

## DETEKSI KEJADIAN LALU LINTAS PADA TEKS *TWITTER* DENGAN PENDEKATAN KLASIFIKASI MULTI-LABEL BERBASIS *DEEP LEARNING*

Luthfi Atikah<sup>1</sup>, Diana Purwitasari<sup>\*2</sup>, Nanik Suciati<sup>3</sup>

<sup>1,2,3</sup>Institut Teknologi Sepuluh Nopember, Surabaya  
Email : <sup>1</sup>luthfiatikah@gmail.com, <sup>2</sup>diana@if.its.ac.id, <sup>3</sup>nanik@if.its.ac.id  
\*Penulis Korespondensi

(Naskah masuk: 27 Juni 2021 , diterima untuk diterbitkan: 31 Januari 2022 )

### Abstrak

Kemacetan merupakan salah satu kejadian yang sering terjadi di kota-kota besar. Hal ini dapat merugikan pengguna jalan, oleh karena itu perlu dilakukan pendeteksian kejadian lalu lintas. Saat ini, *twitter* digunakan sebagai sumber informasi untuk mendeteksi suatu kejadian. Namun, pengguna *twitter* cenderung membagikan beberapa informasi sekaligus, sehingga dalam satu *tweet* bisa memiliki lebih dari satu label. Pada penelitian ini dilakukan klasifikasi multi-label menggunakan 18.000 data dari akun *twitter* terverifikasi di Surabaya. Klasifikasi multi-label pada penelitian ini dilakukan untuk mengidentifikasi banyak situasi lalu lintas seperti kondisi cuaca, kecelakaan lalu lintas, kemacetan lalu lintas, lalu lintas padat, dan lalu lintas lancar. Klasifikasi dilakukan dengan menggunakan pendekatan *deep learning* (CNN dan LSTM) dan *word embedding* (*word2vec* dan *fastText*) dengan augmentasi dan non-augmentasi data. Eksperimen dilakukan dengan 3 skenario berbeda untuk melihat pengaruh data uji yang berbeda pada data latih yang sama. Selanjutnya dilakukan eksperimen untuk menguji pengaruh jumlah label terhadap klasifikasi multi-label pada data uji yang sama. Akurasi tertinggi pada non-augmentasi data adalah 0,75 dan pada augmentasi data adalah 0,95. Dari keseluruhan ujicoba akurasi tertinggi diperoleh dari kombinasi LSTM dan *fastText*.

**Kata kunci:** Klasifikasi Multi-label, *Deep learning*, Deteksi Kejadian Lalu-lintas

## *MULTI-LABEL CLASSIFICATION USING DEEP LEARNING APPROACH ON TWITTER TEXTS FOR TRAFFIC EVENT*

### Abstract

*Congestion is one of the events that often occurs in big cities. This can be detrimental to road users, therefore it is necessary to detect traffic events accurately and efficiently. Currently, Twitter is used as a source of information to detect an incident. However, twitter users tend to share several information at once, so that in one tweet can have more than one label. Therefore, multi-label classification is necessary. This study utilizes 18,000 data from verified twitter accounts in Surabaya. Multi-label classification is carried out to identify many traffic situations, such as weather conditions, events, traffic jams, heavy traffic, and smooth traffic. Classification is performed using deep learning approach (CNN and LSTM) and word embedding (word2vec and fastText) with augmented and non-augmented. Experiments are carried out with 3 different scenarios to see the effect of different uji data on the same train data. Furthermore, the experiments are conducted to examine the effect of the number of labels on the multi-label classification on the same test data. The highest accuracy on non-augmented data is 0,75 and on augmented data is 0,95. All of the highest accuracy is obtained from the combination of LSTM and fastText.*

**Keywords:** *Multi-label Classification, Deep learning, Traffic Events Detection*

## 1. PENDAHULUAN

*Twitter* merupakan salah satu sumber informasi utama terkait peristiwa yang terjadi di masyarakat.

Penelitian terbaru (Mazoyer et al., 2020) menunjukkan bahwa jurnalis *Associated Press*

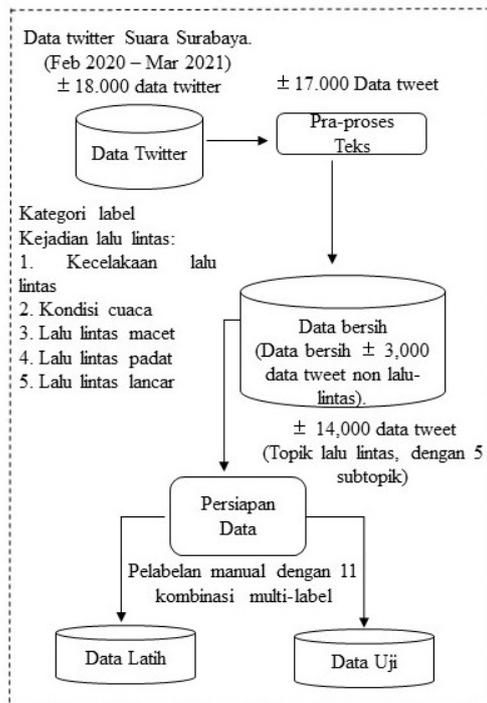
menggunakan *twitter* sebagai sumber informasi. Beberapa manfaat *twitter* adalah menjadi objek untuk membantu peneliti memahami opini publik, mengukur minat opini publik terkait topik tertentu, dan mendeteksi kejadian tak terduga secara *real-time*. Secara bersamaan, pengguna *twitter* dapat berbagi lebih dari 500 juta pesan setiap hari tentang semua topik dan subjek (Chamby-Diaz & Bazzan, 2019), misalnya terkait laporan peristiwa lalu lintas. Kejadian lalu lintas merupakan masalah yang sering terjadi di kota-kota besar karena peningkatan jumlah kendaraan, kecelakaan lalu lintas, dan kejadian tak terduga lainnya. Hal ini dapat merugikan pengguna jalan, oleh karena itu diperlukan strategi untuk mendeteksi kejadian lalu lintas secara efisien (Alomari, Mehmood & Katib, 2019).

Beberapa penelitian sebelumnya telah menggunakan data *twitter* terkait laporan peristiwa lalu lintas sebagai objek untuk mendeteksi peristiwa lalu lintas. Penelitian sebelumnya yang dilakukan oleh (Dabiri & Heaslip, 2018) memanfaatkan teks *twitter* untuk mendeteksi peristiwa lalu lintas dan insiden lalu lintas untuk membantu orang mengetahui peristiwa lalu lintas, pada penelitian tersebut hanya menggunakan satu label. Namun, teks *twitter* dapat memiliki lebih dari satu kategori karena seseorang biasanya membagikan berita sekaligus dalam satu *tweet*. Oleh karena itu diperlukan klasifikasi teks multi-label. Klasifikasi multi-label sendiri bertujuan untuk memprediksi variabel keluaran untuk setiap *instance* masukan (Yogarajan et al., 2020). Ada dua cara untuk melakukan klasifikasi multi-label, yaitu masalah transformasi dan algoritma adaptasi (Boutell et al., 2004). Diketahui bahwa metode tradisional memiliki keterbatasan untuk mengklasifikasikan data berskala besar selain klasifikasi menggunakan metode tradisional yang cenderung *overfitting* (Parwez, Abulaish & Jahiruddin, 2019). Metode *deep learning* adalah solusi untuk menangani data dalam jumlah besar. Selain itu, metode *deep learning* juga mampu menangani klasifikasi multi-label.

Metode yang digunakan dalam penelitian ini adalah *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) dengan beberapa penelitian sebelumnya yang menggunakan metode tersebut untuk mengklasifikasi teks. Penelitian sebelumnya (Parwez, Abulaish, & Jahiruddin, 2019) mengklasifikasikan teks *twitter* yang terkait dengan *tweet* medis dengan menggunakan kombinasi *word embedding* dan hasil klasifikasi CNN memberikan hasil terbaik yaitu 0,95 dibandingkan dengan klasifikasi menggunakan kombinasi lain. Selain itu keunggulan CNN dapat menyimpan informasi tentang urutan kata dan representasi vektor hasil klasifikasi lebih akurat daripada metode tradisional

(Aipe et al., 2018). Dalam penelitian lain (Wang et al., 2016) memanfaatkan metode CNN untuk mendapatkan target probabilitas dari setiap label dari hasil klasifikasi didapatkan akurasi tertinggi yaitu 0,75. Pada penelitian lain (Sari et al., 2020) dilakukan klasifikasi teks multi-label pada artikel menggunakan metode LSTM dengan kombinasi *word2vec*, hasil klasifikasi mendapatkan hasil yang baik yaitu sebesar 95,38. Keunggulan metode LSTM dibandingkan dengan metode *deep learning* lainnya adalah tidak mengabaikan informasi semantik dari urutan kata sebelumnya dan menyimpan informasi tersebut pada lapisan tersembunyi (Juanlin Hu, Xin Kang, & Shun Nishide, 2019).

Kombinasi *word embedding* dan *deep learning* diketahui memberikan hasil terbaik pada klasifikasi teks. Penelitian sebelumnya (Atefeh & Khreich, 2015) menggunakan representasi fitur *bag-of-words* (BOW) untuk mengubah *tweet* menjadi fitur vektor. Namun, BOW dalam aplikasi mengabaikan urutan kata. Selain itu, permasalahan jumlah data *tweet* yang memiliki panjang yang berbeda berpengaruh pada model. Matriks yang dihasilkan memiliki banyak nilai nol (*sparse matrix*), membuat model menjadi tidak efisien. *Word embedding* merupakan solusi untuk menutupi kekurangan BOW. Penelitian sebelumnya (Sari et al., 2020) menggunakan *word2vec* untuk klasifikasi multi-label dan mendapatkan hasil yang lebih baik daripada menggunakan BOW. Keunggulan *word2vec* adalah dapat membuat vektorisasi kata menggunakan dua teknik utama yaitu *Skip-gram* (SK) dan *Continuous Bag of Word* (CBOW) (Bdeir & Ibrahim, 2020). Dalam penelitian lain yang dilakukan (Juanlin Hu, Xin Kang, & Shun Nishide, 2019) melakukan klasifikasi multi-label dengan memanfaatkan *fastText* dalam mengklasifikasikan artikel berita. Selain itu, *fastText* merupakan perpanjangan dari model *word2vec*, yang menampilkan setiap kata sebagai sekumpulan karakter *n-gram* (Khattak et al., 2019). *FastText* merupakan metode *word embedding* bentuk *out of vocabulary*, karena *fastText* dapat memprediksi kata-kata yang tidak ada dalam kamus



Gambar 1. Persiapan Data

kosakata yang dikembangkan oleh *Facebook AI Research* dalam

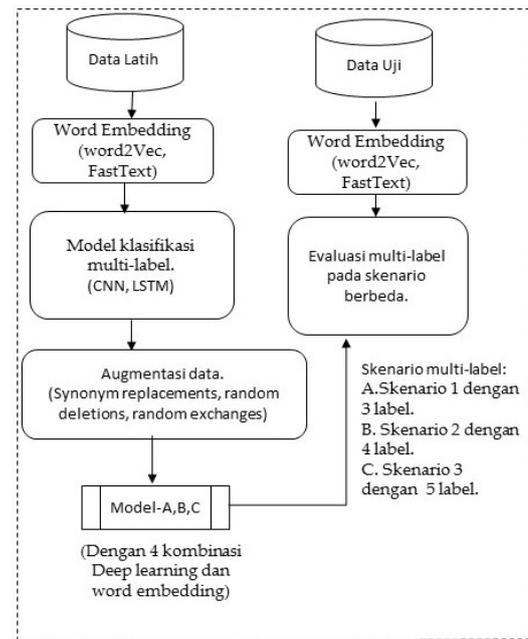
Pada penelitian ini mengusulkan identifikasi peristiwa lalu lintas dalam teks multi-label melalui objek teks *twitter* berdasarkan ekstraksi fitur teks proses dua tahap dan klasifikasi multi-label. Selain itu, untuk mengidentifikasi teks *twitter* yang mengindikasikan adanya multi-label berdasarkan kejadian lalu lintas, dilakukan manual labelling untuk mengidentifikasi data dalam penelitian ini. Tahap selanjutnya adalah klasifikasi multi-label menggunakan variasi *word embedding* dan deep learning. Klasifikasi dilakukan dengan empat pasang kombinasi antara metode penyisipan dua kata yaitu *word2vec* dan *fastText*, dengan dua metode deep learning yaitu CNN dan LSTM. Penelitian kami juga dapat menunjukkan kombinasi metode pembelajaran deep learning dan word embedding untuk proses klasifikasi teks *Twitter*.

## 2. METODE PENELITIAN

Penelitian ini akan melalui beberapa tahapan seperti terlihat pada Gambar 1 yaitu persiapan data dan Gambar 2 merupakan skenario klasifikasi multi-label dalam penelitian ini. Penjelasan rinci dari tahapan ini akan dijelaskan pada bagian berikut:

### 2.1. Persiapan Data

Data dalam penelitian ini menggunakan data *twitter* dari akun terverifikasi Suara Surabaya pada



Gambar 2. Alur Proses Pembuatan Model Klasifikasi

bulan Februari 2020 – Maret 2021. Jumlah data *tweet* yang dikumpulkan dalam penelitian ini berjumlah sekitar 18.000 data, terdiri dari data non lalu lintas dan lalu lintas. *Tweet* diambil dengan teknik *crawler* menggunakan API *twitter* dan disimpan dalam format *csv*. Data *twitter* melalui pra-pemrosesan seperti *case folding*, *tokenizing*, *stemming*, penghapusan tanda baca, dan penghapusan kata tidak dikenal. Salah satu penghapusan kata tidak dikenal dalam penelitian ini adalah penghapusan Bahasa Jawa karena banyaknya penyebutan Bahasa Jawa di akun *twitter* Suara Surabaya.

Kemudian dilakukan pelabelan secara manual menjadi lima kelas, lima kelas label mengacu pada penelitian dari (Brannon et al., 2017) dan juga berdasarkan kondisi data *twitter* akun Suara Surabaya. Berikut ini adalah contoh data *tweet* Suara Surabaya terkait kejadian lalu lintas yang sebenarnya diperoleh dalam bahasa Indonesia pada Tabel 1. Terdapat lima kelas yang terkait dengan kejadian lalu lintas, yaitu kondisi cuaca, insiden, jam lalu lintas, lalu lintas berat, dan lalu lintas ringan. Data terkait trafik sebanyak 14.000 data, selain itu 3.000 data non trafik dihapus. Kemudian 14.000 data dibagi menjadi 7.000 data latih dan 7.000 data uji untuk klasifikasi multi-label. Dalam penelitian ini, identifikasi data multi-label dilakukan secara manual, karena teks *twitter* ditulis dalam berbagai karakter, sehingga proses pelabelan memakan waktu lama. Contoh jumlah data *tweet* multi-label ditunjukkan pada Tabel 2.

Tabel 1. Definisi Contoh *Tweet* Peristiwa Lalu lintas pada Akun *Twitter* Suara Surabaya

No.	Topik*	Definisi Topik*	Kata Kunci	Teks <i>Twitter</i>
1.	Kondisi cuaca	Situasi lalu lintas yang terhenti dikarenakan banjir atau kondisi alam seperti hujan dan angin lebat.	Banjir, hujan deras, genangan air	<ul style="list-style-type: none"> <li>Salah satu jalur di Wisma Tropodo terendam banjir.</li> <li>Banjir di wilayah Baratajaya mulai surut</li> </ul>
2.	Kecelakaan lalu lintas	Situasi lalu lintas terhambat dikarenakan terjadi kecelakaan kendaraan.	Tabrakan, terguling, ban bocor	<ul style="list-style-type: none"> <li>17.44: Wisma Tropodo Waru BANJIR, dengan ketinggian satu ban sepeda motor.</li> </ul>
3.	Lalu lintas macet	Situasi lalu lintas yang ditandai dengan tersendatnya lalu lintas atau berhenti dikarenakan banyaknya jumlah kendaraan melebihi kapasitas jalan yang disediakan.	Macet, tutup jalan, mogok	<ul style="list-style-type: none"> <li>Perlintasan kereta api Dupak Rukun macet.</li> <li>Arah tropodo ke waduk asri macet.</li> </ul>
4.	Lalu lintas padat	Situasi lalu lintas yang ditandai dengan kendaraan yang bergerak antara 10-30 km/jam. pada data <i>tweet</i> keadaan dikarenakan perbaikan lalu lintas, atau kendaraan terhenti dan menutup sebagian jalan.	Padat merayap, lampu lalu-lintas, pipa bocor, merambat, ban bocor	<ul style="list-style-type: none"> <li>Ada demo di depan Grahadi. Posisi pengunjung rasa di jalur kanan. Lalu lintas sangat padat.</li> <li>Jalan kedung cowek menuju kenjeran lalu lintas padat.</li> </ul>
5.	Lalu lintas lancar	Situasi lalu lintas ditandai dengan kendaraan berjalan dengan baik.	Lancar dan lengang.	<ul style="list-style-type: none"> <li>Lalu lintas di kota Surabaya masih terpantau dengan baik.</li> <li>Gerbang Tol Porong arah Sidoarjo situasi terpantau dengan baik.</li> </ul>
6	Kecelakaan lalu lintas dan lalu lintas macet	Situasi lalu lintas terhambat dikarenakan terjadi kecelakaan kendaraan dan situasi lalu lintas yang ditandai dengan tersendatnya lalu lintas atau berhenti	Macet dan kecelakaan	<ul style="list-style-type: none"> <li>Arteri Porong arah Surabaya masih macet, imbas evakuasi truk yang terlibat kecelakaan di jembatan.</li> </ul>
7	Lalu lintas padat dan Lalu lintas lancar	Situasi lalu lintas yang ditandai berjalan lambat dikarenakan perbaikan lalu lintas, atau kendaraan terhenti dan menutup sebagian jalan dan situasi lalu lintas ditandai dengan kendaraan berjalan dengan baik	Macet dan padat	<ul style="list-style-type: none"> <li>Jalur-jalur ini Macet 1. Wiyung arah Lakarsantri Macet 2. Ketintang arah Royal Macet 3. Tol Waru arah Satelit padat 4. Kedamean arah Boboh Macet.</li> </ul>
8	Kecelakaan lalu lintas, lalu lintas macet, dan lalu lintas padat	Situasi lalu lintas terhambat dikarenakan terjadi kecelakaan kendaraan, situasi lalu lintas yang ditandai dengan tersendatnya lalu lintas atau berhenti, dan situasi lalu lintas yang ditandai berjalan lambat dikarenakan perbaikan lalu lintas	Kecelakaan, lancar, dan padat.	<ul style="list-style-type: none"> <li>16.26: 1.Update: Kendaraan yang terlibat kecelakaan di tol Sidoarjo - Porong KM 758 sudah dievakuasi. Lepas lokasi, lalu lintas lancar; 2.Mayjend Sungkono - HR Muhammad padat; <a href="http://3.HR Muhammad - Lenmarc">http://3.HR Muhammad - Lenmarc</a> padat sejak depan Apartemen Beverly. (odp-rt)</li> </ul>
9	Lalu lintas macet, lalu lintas padat, dan lalu lintas lancar.	Situasi lalu lintas yang ditandai dengan tersendatnya lalu lintas atau berhenti, dan situasi lalu lintas yang ditandai berjalan lambat dikarenakan perbaikan lalu lintas, dan situasi lalu lintas ditandai dengan kendaraan berjalan dengan baik.	Macet dan padat merayap.	<ul style="list-style-type: none"> <li>Jalur macet 1. A. Yani arah Raya Waru, imbas volume. Menjelang Layang Waru juga sama padat merayap 2. TL Lakarsantri, info sudah diteruskan ke petugas; 3. Depan Pasar Ikan Gunungsari arah Kedurus, setelah itu lancar.</li> </ul>
10	Kecelakaan lalu lintas dan Lalu lintas lancar	Situasi lalu lintas terhambat dikarenakan terjadi kecelakaan kendaraan dan situasi lalu lintas ditandai dengan kendaraan berjalan dengan baik.	Kecelakaan dan ramai lancar	<ul style="list-style-type: none"> <li>kecelakaan di depan PLN Taman. Aiptu Amin Satlantas BM Polresta Sidoarjo mengatakan, kecelakaan melibatkan mobil Kijang, menabrak tiang. Kecelakaan tunggal. Korban sudah dibawa ke RS Siti Khadijah Sepanjang. Lalu lintas ramai lancar</li> </ul>
11	Kondisi cuaca dan lalu lintas macet	Situasi lalu lintas yang terhenti dikarenakan banjir atau kondisi alam seperti hujan dan angin lebat dan situasi lalu lintas yang ditandai dengan tersendatnya lalu lintas atau berhenti	Banjir dan lalu lintas ditutup	<ul style="list-style-type: none"> <li>Jalan Porong masih ditutup dan dialihkan lewat Arteri Porong, karena Porong banjir. Iptu Cholil Kanit Turjawali Polresta Sidoarjo melaporkan, kendaraan yang ke arah Pasuruan pengalihannya di Bundaran Tanggulangin. Dan lalu lintas yang dari selatan sudah ditutup.</li> </ul>

\* Sumber Topik Identification and Categorization of Public Information in Community Base Social Media (Brannon et al., 2017)

## 2.2. Pembentukan Model Klasifikasi Multi-label

Klasifikasi dilakukan dalam lima kelas, yaitu kondisi cuaca, insiden, kemacetan lalu lintas, lalu lintas berat, dan lalu lintas ringan. Dalam eksperimen ini, kami mencoba dua skenario berbeda untuk melihat bagaimana data uji yang berbeda

memengaruhi data kereta yang sama. Dalam percobaan ini, kami mencoba tiga skenario yang berbeda untuk melihat bagaimana data uji yang berbeda mempengaruhi data latih yang sama dan juga untuk melihat pengaruh jumlah label pada klasifikasi multi-label dari data uji yang sama.

Tabel 2. Jumlah Data Multi-label pada Akun *Twitter* Sura Surabaya

Id	Kondisi Cuaca	Kecelakaan lalu lintas	Kemacetan lalu lintas	Lalu lintas padat	Lalu lintas lancar	Jumlah Tweet
1	√					255
2		√				4865
3			√			891
4				√		3516
5					√	3025
6			√		√	783
7				√	√	376
8		√	√	√		89
9			√	√	√	78
10		√			√	1269
11	√	√				317

Skenario tersebut ditunjukkan pada Tabel 4. Penelitian dilakukan dalam tiga skenario, skenario pertama dijalankan pada data latih dengan data augmentasi dan non-augmentasi menjadi tiga label. Skenario kedua adalah menjalankan data latih augmentasi dan non-augmentasi menjadi empat label. Skenario ketiga adalah menjalankan data latih augmentasi dan non-augmentasi menjadi lima label.

Dalam setiap skenario, penelitian ini mengklasifikasikan menggunakan kombinasi *word embedding* dan pembelajaran mendalam. Metode yang kami gunakan adalah kombinasi dari 2 *word embedding* dan 2 metode *deep learning*. *Word embedding* yang digunakan dalam percobaan ini adalah *word2vec* dan *fastText*. *Word2Vec* adalah pembelajaran tanpa pengawasan yang mengubah sebuah kata menjadi representasi vektor dengan mencari hubungan semantik antar kata dengan

Tabel 3. Skenario Ujicoba

Skenario	Label	Data Latih	ID Uji	Data Uji
1	3 label	Februari- Oktober	S1.1	November
			S1.2	Desember
			S1.3	Januari
			S1.4	Februari
			S1.5	Maret
2	4 label	Februari- Oktober	S2.1	November
			S2.2	Desember
			S2.3	Januari
			S2.4	Februari
			S2.5	Maret
3	5 label	Februari- Oktober	S3.1	November
			S3.2	Desember
			S3.3	Januari
			S3.4	Februari
			S3.5	Maret

menghitung jarak (Rahmawati and Khodra, 2016). Teks cepat adalah model representasi kata

yang dikembangkan dari *word2vec*. Dalam memilih algoritma *fastText* bisa menggunakan CBOW atau *skip-gram* seperti *word2vec*. Model representasi kata ini bekerja dengan memperhatikan struktur internal kata (Lim et al., 2020). Metode *deep learning* yang digunakan adalah CNN dan LSTM.

CNN merupakan metode *deep learning* yang menggunakan filter *convolution* untuk menangkap fitur tersembunyi melalui proses pembelajaran dan pembaruan berkelanjutan dengan memanfaatkan *feed forward*, dan untuk melakukan pembelajaran pembobotan dari kerugian menggunakan *backpropagation* (Parwez et al., 2019). LSTM menggunakan tiga gerbang yaitu gerbang *input*, gerbang *forgot*, dan gerbang *output* untuk mengontrol penggunaan dan pembaruan informasi sebelumnya. Kombinasi metode tersebut adalah *word2vec* + CNN, *word2vec* + LSTM, *fastText* + CNN, dan *fastText* + LSTM. Hasil dari

Tabel 4. Hasil Klasifikasi Multi-label dengan *FastText*

Skenario	ID Uji	Data Uji	CNN dengan <i>FastText</i>		LSTM dengan <i>FastText</i>	
			Non- Augmentasi	Augmentasi	Non- Augmentasi	Augmentasi
1	S1.1	November	0,61	0,71	0,65	0,75
	S1.2	Desember	0,62	0,72	0,63	0,80
	S1.3	Januari	0,60	0,70	0,77	0,71
	S1.4	Februari	0,69	0,79	0,67	0,74
	S1.5	Maret	0,69	0,79	0,68	0,76
2	S2.1	November	0,68	0,74	0,71	0,79
	S2.2	Desember	0,68	0,72	0,71	0,80
	S2.3	Januari	0,65	0,79	0,68	0,81
	S2.4	Februari	0,67	0,81	0,73	0,87
	S2.5	Maret	0,59	0,76	0,65	0,84
3	S3.1	November	0,71	0,90	0,73	0,92
	S3.2	Desember	0,71	0,83	0,70	0,88
	S3.3	Januari	0,70	0,82	0,78	0,95
	S3.4	Februari	0,73	0,89	0,79	0,90
	S3.5	Maret	0,72	0,80	0,73	0,90

metode klasifikasi untuk mendapatkan persentase dengan akurasi terbaik.

dilakukan dengan menggunakan augmentasi dan non augmentasi data data latih. Skenario terakhir adalah pengklasifikasian teks *twitter* menjadi 5 label

Table 5. Hasil Klasifikasi Multi-label dengan *Word2Vec*

Label	ID Uji	Uji Data	CNN dengan <i>Word2Vec</i>		LSTM dengan <i>Word2Vec</i>	
			Non- Augmentasi	Augmentasi	Non- Augmentasi	Augmentasi
3 label	S1.1	November	0,51	0,64	0,66	0,70
	S1.2	Desember	0,55	0,65	0,65	0,68
	S1.3	Januari	0,59	0,63	0,61	0,69
	S1.4	Februari	0,59	0,65	0,63	0,69
	S1.5	Maret	0,52	0,60	0,64	0,62
4 label	S2.1	November	0,66	0,78	0,67	0,79
	S2.2	Desember	0,62	0,77	0,70	0,78
	S2.3	Januari	0,69	0,75	0,75	0,80
	S2.4	Februari	0,61	0,77	0,67	0,72
	S2.5	Maret	0,62	0,75	0,64	0,77
5 label	S3.1	November	0,72	0,78	0,75	0,89
	S3.2	Desember	0,73	0,77	0,69	0,79
	S3.3	Januari	0,74	0,84	0,78	0,78
	S3.4	Februari	0,72	0,77	0,71	0,84
	S3.5	Maret	0,69	0,71	0,68	0,73

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil

Penelitian ini melakukan tiga skenario berbeda yang bertujuan untuk mengetahui pengaruh jumlah data latih dan label yang diberikan pada teks *twitter*. Skenario 1 mengklasifikasikan data ke dalam 3 label yang berbeda, yaitu Kondisi cuaca, kecelakaan lalu lintas, dan Kemacetan lalu lintas. Skenario 2 mengklasifikasikan data ke dalam 4 label yang berbeda, yaitu kondisi Cuaca, Kecelakaan lalu lintas, Kemacetan lalu lintas, dan Lalu Lintas Padat. Pada skenario 3 mengelompokkan data ke dalam 5 label yang berbeda, yaitu kondisi Cuaca, Insiden, Kemacetan lalu lintas, Lalu Lintas Padat, dan Lalu Lintas Lancar. Setiap skenario klasifikasi multi-label diproses menggunakan augmentasi dan non-augmentasi untuk data latih. Metode augmentasi digunakan untuk mengetahui pengaruh augmentasi data terhadap hasil akurasi klasifikasi. Metode augmentasi yang digunakan adalah *Synonym replacement*, *Random deletion*, dan *Random exchange*. Proses klasifikasi dilakukan dengan menggunakan kombinasi metode *deep learning* yaitu CNN dan LSTM dengan *fastText* dan *word2vec* sebagai metode *word embedding*. Skenario 1 adalah skenario yang melakukan klasifikasi teks *twitter* yang memiliki 2 label pengujian dilakukan pada data latih yang telah dilakukan augmentasi dan tanpa augmentasi. Skenario selanjutnya adalah skenario 2 yaitu skenario klasifikasi teks *twitter* ke 3 label yang berbeda. Dalam skenario ini, pengujian juga

yang berbeda juga klasifikasi dengan augmentasi dan tanpa augmentasi data data latih. Hasil ujicoba ditunjukkan pada Tabel 4. dan Tabel 5

Tabel 1 menunjukkan hasil ujicoba yang dilakukan menggunakan skenario 1. Hasil ujicoba menunjukkan bahwa metode CNN mencapai akurasi tertinggi bila dikombinasikan dengan *fastText* sebagai metode *word embedding* dengan nilai akurasi 0,69. Sedangkan akurasi terendah hanya 0,51 yang dihasilkan oleh kombinasi metode CNN dan *word2vec*. Kemudian, ujicoba menggunakan metode LSTM menunjukkan hasil akurasi tertinggi mencapai 0,68 bila dikombinasikan dengan *fastText* sebagai metode *word embedding*. Sedangkan akurasi terendah hanya 0,61 bila dikombinasikan dengan *word2vec* sebagai metode *word embedding*.

Selanjutnya, uji klasifikasi dengan augmentasi yang diterapkan pada data latih menunjukkan bahwa metode CNN mencapai akurasi tertinggi 0,79 bila dikombinasikan dengan *fastText* sebagai metode *word embedding*. Sedangkan metode CNN mencapai akurasi terendah dengan nilai 0,60 jika digabungkan dengan *word2vec*. Kemudian, kombinasi metode LSTM dan *fastText* ketika dikombinasikan dengan metode *word2vec*. Selanjutnya, kombinasi metode LSTM dan *fastText* menunjukkan hasil akurasi tertinggi mencapai 0,95. Sedangkan hasil akurasi terendah diperoleh pada kombinasi LSTM dan *word2vec* dengan nilai akurasi hanya 0,73.

Pada ujicoba skenario 2 *fastText* sebagai metode *word embedding*. Pengujian menggunakan metode LSTM menunjukkan hasil akurasi tertinggi mencapai 0,73 bila digabungkan dengan metode

*fastText*. Kemudian mencapai akurasi terendah dengan nilai 0,64 jika digabungkan dengan *fastText*.

Sementara itu, uji klasifikasi dengan augmentasi yang diterapkan pada data latih menunjukkan bahwa metode CNN mencapai akurasi tertinggi sebesar 0,80 bila dikombinasikan dengan *fastText*. Sedangkan akurasi terendah mencapai nilai 0,72 jika digabungkan dengan metode *fastText*. Kemudian, kombinasi metode LSTM dan *fastText* menunjukkan hasil akurasi tertinggi mencapai 0,73. Sedangkan hasil akurasi terendah diperoleh pada kombinasi LSTM dan *word2vec* dengan nilai akurasi sebesar 0,70.

Hasil ujicoba skenario 3 tanpa menggunakan augmentasi menunjukkan bahwa metode CNN mencapai akurasi tertinggi sebesar 0,74 bila dikombinasikan dengan *fastText*. Sedangkan metode CNN mencapai akurasi terendah sebesar 0,70 bila dikombinasikan dengan *fastText* sebagai metode *word embedding*. Pengujian menggunakan metode LSTM menunjukkan hasil akurasi tertinggi mencapai 0,79 jika digabungkan dengan metode *fastText*. Kemudian mencapai akurasi terendah dengan nilai 0,68 bila digabungkan dengan *word2vec*.

Sedangkan uji klasifikasi dengan augmentasi yang diterapkan pada data latih menunjukkan bahwa metode CNN mencapai akurasi tertinggi sebesar 0,90 bila dikombinasikan dengan *fastText*. Kemudian, akurasi terendah mencapai nilai 0,71. Semua pengujian yang dilakukan dengan menggunakan skenario 1, skenario 2, dan skenario 3 menunjukkan bahwa hasil akurasi klasifikasi multi-label tertinggi diperoleh ketika data latih yang dilakukan augmentasi. Hasil pengujian juga menunjukkan bahwa jumlah label yang digunakan dalam klasifikasi mempengaruhi hasil akurasi.

### 3.2. Pembahasan

Kemacetan merupakan hal yang biasa terjadi di kota-kota besar. Hal ini dapat merugikan pengguna jalan, oleh karena itu perlu dilakukan pendeteksian kejadian lalu lintas secara akurat dan efisien. Saat ini media sosial banyak digunakan sebagai sumber informasi untuk mendeteksi suatu kejadian, khususnya *twitter*. Setiap menit, pengguna *twitter* dapat berbagi informasi terkait peristiwa di sekitarnya seperti peristiwa lalu lintas. Namun, karena beragamnya diskusi pada teks *twitter*, maka perlu dilakukan klasifikasi teks *twitter* untuk mendapatkan informasi yang relevan. Pengguna *twitter* cenderung berbagi beberapa informasi sekaligus, sehingga satu *tweet* dapat memiliki lebih dari satu label. Oleh karena itu, perlu dilakukan pengklasifikasian multi-label pada masalah ini. Dalam hal kejadian lalu lintas, penelitian ini dapat

mengidentifikasi banyak situasi lalu lintas, seperti kondisi cuaca, kecelakaan lalu lintas, kemacetan lalu lintas, lalu lintas padat, lalu lintas lancar.

Penelitian ini menggunakan tiga skenario. Skenario pengujian ditunjukkan pada Tabel 3. Skenario pertama menjalankan data latih dengan augmentasi dan non-augmentasi menjadi tiga label. Skenario kedua adalah menjalankan data latih dengan augmentasi dan non-augmentasi menjadi empat label. Skenario ketiga adalah menjalankan data latih augmentasi dan non-augmentasi menjadi lima label. Keseluruhan skenario adalah untuk mengetahui pengaruh data uji yang berbeda pada data latih yang sama dan untuk mengetahui pengaruh jumlah label data uji pada data uji yang sama. dalam skenario ini menunjukkan bahwa semakin banyak data latih pada pembuatan model, maka semakin baik hasil akurasi percobaan.

Pengklasifikasian ketiga skenario dilakukan dengan augmentasi dan non-augmentasi pada data latih, percobaan dilakukan untuk melihat pengaruh augmentasi pada data latih, augmentasi yang digunakan dalam penelitian ini adalah *Synonym replacement*, *Random exchange*, dan *Random dilation*. Karena data penelitian yang digunakan dalam bahasa Indonesia, maka penambahan sinonim penggantian menggunakan daftar sinonim dari Kamus Besar Bahasa Indonesia (KBBI). Hasil akurasi keseluruhan yang diperoleh dengan menggunakan augmentasi lebih baik daripada klasifikasi non-augmentasi. Dari analisis yang dilakukan data latih yang dilakukan augmentasi dan data latih tanpa dilakukan augmentasi menunjukkan bahwa data dengan augmentasi mempengaruhi hasil klasifikasi pada teks *twitter*.

Pada hasil klasifikasi data latih tanpa menggunakan augmentasi menghasilkan klasifikasi yang tidak berkerja dengan baik. Seperti contoh data teks *twitter* “update truk yang terguling di lajur kiri kecelakaan tol waru arah perak sedang proses evakuasi” merupakan tweet bulan November 2020 yang seharusnya masuk pada topik kecelakaan lalu lintas. Namun pada hasil klasifikasi *tweet* tersebut masuk pada topik lalu lintas macet. selain itu terdapat berbagai topik *twitter* yang termasuk lalu lintas macet masuk kedalam lalu lintas lancar. Seperti pada contoh *tweet* “lima jalur ini macet citraraya unesa dua arah macet raya lontar arah citraland macet tol waru arah satelit macet dukuh kupang arah ngesong macet mayjen sungkono arah satelit macet “ berikut adalah contoh *tweet* pada bulan Januari 2021 yang termasuk kedalam topik

Tabel 6. Kesalahan Data Target

No	Teks <i>Twitter</i>	Target
1.	<b>Waspada</b> i kawan. Sebelum Pasar Menganti arah ke Lakarsantri MACET. Ekor kemacetan di depan Kantor Kecamatan Menganti.	[0,0,1,0,0] Lalu lintas macet
2.	14.29: <b>Waspada</b> i kepadatan di jalur-jalur ini 1. Underpass Satelit arah Mayjen Sungkono padat 2. HR Muhammad arah Mayjen Yono Suwoyo padat 3. Citraland Unesa dua arah 4. Raya Lontar dua arah padat	[0,0,0,1,0] Lalu lintas padat
3.	09.39: <b>Waspada</b> i kepadatannya 1. Mayjen Sungkono arah Adityawarman padat, simpul kepadatan di TL TVRI 2. Aloha arah Waru padat 3. Kedamean arah Bringkang padat	[0,0,1,0,0] Lalu lintas padat
4.	09.03: <b>Waspada</b> i kepadatannya. Aiptu I Gusti Putu - BM Lintas Polsek Porong mengabarkan, ada Dump Truk terguling di Arteri Porong arah Pasuruan. Posisinya di sebelah utara jembatan kebon agung.	[1,0,1,0,0] Lalu lintas padat

lalu lintas macet menjadi topik kecelakaan lalu lintas sehingga hasil akurasi data *tweet* dengan data augmentasi dan non augmentasi berbeda. Selain itu hasil dari kesalahan *tweet* tanpa augmentasi dikarenakan kesalahan topik juga menyebabkan akurasi yang rendah dibandingkan data dengan augmentasi. contoh data yang telah dilakukan augmentasi terdapat pada Tabel 7. Pada Tabel 7 bagian contoh dilakukan augmentasi menggunakan *Synonym replacement*, *Random exchange*, dan *Random deletion*. *Synonym replacement* sendiri yaitu mengganti kata n pada data latih dari kamus secara acak dengan sinonim kata n tersebut. Kemudian pada bagian 2 adalah contoh *Random exchange* adalah mengganti kata secara acak dengan sinonim yang ada pada data latih tersebut. Bagian 3 adalah contoh *Random deletion* adalah menukar posisi kata n secara acak pada suatu kalimat.

Pada skenario pertama dan kedua, klasifikasi ke dalam tiga kelas label dan empat label menunjukkan nilai akurasi yang baik karena data yang diolah hanya diklasifikasikan menjadi dua kelas dan tiga kelas. Hal ini disebabkan karena distribusi data yang tidak seimbang dikarenakan sedangkan model data latih memiliki lima label. Hasil klasifikasi tahap pertama ditunjukkan pada Tabel 4 dilakukan klasifikasi menggunakan metode *fastText* dan kombinasi *deep learning* (CNN dan LSTM). Tabel tersebut menunjukkan bahwa semua akurasi menunjukkan hasil yang baik. Sedangkan pada klasifikasi tahap kedua, eksperimen skenario kedua menunjukkan hasil akurasi yang kurang baik

Tabel 7. Sampel Data Augmentasi

No	Teks <i>Twitter</i>	Hasil Augmentasi	Jenis Augmentasi
1.	info awal kecelakaan di tol porong arah sidoarjo.	<b>informasi</b> awal <b>insiden</b> di tol porong arah sidoarjo	<i>Synonym replacement</i>
2.	info awal truk mogok di lajur kanan rangkai kenjeran waspada kawan.	<b>informasi</b> awal truk <b>macet</b> di <b>banjar</b> kanan rangkai <b>hati-hati</b> teman	<i>Random exchange</i>
3.	info awal kecelakaan beruntun di km tol porong arah sidoarjo melibatkan sekitar kendaraan info sudah diteruskan ke petugas	info <b>kendaraan</b> kecelakaan beruntun di km tol porong arah sidoarjo melibatkan sekitar <b>awal</b> info sudah diteruskan ke petugas	<i>Random deletion</i>

seperti yang ditunjukkan pada Tabel 5 dilakukan klasifikasi menggunakan metode *word2vec* dan kombinasi *deep learning* (CNN dan LSTM).

Hasil akurasi menunjukkan bahwa kombinasi *deep learning* dengan *fastText* menunjukkan akurasi yang lebih baik daripada menggunakan *word2vec*. Dari semua percobaan klasifikasi skenario pertama menunjukkan bahwa 38 dari 40 percobaan menunjukkan bahwa word embedding *fastText* lebih baik dari *word2vec* semua kombinasi metode yang tersedia. Selain itu klasifikasi skenario kedua, 37 dari 40 percobaan dan untuk klasifikasi tahap kedua menunjukkan bahwa *fastText* memiliki akurasi terbaik. Klasifikasi skenario ketiga, 35 dari 40 percobaan dan untuk klasifikasi tahap kedua menunjukkan bahwa *fastText* memiliki akurasi terbaik dari *word2vec*. *fastText* adalah algoritma yang memperhitungkan struktur internal kata dan representasi kata. Pendekatan representasi model *word embedding fastText* berbeda dengan *word embedding* lain seperti *word2vec*. *Word2Vec* menggunakan setiap kata sebagai unit terkecil, sedangkan *fastText* mengasumsikan kata-kata disusun oleh *n-gram*. selain itu keunggulan *fastText* adalah karena metode ini menyimpan kata-kata vektor sebagai *n-gram* karakter, dapat menemukan representasi vektor untuk kata-kata yang tidak langsung ditemukan dalam kamus (Kuyumcu, Aksakalli & Delil, 2019).

Namun terdapat kombinasi *word embedding* menggunakan *word2vec* lebih baik dibandingkan dengan *fastText*. Seperti hasil ujicoba

yang telah dilakukan pada skenario 1 terdapat 2 kombinasi menggunakan *word2vec* lebih baik. Pada skenario 2, terdapat 3 kombinasi dan Pada skenario 3 terdapat 5 kombinasi *word2vec* lebih baik dibandingkan kombinasi yang ada. Dari analisis yang dilakukan diketahui *fastText* dikembangkan dari model *word2vec* dengan memetakan suku kata. Namun dari hasil model klasifikasi perbedaan kinerja kedua *word embedding* yang tidak begitu signifikan menunjukkan bahwa kedua *word embedding* memiliki kinerja yang kompetitif pada saat klasifikasi.

Dari semua percobaan klasifikasi skenario pertama menunjukkan bahwa 37 dari 40 percobaan menunjukkan LSTM lebih baik dari semua kombinasi metode yang tersedia. Sedangkan untuk klasifikasi skenario kedua, 39 dari 40 percobaan dan untuk klasifikasi tahap kedua menunjukkan bahwa LSTM memiliki akurasi terbaik dari semua pasangan kombinasi yang ada. Klasifikasi skenario ketiga, 35 dari 40 percobaan dan untuk klasifikasi tahap kedua menunjukkan bahwa LSTM memiliki akurasi terbaik dari semua pasangan kombinasi yang tersedia. Hal ini karena LSTM tidak mengabaikan informasi semantik dari urutan kata sebelumnya dan menyimpan informasi tersebut dalam lapisan tersembunyi (Bdeir & Ibrahim, 2020). Menurut (Yan, Wang, Chun, & Yin, 2017) metode LSTM lebih unggul apabila digunakan untuk model time series. Sehingga LSTM cocok untuk menyelesaikan masalah teks *twitter* dengan model *sequence*.

Pada hasil klasifikasi, model yang dibangun menggunakan *word2vec* dan LSTM dari 1467 data pada uji bulan Februari 2020 sebanyak 54 data mengalami kesalahan klasifikasi. kesalahan klasifikasi seperti pada contoh data *tweet* "Waspadai kepadatannya. MERR Galaxy arah Rungkut padat" yang diklasifikasi pada label lalu lintas padat, namun masuk kedalam label lalu lintas macet. Kesalahan klasifikasi tersebut dikarenakan kata terdapat kata "waspadai" yang sering berada pada data *tweet* dan lalu lintas macet seperti pada contoh data *tweet* "Waspadai jalur-jalur ini underpass Satelit arah Mayjen Sungkono macet". Dari analisis yang dilakukan proses pra-proses yaitu menghilangkan kata - kata yang sering muncul (*stop word*) perlu dilakukan analisis terhadap data pada saat pra-proses dan pelabelan, seperti terlihat pada Tabel 6.

Analisis lain dilakukan terhadap skenario ujicoba yaitu skenario 1, skenario 2, dan skenario 3.

dari hasil ujicoba yang telah dilakukan dan didapatkan bahwa skenario 3 memiliki hasil akurasi lebih baik dibandingkan dengan skenario 1 dan skenario 2. Hal tersebut dikarenakan model data latih yang sama diuji terhadap data dengan jumlah label yang berbeda sehingga kesalahan klasifikasi label terjadi pada skenario 1 dan skenario 2. Sedangkan pada skenario 3 hasil dari model klasifikasi mendapatkan hasil nilai terbaik. Hasil dari ujicoba yang telah dilakukan klasifikasi multi-label menggunakan kombinasi *deep learning* dan *word embedding* didapatkan bahwa kombinasi *fastText* dan LSTM lebih baik dibandingkan dengan kombinasi lain. Pada multi-label non augmentasi akurasi tertinggi adalah 0,75. Sedangkan klasifikasi dengan augmentasi akurasi tertinggi yaitu 0,95. Jumlah banyak label dan augmentasi juga mempengaruhi hasil klasifikasi.

#### 4. KESIMPULAN

Klasifikasi multi-label dilakukan ke dalam lima kelas multi-label dengan menggunakan kombinasi metode *deep learning* (CNN dan LSTM) dan *word embedding* (*word2vec* dan *fastText*) dengan data augmentasi dan non-augmentasi. Hasil ujicoba Akurasi tertinggi diperoleh dari kombinasi LSTM dan *fastText*. Dikarenakan *fastText* dapat menemukan representasi vektor untuk kata-kata yang tidak langsung ditemukan dalam kamus. Hasil pengujian juga menunjukkan bahwa jumlah label yang digunakan dalam klasifikasi mempengaruhi hasil akurasi. Dikarenakan model data latih yang digunakan menggunakan 5 label namun diuji pada data 2 dan 3 label, sehingga label lain mengalami kesalahan pada saat dilakukan ujicoba. Selain itu, data latih yang dilakukan augmentasi secara keseluruhan membantu meningkatkan akurasi multi-label dalam ujicoba. Augmentasi melakukan penambahan data latih dengan cara menemukan sinonim pada data latih.

Pada penelitian ini didapatkan permasalahan pada proses ujicoba yaitu pelabelan pada data latih dilakukan secara manual. Pelabelan manual dilakukan analisis satu persatu pada data *tweet* dikarenakan dalam satu kalimat memiliki makna yang beragam, sehingga lebih akurat apabila dilakukan secara manual. Namun pelabelan secara manual cukup memakan waktu. Untuk itu diharapkan pada penelitian selanjutnya dapat mengeksplorasi terkait pelabelan pada data banyak. Sehingga penelitian selanjutnya dapat mengembangkan pelabelan pada data latih terkait lalu lintas yang lebih efisien.

## DAFTAR PUSTAKA

- AIPE, A., MUKUNTHA, N. S., EKBAL, A., & KUROSHASI, S. 2018. Deep learning approach towards multi-label classification of crisis related tweets. *Proceedings of the International ISCRAM Conference, 2018-May*(May), 705–717.
- ALOMARI, E., MEHMOOD, R., & KATIB, I. 2019. Road Traffic Event Detection Using Twitter Data , Machine Learning , and Apache Spark. <https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00332>
- ATEFAH, F., & KHREICH, W. 2015. A survey of techniques for event detection in Twitter. *Computational Intelligence, 31*(1), 133–164. <https://doi.org/10.1111/coin.12017>
- BDEIR, A. M., & IBRAHIM, F. 2020. A framework for arabic tweets multi-label classification using word embedding and neural networks algorithms. *ACM International Conference Proceeding Series*, 105–112. <https://doi.org/10.1145/3404512.3404526>
- BOUTELL, M. R., LOU, J., SHEN, X., & BROWN, C. M. 2004. Learning multi-label scene classification. *37*, 1757–1771. <https://doi.org/10.1016/j.patcog.2004.03.009>
- BRANNON, R. M., LIENGME, B. V., KRITHIKA, L. B., ROY, P., & JERLIN, M. A. 2017. Topic Identification and Categorization of Public Information in Community-Based Social Media Topic Identification and Categorization of Public Information in Community-Based Social Media. <https://doi.org/10.1088/1742-6596/755/1/011001>
- CHAMBY-DIAZ, J. C., & BAZZAN, A. L. C. 2019. Identifying traffic event types from twitter by multi-label classification. *Proceedings - 2019 Brazilian Conference on Intelligent Systems, BRACIS 2019*, 806–811. <https://doi.org/10.1109/BRACIS.2019.00144>
- DABIRI, S., & HAESLIP, K. 2019. Developing a Twitter-based traffic event detection model using deep learning architectures. *Expert Systems with Applications, 118*, 425–439. <https://doi.org/10.1016/j.eswa.2018.10.017>
- GU, Y., QIAN, Z., & CHEN, F. 2016. From Twitter to detector: Real-time traffic incident detection using social media data. *Transportation Research Part C: Emerging Technologies, 67*, 321–342. <https://doi.org/10.1016/j.trc.2016.02.011>
- JANG, B., KIM, I., & KIM, J. W. 2019. Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS ONE, 14*(8), 1–20. <https://doi.org/10.1371/journal.pone.0220976>
- JUANLIN, HU, XIN KANG, SHUN NISHIDE, F. R. 2019. Text multi-label sentiment analysis based on Bi-LSTM. *Proc. of CCIS*, 16–20.
- KHATTAK, F. K., JEBLEE, S., POU-PROM, C., ABDALLA, M., MEANEY, C., & RUDZICZ, F. 2019. A survey of word embeddings for clinical text. *Journal of Biomedical Informatics: X, 4*(October), 100057. <https://doi.org/10.1016/j.yjbinx.2019.100057>
- KUYUMCU, B., AKSAKALLIL, C., & DELIL, S. 2019. An automated new approach in fast text classification (*fastText*): A case study for Turkish text classification without pre-processing. *ACM International Conference Proceeding Series*, 1–4. <https://doi.org/10.1145/3342827.3342828>
- LIM, E., ISTTS, T. I., SETIAWAN, E. I., & ISTTS, T. I. 2020. Stance Classification Post Kesehatan di Media Sosial Dengan FastText Embedding dan Deep learning. 65–73.
- MAZOYER, B., CAGE, J., HERVE, N., & HUDELLOT, C. 2020. A french corpus for event detection on twitter. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, (May), 6220–6227.
- MOHAMMAD, F. 2018. Is preprocessing of text really worth your time for toxic comment classification? *2018 World Congress in Computer Science, Computer Engineering and Applied Computing, CSCE 2018 - Proceedings of the 2018 International Conference on Artificial Intelligence, ICAI 2018*, 447–453.
- PARWEZ, M. A., ABDULISH, M., & JAHIRUDDIN. 2019. Multi-Label Classification of Microblogging Texts Using Convolution Neural Network. *IEEE Access, 7*, 68678–68691. <https://doi.org/10.1109/ACCESS.2019.2919494>
- RAHMAWATI D., & KHODRA, M. L. 2016. Word2vec semantic representation in multi-label classification for Indonesian news article. *4th IGNITE Conference and 2016 International Conference on Advanced Informatics: Concepts, Theory and Application, ICAICTA 2016*, 0–5. <https://doi.org/10.1109/ICAICTA.2016.7803115>
- RAO, A., & SPASOJEVIC, N. 2016. Actionable and Political Text Classification using Word embeddings and LSTM. *Research Gate*, (July), 2–10.