dengan berbagai metode lainnya untuk mengetahui nilai terbaik dalam NER [7].
4. Dalam penelitian NER dengan *machine learning* yang penerapannya pada catatan klinis, yang terfokus kepada obat-obatan dan narkoba. Penelitian ini hanya berfokus kepada kata yang komprehensif representasi atau inputan dari jaringan saraf. Representasi tersebut hanya kepada kata-kata yang memiliki fungsi untuk meningkatkan kinerja dasar dari LSTM. Dengan hasil yang didapatkan dalam proses input jaringan saraf dengan nilai baik untuk sebatas pelatihan, namun agar NER lebih baik, maka penerapannya harus digabungkan dengan RNN untuk memberikan nilai bobot yang lebih penting dan juga tidak sebagai kinerja teratas sebagai acuan mendeteksi entitas medis dari catatan klinis [8].

5. Dalam penelitian NER dengan *machine learning* yang penerapannya pada bahasa Telugu dengan pendekatan pendekatan berbasis aturan, *machine learning*, hibrida, dan Suffix, bahasa ini berasal dari India yang dimiliki keluarga dari Dravida dengan nilai linguistik yang beragam, dalam hal ini CRF sebagai algoritma dalam bahasa Telugu, yang dapat menghasilkan nilai baik dengan pendekatan berbasis aturan *pada machine learning*, dan bidang acak bersyarat dengan kelas nama orang, nama lokasi, dan nama organisasi. Namun agar dapat maksimal dan terbaik yang bertujuan untuk memberikan hasil akurasi yang tinggi dapat menggunakan corpus yang lebih banyak dan kelas yang banyak dengan pendekatan hybrid [9].

5. KESIMPULAN

Dari berbagai penelitian yang telah dilakukan, dapat disimpulkan penerapan NER diantaranya:

1. Dapat dihasilkan bahwa NER dapat digunakan diberbagai tujuan untuk mengetahui berbagai bentuk bahasa yang akan di ekstraksi menjadi suatu informasi yang berguna untuk mewakili suatu kata yang mudah dikenali oleh manusia.
2. Untuk saat ini penggunaan NER banyak diapliaksikan untuk nama, organisasi, nama tempat, lokasi, dan lainnya khususnya dalam bahasa Indonesia, India, Arab, Telugu, dalam bidang medis, dan bahasa sosial media.

Berbagai hal yang telah simpulkan, namun dalam hal ini terdapat saran dalam NER diantaranya ialah

1. Perlu ditambahkannya corpus dalam bahasa Indonesia, dan penggunaan algoritma BLSTMs-CNNs agar dapat memiliki perbandingan hasil yang terbaik.
2. NER dalam bidang kesehatan hanya membahas biomedis yang diantaranya gen, protein, DNA dan RNA serta basis yang tidak portable yang mengakibatkan perlu pengembangan yang praktis agar dapat digunakan dengan mudah, efektif dan efisien untuk mengetahui NER dalam bidang biomedis.
3. NER dalam bahasa Arab hanya membahas teks Arab, tidak membahas teks arab secara media sosial yang mudah di mengerti dalam penerapan NER.
4. NER dalam bidang kesehatan khususnya belum dilakukan dengan NER representasi dari kalimat, dan perlu pengembangan lebih tinggi untuk jumla corpus dalam bidang obat-obatan.
5. NER dalam bahasa Telugu, perlu pengembangan dibidang bahasa media sosial, dikarenakan untuk mengetahui akurasi dari NER dibidang media sosial.

DAFTAR PUSTAKA

[1] A. Willyawan, "Named Entity Recognition (NER) Bahasa Indonesia Menggunakan Conditional Random Field dan Pos-Tagging," Universitas Sumatera Utara, Medan, 2018.

[2] H. A. Ramadhan and D. A. Putri, "Big Data, Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia: Usulan Desain, Prinsip, dan Rekomendasi Kebijakan".*Centre for Innovation Policy and Governance (CIPG)*: Jakarta, 2018.

[3] R. Ashrovy, "Pengenalan dan Tutorial RNN menggunakan Python" 2017. [Online]. Available:https://medium.com/@ashrovy/recurrent-neural-network-part-one-822f1341fec.

[4] N. Sofia, "Convolutional Neural Network" 2018. [Online]. Available: https://medium.com/@nadhifasofia/1-convolutional-neural-network-convolutional-neural-network-merupakan-salah-satu-metode-machine-28189e17335b

[5] W. Gunawan, D. Suhartono, F. Purnomo, and A. Ongko, "Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs," in *Procedia Computer Science*, 2018, doi: 10.1016/j.procs.2018.08.193.

[6] U. J. A.Salini, "*Named Entity Recognition* Using Machine Learning Approaches," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 6, no. 11, pp. 491–501, 2017.

[7] N. F. Mohammed and N. Omar, "Arabic *Named Entity Recognition* using artificial neural network," *J. Comput. Sci.*, 2012, doi: 10.3844/jcssp.2012.1285.1293.
[8] E. Florez, F. Precioso, M. Riveill, F. Liu, A. Jagannatha, and H. Yu, "*Named Entity Recognition* using Neural Networks for Clinical Notes," in *Proceedings of Machine Learning Research*, 2018.
[9] M. H. Khanam, M. A. Khudhus, and M. S. P. Babu, "*Named Entity Recognition* using Machine learning techniques for Telugu language," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2016, doi: 10.1109/ICSESS.2016.7883220.

Biodata Penulis

Dwi Swasono Rachmad, Lahir di Bekasi, 15 Maret 1990, Meraih Gelar Sarjana Teknik (ST) dari Universitas Gunadarma pada tahun 2013, dan juga telah menyelesaikan program Magister Sistem Informasi (MMSI) pada program studi Sistem Informasi Bisnis pada tahun 2016, dan saat ini sedang melanjutkan studi Doktoral Teknologi Informasi melalui Beasiswa LPDP dari Kementerian Keuangan Republik Indonesia pada tahun 2019.