# BIDIRECTIONAL LSTM-CNNs UNTUK EKSTRAKSI ENTITY LOKASI KEBAKARAN PADA BERITA ONLINE BERBAHASA INDONESIA

(BIDIRECTIONAL LSTM-CNNs FOR ENTITY EXTRACTION OF FIRE LOCATION IN INDONESIAN ONLINE NEWS)

Studi Kasus di Provinsi DKI jakarta (Case Study in DKI Jakarta Province)

## Alif Andika Putra<sup>1</sup>, Robert Kurniawan<sup>2</sup>

Politeknik Statistika STIS<sup>1,2</sup> Jakarta E-mail: 16.8994@stis.ac.id

## **ABSTRAK**

Provinsi DKI Jakarta merupakan salah satu daerah rawan terjadi kebakaran. BPBD DKI Jakarta sebagai Lembaga penanggulangan bencana memiliki salah satu misi yaitu meningkatkan kesiagaan masyarakat kota Jakarta terhadap bencana, salah satunya bencana kebakaran. Peningkatan kesiagaan terhadap bencana kebakaran dapat dilakukan dengan penyajian informasi mengenai lokasi rawan terjadinya kebakaran. BPBD DKI Jakarta dalam hal ini dapat memanfaatkan peluang dengan adanya perkembangan teknologi informasi dan komunikasi, seperti internet sebagai sumber daya informasi. Persebaran informasi melalui internet salah satunya dimuat dalam bentuk web berita online. Informasi yang terdapat pada artikel berita online dapat dijadikan sebagai sumber informasi dalam memperoleh data. Suatu rangkaian proses diperlukan untuk dapat mengekstraksi informasi yang ada didalam artikel berita online. Pada penelitian ini, Eksraksi informasi pada artikel berita online dilakukan dengan mengklasifikasi entity ke dalam kelas-kelas tertentu menggunakan Name Entity Recognition (NER) dengan pendekatan deep learning hybrid network model Bidirectional LSTM-CNNs (BLSTM-CNNs). Penelitian ini menunjukan model NER dengan BLSTM-CNNs memiliki performa yang baik berdasarkan hasil perhitungan F1-score, presisi dan recall. Kemudian, dilakukan pemetaan berdasarkan entity lokasi yang terdapat dalam artikel berita online hasil klasifikasi menggunakan model NER dengan BLSTM-CNNs.

**Kata kunci**: kebakaran, Named Entity Recognition, Bidirectional LSTM-CNNs

#### **ABSTRACT**

DKI Jakarta Province is one of the areas prone to fire. BPBD DKI Jakarta as a Disaster Management Institution has one mission which is to increase the preparedness of the people of Jakarta for disasters, one of which is the fire disaster. Increased preparedness for fire disasters can be done by presenting information about locations prone to fire. BPBD DKI Jakarta in this case can take advantage of opportunities with the development of information and communication technology, such as the internet as an information resource. Dissemination of information through the internet one of which is published in the form of an online news web. The information contained in online news articles can be used as a source of information in obtaining data. A series of processes is needed to be able to extract information contained in online news articles. In this study, information contraction in online news articles is done by classifying entities into certain classes using Name Entity Recognition (NER) with the deep learning hybrid network approach model of Bidirectional LSTM-CNNs (BLSTM-CNNs). This study shows the NER model with BLSTM-CNNs has good performance based on the results of F1-score, precision and recall calculation. Then, mapping is done based on the location entity contained in the online news article classification results using the NER model with BLSTM-CNNs.

Keywords: fire disaster, Named Entity Recognition, Bidirectional LSTM-CNNs

#### **PENDAHULUAN**

Provinsi DKI Jakarta merupakan salah satu daerah yang rawan terjadi kebakaran. Menurut data Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta tahun 2016, tercatat 607 kejadian kebakaran yang terjadi dan terus meningkat setiap tahunnya. Pada tahun 2018 jumlah

kejadian kebakaran di provinsi DKI Jakarta meningkat hingga mencapai 692 kejadian kebakaran yang terjadi sepanjang tahun dengan total korban jiwa sebanyak 11.072 jiwa yang tersebar pada masing-masing kota administrasi (BPBD DKI Jakarta, 2019).

BPBD DKI Jakarta merupakan salah satu Lembaga penanggulangan bencana yang melakukan tugas pokok dalam penyusunan prosedur penanganan, dan penyelenggaraan penanganan bencana di provinsi DKI Jakarta. BPBD DKI Jakarta memiliki salah satu misi, yaitu meningkatkan kesiagaan masyarakat kota Jakarta terhadap bencana, salah satunya bencana kebakaran (BPBD DKI Jakarta, 2019). Akan tetapi, belum tersedianya informasi mengenai data kebakaran pada website resmi BPBD DKI Jakarta menyebabkan masyarakat sulit untuk mengakses dan memperoleh informasi yang aktual dan cepat mengenai kebakaran.

Kebakaran adalah suatu bahaya yang mengancam keselamatan jiwa manusia atau harta benda jika terdapat nyala api yang tidak terkendali (Rahmad, dkk, 2016). Bencana kebakaran merupakan kejadian yang tidak dapat diperkirakan atau diprediksikan kapan terjadinya dan penyebabnya. Pencegahan yang dapat dilakukan yaitu dengan persiapan dan peringatan dini (Januandari, Rachmawati, & Sufianto, 2017). Oleh karena itu, pemetaan lokasi rawan kebakaran, penyebaran informasi penggulangan bencana kebakaran dan manajemen penanggulangan kebakaran diperlukan agar masyarakat lebih tanggap dan waspada terhadap bahaya bencana kebakaran dengan memanfaatkan adanya peluang dari perkembangan teknologi informasi dan komunikasi seperti internet sebagai sumber daya informasi.

Perkembangan teknologi informasi dan komunikasi yang pesat membuat informasi menjadi suatu hal yang sangat mudah diperoleh, salah satunya melalui internet. Internet merupakan media sumber informasi yang paling *up-to-date* dan memberikan kemudahan dalam memperoleh informasi sehingga manusia dapat mengetahui apa saja hal yang sedang terjadi (Rochmawati, 2012). Persebaran berbagai macam informasi dimuat dalam berbagai macam media internet secara luas dan banyak. Media internet yang digunakan oleh masyarakat yaitu berupa situs-situs *website*, diantaranya media sosial, situs lembaga pemerintahan, situs penelitian pendidikan, situs berita, situs perusahaan, dan situs lainnya. Salah satu situs yang menyajikan informasi yang aktual mengenai suatu kejadian dalam suatu waktu adalah situs berita *online*.

Kumpulan informasi yang terdapat dalam berita *online* adalah bagian dari data yang tersebar luas secara masif di dunia maya, salah satunya yaitu berita *online* mengenai kebakaran. Informasi mengenai suatu konten pada berita *online* kebakaran mengandung beberapa unsur dasar tentang apa (*what*), siapa (*who*), kapan (*when*), dimana (*where*), mengapa (*who*), dan bagaimana (*how*) (5W1H) (Nurdin & Maulidevi, 2017). Informasi yang terdapat pada artikel berita *online* kebakaran dapat dijadikan sebagai sumber informasi dalam memperoleh data.

Suatu rangkaian proses diperlukan untuk dapat mengekstraksi informasi yang ada di dalam artikel berita *online* kebakaran. Ekstraksi informasi pada artikel berita online dilakukan dengan mengklasifikasi *entity* pada artikel berita *online* kebakaran ke dalam kelas-kelas tertentu menggunakan Name Entity Recognition (NER) dengan pendekatan *deep learning hybrid network model Bidirectional* LSTM-CNNs (BLSTM-CNNs). Proses NER membantu *user* untuk menghasilkan *corpus* dan mengklasifikasikannya ke dalam grup seperti nama orang, organisasi, lokasi dan sebagainya (Sulistiyawan, dkk, 2016). Penggunaan BLSTM-CNNs pernah dilakukan terhadap dataset OntoNotes 5.0 yang dapat mencapai hasil terbaik pada saat diterapkan pada NER dengan sedikit tambahan *fiture engineering* dan melakukan pelatihan yang banyak sehingga mampu mempelajari hubungan kompleks dari sejumlah data yang besar (Chiu & Nichols, 2016).

Pengkajian pembangunan dan penggunaan model NER dengan metode *hybrid network models Bidirectional* LSTM-CNNs dilakukan untuk mengetahui *performance* dari model NER BLSTM-CNNs. Kemudian, juga dilakukan ekstraksi *entity* lokasi dan waktu pada berita *online* berbahasa Indonesia untuk melihat dan memetakan persebaran kejadian kebakaran provinsi DKI Jakarta. Pemetaan dilakukan berdasarkan waktu dan lokasi berupa nama kecamatan dan kota administrasi di provinsi DKI Jakarta.

#### **METODE**

## Filtering

Data artikel berita *online* yang telah diambil diseleksi secara manual. Proses *filtering* dilakukan dengan tujuan untuk mereduksi data artikel berita *online* yang tidak sesuai dengan apa yang diinginkan, seperti terdapat adanya data artikel berita *online* yang topiknya tidak membahas mengenai kebakaran, lokasi kejadian kebakaran tidak berada di daerah provinsi DKI Jakarta, terdapat berita yang membahas mengenai kebakaran hutan dan terdapat berita yang membahas kejadian kebakaran yang sama dari dua atau lebih artikel berita *online* yang berbeda.

## **Preprocessing**

Proses *preprocessing* yang dilakukan pada penelitian ini adalah sebagai berikut:

## 1. Case Folding

Case folding dilakukan dengan tujuan untuk menghilangkan karakter tanda baca dalam teks artikel berita *online* selain tanda titik ("."), koma (","), kurung ("()"), garis miring ("/"), huruf "a" sampai "z", dan karakter angka "0...9" dan menghilangkan *whitespace* atau karakter kosong pada data teks artikel berita *online* kebakaran (Mustaghfiri, dkk, 2012).

#### 2. Tokenization

*Tokenization* dilakukan untuk melakukan pemisahan teks pada data artikel berita *online* kebakaran yang telah melalui proses *case folding* menjadi potongan-potongan kata yang disebut token.

## **Pelabelan Data**

Kata-kata hasil tokenisasi kemudian dilabeli dengan menggunakan notasi BIO (*Begin, Inside* dan *Other*) sebagai skema pelabelan yang menandakan urutan yang kemudian diklasifikasikan menjadi sembilan kelas yaitu B-LOC, I- LOC, B-PER, I-PER, B-ORG, I-ORG, B-MISC, I,MISC dan O. Kemudian, data yang telah dilakukan pelabelan akan di *split* menjadi *train* data, *test* data dan *validation* data. Pada tabel 1, dapat dilihat distribusi jumlah label pada data token artikel berita *online* kebakaran yang telah dilakukan pelabelan menggunakan skema BIO berdasarkan *training*, *test* dan *validation* data.

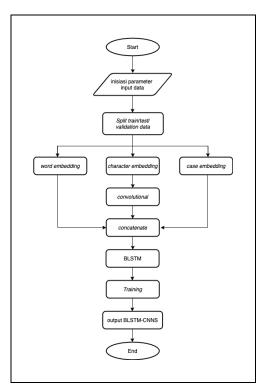
**Tabel 21.** Distribusi Jumlah Label berdasarkan skema BIO.

Training Set	Test Set	Validation Set	
819	227	182	
570	156	129	
314	81	60	
240	37	39	
681	164	180	
159	33	23	
445	115	66	
57	5	6	
9211	2357	2112	
12506	3175	2797	
	819 570 314 240 681 159 445 57	819 227 570 156 314 81 240 37 681 164 159 33 445 115 57 5 9211 2357	

## **Pemodelan NER dengan BLSTM-CNNs**

Pemodelan model NER dengan Algoritma BLSTM-CNNs dilakukan dengan menggabungkan dua algoritma *deep learning* yaitu *Convolutional Neural Network* (CNN) dan *Bidirectional* LSTM. CNN digunakan untuk mengekstrak informasi sintaksis dan token semantik dalam kalimat dan Bidirectional LSTM untuk mempelajari hubungan antar token dalam urutan (Gunawan, Suhartono

& Purnomo, 2018). Pembangunan model dilakukan berdasarkan arsitektur model yang telah diterapkan pada penelitian yang dilakukan oleh Chiu dan Nichols (2016). Langkah-langkah pemodelan NER dengan algoritma BLSTM-CNNs dapat dilihat pada gambar 1.



Gambar 38. Flowchart NER dengan BLSTM-CNNs

Pemodelan NER dengan algoritma BLSTM-CNNs dimulai dengan menginisisasi parameter input data. Kemudian, dataset yang merupakan data artikel berita *online* yang telah dilakukan pelabelan dan dibagi menjadi *train* data, *test* data dan *validation* data, untuk setiap kata pada data dilakukan proses *embedding* yang terdiri dari *word embedding*, *character embedding* dan *additional word-level features* atau Case Embedding. Kemudian, *character embedding* dengan penambahan *additional character featurers* yang telah melalui proses ekstraksi fitur karakter pada setiap kata diteruskan ke jaringan saraf *convolutional* (CNN). Selanjutnya, ketiganya di satukan (*concatenate*) dan diteruskan ke jaringan BLSTM. Terakhir, dilakukan *training* model yang kemudian model NER dapat digunakan untuk melakukan klasifikasi entitas pada artikel berita *online* kebakaran.

## Word Embedding

Word embedding yang digunakan yaitu 50 dimensional GloVe word embedding vektor dari penelitian Irfan Hanif (2018). Jika terdapat data token artikel berita online kebakaran yang tidak ditemukan didalam GloVe word embedding vektor, maka word embedding vektor baru akan diinisialisasi secara random menggunakan distribusi uniform dengan interval -0,25 sampai 0,25.

### Character Embedding

Proses *character embedding* peneliti menginisiasi vektor secara random dengan nilai yang diambil menggunakan distribusi *uniform* dengan interval -0.5 sampai 0,5. *Character embedding* yang dilakukan akan menghasilkan *output* 30 dimensional *character embedding* vektor.

#### Additional Character Features

Additional character feature digunakan untuk mengidektifikasi tipe karakter pada setiap data token artikel berita online kebakaran. Character feature yang digunakan untuk mengidetifikasi tipe karakter yaitu terdiri huruf besar, huruf kecil, angka, tanda baca dan lainnya. Pada gambar 2 memuat hasil additional character features, dimana semua karakter huruf besar, huruf kecil, angka

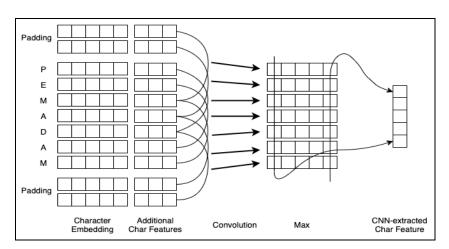
dan tanda baca ditandai dengan kode angka yang digunakan untuk mengidentifikasi tipe karakter pada setiap data token artikel berita *online* kebakaran.

	{'PADDING': 0,	's': 31,	'W': 61,		
	'UNKNOWN': 1,	't': 32,	'X': 62,		
	' ': 2,	'u': 33,	'Y': 63,		
	'0': 3,	'v': 34,	'Z': 64,		
	'1': 4,	'w': 35,	'.': 65,		
	'2': 5,	'x': 36,	',': 66,		
	'3': 6,	'y': 37,	'-': 67,		
	'4': 7,	'z': 38,	'_': 68,		
	'5': 8,	'A': 39,	'(': 69,		
	'6': 9,	'B': 40,	')': 70,		
	'7': 10,	'C': 41,	'[': 71,		
	'8': 11,	'D': 42,	']': 72,		
	'9': 12,	'E': 43,	'{': 73,		
	'a': 13,	'F': 44,	'}': 74,		
	'b': 14,	'G': 45,	'!': 75,		
	'c': 15,	'H': 46,	'?': 76,		
	'd': 16,	'I': 47,	':': 77,		
	'e': 17,	'J': 48,	';': 78,		
	'f': 18,	'K': 49,	'#': 79,		
	'g': 19,	'L': 50,	"'": 80,		
	'h': 20, 'i': 21,	'M': 51,	'"': 81,		
	'j': 22,	'N': 52,	'/': 82,		
	'k': 23,	'0': 53,	'\\': 83,		
	'l': 24,	'P': 54,	'%': 84,		
	'm': 25,	'Q': 55,	'\$': 85,		
	'n': 26,	'R': 56,	'`': 86,		
	'o': 27,	'S': 57,	'&': 87,	141 04	
	'p': 28,	'T': 58,	'=': 88,	'@': 91,	
	'q': 29,	'U': 59,	'*': 89,	'~': 93,	
	'r': 30,	'V': 60,	'+': 90,	' ': 94}	

**Gambar 2**. Additional Character Features

## Extracting Character Features dengan CNN

Penerapan penggunaan konvulsi dan *max layer* dilakukan untuk mengekstraksi vektor fitur baru dari vektor fitur per-karakter hasil proses *embedding character* dan penambahan *additional character features*. Pada model dan algoritmanya, CNN menggunakan *convolution layer* (mengurangi kompleksitas perhitungan dengan mengurangi dimensi) berdasarkan prinsip *sliding window* dan *weight sharing*, serta *pooling layer* yang digunakan untuk meringkas informasi yang didapatkan dari hasil *convolution* (Gotama, 2019). Pada gambar 3, memuat proses ektraksi fitur karakter dengan CNN. Token yang berupa kata-kata diisi dengan sejumlah karakter *padding* khusus di kedua sisi yang bersesuaian dengan ukuran kernel dari CNN yang telah ditentukan. Selanjutnya, hasil *character embedding* dengan penambahan *additional character features* pada data token artikel berita *online* kebakaran kemudian diteruskan ke CNN.



Gambar 3. Ekstraksi fitur karakter setiap token dengan Convolutional Neural Network

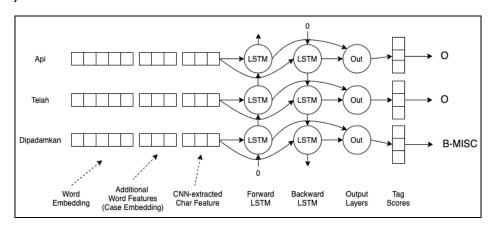
#### Additional Word-Level Features

Additional word-level feature atau Case Embedding digunakan untuk mengidentifikasi case untuk setiap kata pada data token artikel berita online kebakaran. Pengidentifikasian case dilakukan dengan menggunakan tabel pencarian sebagai case feature. Case feature memiliki

beberapa *option* yaitu *numeric* dengan kode 0, *allLower* dengan kode 1 *allUpper* dengan kode 2, *initialUpper* dengan kode 3, *other* dengan kode 4, *mainly\_numeric* dengan kode 5, *contains\_digits* dengan kode 6 dan *PADDING\_TOKEN* dengan kode 7. Kemudian peneliti menginisisasi *case embedding* vektor menggunakan matriks indentitas dengan dimensi yang bersesuaian dengan jumlah *option case*.

## Concatenate Layers

Setelah semua *layers* telah dibentuk, vektor dari setiap kata hasil pemrosesan pada setiap *layers* mulai dari *character embedding layer* dengan *additional character feature* dan CNN-extracted char feature, word embedding layers dan case embedding layers digabungkan menjadi satu. Selanjutnya, vektor tersebut diteruskan ke jaringan BLSTM dan kemudian yang terakhir ke *output layer*. Gambar 4 memuat ilustrasi *concatenate layers* yang diteruskan ke jaringan BLSTM dan *output layer*.



Gambar 4. Concatenate layers diteruskan ke jaringan BLSTM dan output layer

## **Evaluasi Model**

Pengujian model yang dilakukan untuk mengetahui *performance* model NER dengan BLSTM-CNN yang dibuat diukur dengan menggunakan perhitungan presisi, *recall* dan F1-*score* (akurasi) dengan menggunakan *test* dan *validation* data.

## **Pemetaan Entitas Lokasi Kebakaran**

Pemetaan entitas lokasi kebakaran dilakukan berdasarkan hasil dari klasifikasi data teks artikel berita *online* menggunakan model NER BLSTM-CNNs. Pemetaan dilakukan untuk melihat persebaran lokasi kejadian kebakaran yang terjadi pada periode waktu 1 Januari 2020 sampai dengan 20 April 2020.

### **HASIL DAN PEMBAHASAN**

Data artikel berita *online* kebakaran dikumpulkan dari artikel berita *online* kebakaran pada web berita Detik.com yang terjadi pada bulan Januari 2018 sampai dengan Desember 2019. Data artikel berita *online* yang didapatkan dari hasil *scraping* pada web berita Detik.com sebanyak 2377 data artikel berita *online* kebakaran. Selanjutnya, data artikel berita *online* kebakaran yang telah didapatkan dilakukan proses *filtering*. Dari proses *filtering* data, didapatkan data artikel berita *online* kebakaran sebanyak 829 data artikel berita *online* kebakaran.

## **Implementasi Model NER BLSTM-CNNs**

Performance Evaluation

Data	Precision (P)	Recall (R)	F <i>Measure</i>

Testing Data	0,878	0,816	0,846
Validation Data	0,789	0,802	0,796

Berdasarkan nilai pada tabel 2, hasil perhitungan evalusi model NER BLSTM-CNNs menggunakan *testing* data menghasilkan nilai F-*measure* sebesar 0,846, nilai *recall* sebesar 0,816 dan nilai presisi sebesar 0,878. Sedangkan, pada perhitungan evaluasi model NER BLSTM-CNNs menggunakan *validation* data menghasilkan nilai F-*measure* sebesar 0,796, nilai *recall* sebesar 0,802 dan nilai presisi sebesai 0,789.

## Hasil klasifikasi Model NER dengan BLSTM-CNNs

Setelah proses pembangunan model NER BLSTM-CNNs telah dilakukan, peneliti melakukan klasifikasi teks artikel berita *online* kebakaran menggunakan model NER BLSTM- CNNs. Klasifikasi teks berita *online* kebakaran dilakukan pada data artikel berita *online* kebakaran yang memuat kejadian kebakaran yang terjadi di provinsi DKI Jakarta dalam rentang waktu 1 Januari 2020 sampai dengan 20 April 2020. Hasil klasifikasi teks pada data artikel berita *online* kebakaran dimuat pada gambar 5.

```
('jalan', 'B-LOC'),
('mirah', 'I-LOC'),
('kencana', 'I-LOC'),
('blok', 'I-LOC'),
('c', 'I-LOC'),
(',', '0'),
[('jakarta', 'B-LOC'),
      '-', '0'),
'sebuah', '0'),
'rumah', 'B-LOC'),
'di', '0'),
'kawasan', '0'),
'kebayoran', 'B-LOC
                                                                                                                                                                                                        'lantai',
'.', '0'),
'``', '0'),
                                                                                                                                           'T-LOC').
                                                                                                                                                                                                        'iya', '0')
'benar', '0
                                                                                                                                                                                                       ['benar', '0'),
['terjadi', '0'),
('kebakaran', 'B-MISC'),
('.', '0'),
('10', '0'),
('unit', '0'),
('(', '0'),
('diforphan', '0'),
                                                                                                          'kaboyaran', 'B-LOC'),
'lama', 'I-LOC'),
                                               'B-LOC'),
      'lama', 'I-LOC'),
'jakarta', 'B-LOC'),
'selatan', 'I-LOC'),
'terbakar', 'B-MISC'),
                                                                                                      ('lama', 'I-LOC'),
(',', '0'),
('jakarta', 'B-LOC'),
('selatan', 'I-LOC'),
                               'I-LOC'),
                                                                                                      (',', '0'),
('sabtu', '0'),
('(', '0'),
('4/1/2020', '0'),
                                                                                                                                                                                                       ('dikerahkan', '0'),
                                                                                                                                                                                                     ('dikerahkan', '0'),
(')', '0'),
(')', '0'),
("'", '0'),
("kata', '0'),
('petugas', 'B-ORG'),
('bagian', 'I-ORG'),
('umum', 'I-ORG'),
('damkar', 'I-ORG'),
('damkar', 'I-ORG'),
('jakarta', 'B-LOC'),
('selatan', 'I-LOC'),
        .', '0'),
sebanyak', '0'),
    ('10', '0'),
('unit', '0'),
('mobil', '0'),
                                                                                                    ('1,','0'),
('pada','0'),
('pukul','0'),
('04.35','0'),
('wib','0'),
        'pemadaman', 'B-ORG'),
'kebakaran', 'I-ORG'),
'dikerahkan', 'O'),
                                                                                                      ('04...
('wib', 'u
'.'0'),
        'ke', '0'),
'lokasi.kebakaran', '0'),
                                                                                                      ('dilaporkan', '0'),
                                                                                                    ('dlaporkan', '0')
(',', '0'),
('rumah', '0'),
('yang', '0'),
('terbakar', '0'),
('memiliki', '0'),
                                                                                                                                                                                                    ('selatan', 'I-LOC'),
(',', '0'),
('suparman', 'B-PER'),
('saat', '0'),
('dikonfirmasi', '0'),
('.', '0')]
      'rumah', '0'),
'tersebut', '0'),
'terjadi', '0'),
  ('di', '0').
```

Gambar 5. Hasil klasifikasi teks dengan NER BLSTM-CNNs

## **Pemetaan Entitas Lokasi Kebakaran**

Generate Longitude dan Latitude data

Data entitas lokasi kebakaran hasil klasifikasi teks menggunakan model NER BLSTM-CNNs merupakan data alamat titik lokasi kebakaran yang berupa teks. Untuk dapat melakukan pemetaan data entitas lokasi kebakaran diperlukan informasi lain yang berupa koordinat *longitude* dan *latitude* data entitas lokasi kebakaran pada peta. Untuk memperoleh data koordinat *longitude* dan *latitude* data entitas lokasi kebakaran, dilakukan *genarate longitude* dan *latitude* dengan menggunakan *geocoding* pada setiap data entitas lokasi kebakaran. Pada tabel 3 ditampilkan contoh cuplikan hasil proses *geocoding*.

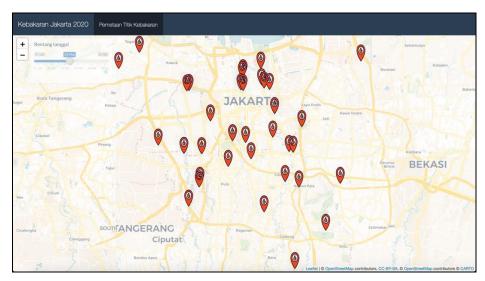
**Tabel 3.** Data hasil *geocoding* entital lokasi kebakaran hasil klasifikasi

Alamat	Tanggal	Longitude	Latitude
spbu jalan daan mogot jakarta barat jakarta	01/01/20	106,7659996	-6,158296
Jatinegara Barat kampung melayu jakarta timur	03/01/20	106,8629524	-6,2155852
kebayoran Lama Jakarta Selatan	04/01/20	106,7777818	-6,24912775

pedurenan masjid karet kuningan setiabudi jakarta selatan	10/01/20	106,827089	-6,2225158
pasar baru sawah besar jakarta pusat	11/01/20	106,8346347	-6,165709

#### Pemetaan titik lokasi kebakaran

Pemetaan titik lokasi kebakaran dilakukan dengan menggunakan *Dashboard* Rshiny dengan bahasa pemrograman R. Berikut merupakan tampilan pemetaan lokasi titik kebakaran pada *dashboard* Rshiny.



Gambar 6. Hasil klasifikasi teks dengan NER BLSTM-CNNs

## **KESIMPULAN**

Berdasarkan pemaparan hasil dan pembahasan diatas dapat disimpulkan bahwa:

- Telah berhasil dibangun model Name Entity Recognition (NER) dengan Bidirectional LSTM-CNNs. Entitas pada artikel berita online kebakaran yang telah berhasil diekstraksi yaitu lokasi, orang, organisasi dan detail kejadian yang masing-masing diklasifikasikan ke dalam kelaskelas tertentu dengan label LOC, PER, ORG dan MISC.
- Model Name Entity Recognition (NER) dengan BLSTM-CNNs yang dibangun dan digunakan untuk klasifikasi teks artikel berita online kebakaran menghasilkan performa hasil yang baik dengan hasil perhitungan F-measure/Akurasi, recall dan presisi lebih dari 75 persen, baik pada hasil evalusi menggunakan testing data maupun validation data.
- Persebaran lokasi titik kebakaran hasil klasifikasi teks artikel berita online yang memuat kejadian kebakaran yang terjadi pada rentang waktu 1 Januari 2020 sampai dengan 20 April 2020, paling banyak tersebar pada wilayah administrasi Jakarta Pusat yang terjadi pada bulan Januari hingga Februari 2020.

## **DAFTAR PUSTAKA**

Badan Penanggulangan Bencana Daerah DKI Jakarta. (2019). *Data Kebakaran*. Jakarta: BPBD DKI Jakarta. Chiu, Jason P.C & Nichols, E. (2016). *Named Entity Recognition with Bidirectional LSTM-CNNs. Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357–370.

Deng, L & Yu, Dong. (2013). *Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing*, vol. 7, nos. 3–4, pp. 197–387.

Gunawan, W, dkk. (2018). *Named-Entity Recognition for indonesian Language using Bidirctional LSTM-CNNs. Procedia Computer Science 135.* 

- Gotama J.W. (2019). *Pengenalan Konsep Pembelajaran Mesin dan Deep learning*. Department of Computer Science, Tokyo Institute of Technology.
- Januandari M.U, dkk. (2017). *Analisa Risiko Bencana Kebakaran Kawasan Segiempat Tunjungan Surabaya*. Jurnal Pengemangan Kota UNDIP, Vol 5, No 2 (149-158).
- Mustaqfiri M, dkk. (2012). *Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance*. ResearchGate.
- Nurdin, A & Maulidevi, N.U. (2018). 5W1H Information with CNN-Bidirectional LSTM. Journal of Physics: Conf. Series 978.
- Rahmad, A., Kristiawan, S.A., & Sambowo, K.A. (2016). *Pengaruh Fire Safety Management Terhadap Kehandalan Bangunan dalam Mengantisipasi bahaya Kebakaran padas Bangunan Rumah Susun di Makassar*. Jurnal Teknik Sipil 1(1).
- Rochmawati, W. (2012). *Perilaku Pemanfaatan Internet (Internet Utilization Of Behavior) (Studi Deskriptif tentang Pemanfaatan Internet Untuk Kepentingan Hiburan dan Akademik di Kalangan Anak-Anak di Kota Surabaya)*. Universitas Airlangga Repositori.
- Sihwi, S.W & Sulistiyawan, N. (2016). *Ekstraksi Entity Lokasi Tindak Kriminal pada Surat kabar Online Berbahasa Indonesia*. ReasearchGate.