

Ekstraksi Informasi Kecelakaan dari Twitter untuk Pemantauan Spasial Kecelakaan Lalu Lintas di Indonesia

Information Extraction from Twitter for Spatial Monitoring of Traffic Accidents in Indonesia

Della Fitrayani Budiono

Universitas Sebelas Maret
Surakarta, Indonesia

dellafitrayani@student.uns.ac.id

Lia Ristiana

Universitas Sebelas Maret
Surakarta, Indonesia

lia.ristiana@student.uns.ac.id

Ig. Donny Fernando

Universitas Sebelas Maret
Surakarta, Indonesia

donnyfernando@student.uns.ac.id

Abstract— Social media has now become one of the most significant source of information in the web today. Twitter, as one of the most popular social media platform in the world, enables its user to share and spread information to the masses in a fast and easy way. As of mid 2015, there were estimated to be around 50 millions of Twitter users in Indonesia, in which 77% of them were actively using Twitter everyday. Considering the huge number of Twitter users in Indonesia, the tweets shared by Twitter users can be used for event-detection. The data obtained from Twitter is extracted using some data mining technique into valuable information. In this paper we propose a method using text mining and data mining techniques to extract information of traffic accidents in Indonesia from tweets. The tweets fetched from Twitter is first preprocessed using text mining technique. The preprocessed data is then classified using Naive Bayes classifier into positive and negative, positive means the tweet is positively-linked to traffic accident in Indonesia and negative means the tweet is negatively-linked to traffic accident. The positively classified data is then extracted to obtain the location of the accident to be visualized in a map. From 600 tweets, with 224 classified as positive and 376 negative, used in the experiment using 10 fold cross validation, the accuracy obtained is 95% with 130 points of location successfully extracted and visualized into a map.

Keywords—data mining; Naive Bayes Classifier; Text Mining; Twitter

I. LATAR BELAKANG

Twitter sebagai salah satu media sosial yang paling populer di masyarakat saat ini berperan penting terhadap penyebaran dan penyampaian informasi di kalangan pengguna internet. Penyampaian informasi di Twitter dilakukan dalam bentuk teks dengan limit sebesar 140 karakter yang disebut dengan *tweet*. Sebagai microblogging service yang berdiri sejak tahun 2006, Twitter dikenal memiliki karakteristik berupa sifatnya yang terbuka untuk publik dan tersedia di web. Pengguna internet yang memiliki akun di Twitter dapat menyampaikan

informasi dalam bentuk *tweet*, mulai dari pengguna akun pribadi hingga akun-akun resmi milik Pemerintah dan instansi-instansi lainnya. Di Indonesia sendiri, diperkirakan terdapat 50 juta pengguna Twitter pada pertengahan 2015, di mana 77% di antaranya aktif menggunakan Twitter setiap hari (Maulana, 2016).

Melihat dari besarnya basis pengguna Twitter, khususnya di Indonesia, dan sifatnya yang memungkinkan untuk diakses oleh siapa saja dan kapan saja secara cepat dan mudah, dapat memberikan banyak keuntungan berupa informasi yang bermanfaat. Keberadaan media sosial kini memungkinkan adanya deteksi bermacam-macam kejadian, bahkan kejadian yang sangat kecil sekalipun, seperti adanya mobil rusak di tepi suatu jalan (Gutiérrez et al, 2015). Mekanisme pendeteksian suatu kejadian menggunakan data mining pada media sosial seperti Twitter ini sering disebut dengan istilah deteksi kejadian (*event detection*).

Dalam *event-detection* secara umum, terdapat dua macam *events*, yaitu *small-scale events* seperti laporan lalu lintas, kebakaran, dan kecelakaan lalu lintas, dan *large-scale events* seperti bencana alam, gempa bumi, pemilihan presiden, dan sebagainya. Pada penelitian ini penulis memfokuskan pada deteksi kejadian kecelakaan lalu lintas yang termasuk ke dalam kategori *small-scale events*. Ciri khas dari *small-scale events* yang perlu diperhatikan di sini adalah ia hanya memiliki jumlah *tweets* yang sedikit, terjadi di suatu lokasi

geografis yang spesifik, dan berlangsung pada interval waktu yang cenderung singkat (Sonune et al, 2016).

Beberapa penelitian mengenai *event-detection* pada Twitter telah dilakukan sebelumnya, seperti melakukan deteksi kejadian yang berhubungan dengan lalu lintas di United Kingdom untuk tweets berbahasa Inggris menggunakan algoritma support vector machine (SVM) (Gutiérrez et al, 2015), deteksi lalu lintas *tweets* berbahasa Inggris yang real-time (Boshale & Kokate, 2014). Namun, penelitian mengenai kejadian kecelakaan di Indonesia berdasarkan *tweets* berbahasa Indonesia masih belum dilakukan.

Informasi mengenai kejadian kecelakaan dapat dengan mudah tersampaikan dan tersebar di Twitter dalam bentuk informasi lokasi dan waktu kejadian. Di Indonesia *tweet* mengenai kejadian kecelakaan aktif diberitakan oleh akun-akun Twitter seperti portal berita (@kompascom, @OfficialNETNews, @antaranews, dsb), channel radio (@RadioElshinta, @rpkd fm, dsb), Traffic Management Center (salah satu divisi Polri) (@TMCPoldaMetro, @RTMC_Jogja, @TMCPolresBogor, dsb), dan situs penyedia informasi lalu lintas publik (@infomacetcom, @lewatmana, @PasangMata, dsb), serta akun-akun pribadi.

Namun, tidak semua *tweet* dengan kata 'kecelakaan' yang dihasilkan baik oleh akun-akun tersebut di atas maupun akun-akun lain di Twitter benar-benar berhubungan dengan kecelakaan. Secara umum, *tweet* yang mengandung kata 'kecelakaan' dapat dibagi menjadi dua jenis, yaitu positif dan negatif. Positif berarti *tweet* tersebut memiliki kaitan dengan kecelakaan lalu lintas, sementara negatif berarti tidak berkaitan dengan kecelakaan lalu lintas. *Tweet* yang positif berkaitan dengan kejadian kecelakaan umumnya mengandung dua macam informasi, yaitu berupa lokasi dan waktu kejadian kecelakaan. Dalam makalah ini, penulis mengajukan metode ekstraksi data lokasi kejadian kecelakaan untuk dipetakan berdasarkan *tweet* yang telah diklasifikasi.

Dalam melakukan data mining pada Twitter, perlu diperhatikan bahwa data yang ada semuanya berbentuk

unstructured, sehingga dalam hal ini proses ekstraksi data masuk ke dalam konteks big data. Data mining pada Twitter memiliki beberapa tantangan khusus, seperti 1) jumlah karakter pada *tweet* yang terbatas sebesar 140 karakter membuat teks yang dihasilkan memiliki banyak singkatan, 2) *tweet* banyak menggunakan bahasa tidak formal, 3) kemungkinan adanya kesalahan ketik atau eja pada kata-kata dan kesalahan tata bahasa, 4) sifat data yang tidak terstruktur (Telange, 2016). Tantangan ini dapat ditangani dengan melakukan *text mining* terhadap *tweets* yang diperoleh. Secara umum, pertama dilakukan *fetching tweets* menggunakan Twitter API yang tersedia, lalu *tweets* yang telah diperoleh akan diklasifikasikan menjadi positif atau negatif menggunakan metode Naive Bayes Multinomial dengan 10 fold cross validation. Selanjutnya dilakukan ekstraksi data terhadap *tweets* untuk memperoleh lokasi kejadian kecelakaan. Terakhir, hasil ini akan divisualisasikan dalam peta untuk menunjukkan lokasi kecelakaan.

II. TUJUAN DAN MANFAAT

Tujuan dari penelitian adalah untuk memperoleh informasi kejadian kecelakaan di Indonesia berdasarkan *tweet* yang telah diklasifikasi antara positif berhubungan dengan kecelakaan lalu lintas atau negatif untuk dipetakan berdasarkan lokasinya. Informasi yang diperoleh dari data mining diharapkan dapat memberikan gambaran kepada Pemerintah maupun masyarakat mengenai lokasi kejadian kecelakaan di Indonesia.

III. BATASAN PENELITIAN

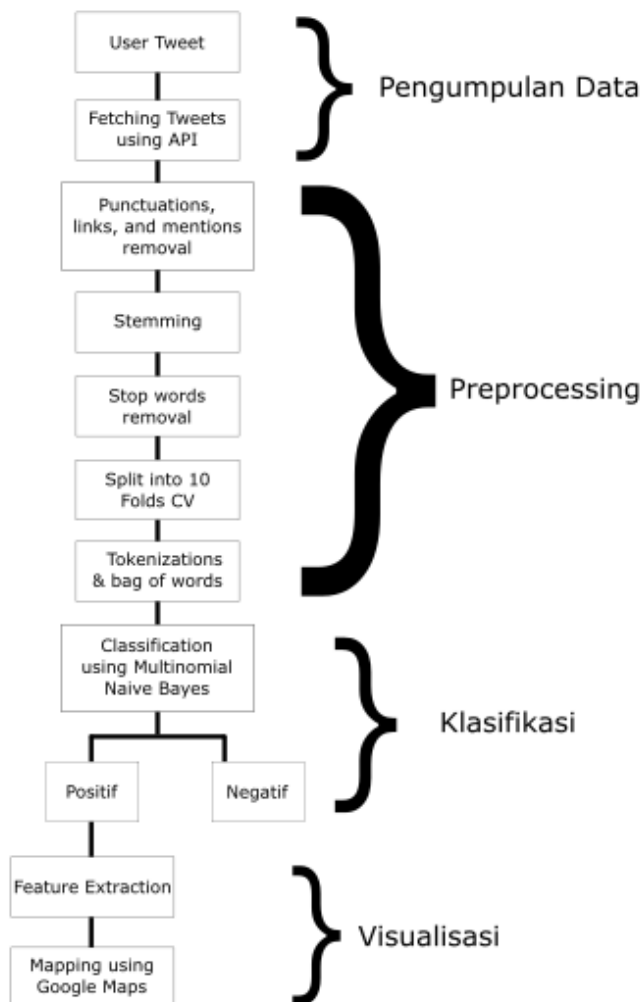
Penelitian dilakukan dengan batasan sebagai berikut.

1. Data yang diolah adalah *tweets* yang menggunakan Bahasa Indonesia, baik formal maupun non formal.
2. *Tweet* yang dianggap positif berhubungan dengan kecelakaan lalu lintas adalah *tweet* yang mengandung informasi fakta kecelakaan yang terjadi di Indonesia saja.
3. Fokus penelitian adalah ekstraksi informasi lokasi kejadian kecelakaan dengan mengabaikan waktu kejadian.
4. Pengambilan data dari Twitter dilakukan dengan

mengakses 3200 *tweets* terakhir (batasan Twitter API) yang dipos oleh masing-masing akun tanpa menspesifikasikan rentang waktu, lalu difilter untuk mengambil *tweet* yang mengandung kata 'kecelakaan'.

IV. METODE YANG DIGUNAKAN

Metode penelitian dilakukan dalam beberapa tahap. Secara garis besar tahapan dibagi menjadi 4 bagian yaitu pengumpulan data, preprocessing, klasifikasi, dan visualisasi. Secara garis besar, tahapan yang digunakan pada penelitian ini bisa dilihat pada skema di bawah ini :



Gambar 4.1. Skema tahapan penelitian.

Pengumpulan data pada penelitian ini dilakukan dengan menggunakan *Application Programming Interface* (API) yang disediakan oleh Twitter. Dengan menggunakan

API Twitter, dilakukan *query* untuk melakukan pengambilan *tweets* yang mengandung kata 'kecelakaan' di dalamnya. Hasil dari proses ini adalah sebuah file JSON yang berisi data *tweets* beserta metadatanya. Salah satu objek JSON tersebut adalah *text* yang merupakan isi *tweet* pengguna.

Pada tahap selanjutnya, yaitu preprocessing, teks *tweet* yang diperoleh kemudian diolah menggunakan metode text mining sebagai berikut.

- Punctuations, mentions, and links removal

Suatu *tweet* umumnya mengandung tidak hanya teks, namun juga hal-hal seperti tautan, tanda baca, dan *mention* (penyebutan pengguna lain dengan tanda (@) diikuti username). Agar memudahkan proses selanjutnya, maka atribut-atribut kurang penting ini perlu dihilangkan.

- Stemming

Dari tahap pembersihan *tweet* tadi, selanjutnya dilakukan stemming yang bertujuan untuk menemukan kata dasar dari setiap kata pada teks dengan menghilangkan imbuhan, baik awalan, sisipan, maupun akhiran.

- Stop word removal

Setelah stemming, data dimasukkan ke dalam proses *stop word removal* atau filtering, yang bertujuan untuk menghilangkan kata-kata tidak penting yang sering muncul, yang dapat mempengaruhi proses klasifikasi. Dalam Bahasa Indonesia, stop words ini contohnya kata 'yang', 'ke', 'adalah', dan sebagainya.

- Tokenisasi

Tokenisasi adalah pemecahan dokumen ke dalam bagian – bagian tertentu. Dalam kasus ini yang digunakan adalah unigram, dengan kata lain satu bagian terdiri dari 1 kata tunggal.

- *Bag of words*

Bag of words merupakan tabel daftar kata-kata yang ada di dalam dokumen. Dengan kata lain, setelah ditokenisasi (dalam kasus ini unigram), tiap kata yang sudah dipecah dimasukkan ke dalam tabel kata, dan dalam kasus ini (twitter), tiap kata yang ada tersebut akan dihitung frekuensi kemunculan pada tiap *tweet*.

Tahap berikutnya yaitu klasifikasi. Data yang telah melalui preprocessing kemudian diklasifikasikan ke dalam positif atau negatif menggunakan data mining metode Naive Bayes multinomial. Pengkategorian *tweet* untuk training data dilakukan secara manual dengan contoh sebagai berikut.

No	Isi tweet	Kategori
1.	9 orang tewas akibat kecelakaan di Jalan Kolonel Masturi. Kecelakaan diduga akibat rem tidak berfungsi.	Positif
2.	Gunakan jalur sesuai ketentuan, jangan melawan arus krn rawan kecelakaan. Kesadaran kita keselamatan semua.	Negatif
3.	06.35 Penanganan kecelakaan tunggal mobil Sedan B 602 SD di Km 00 Interchange Cawang arah Pancoran. Hati2 bila melintas	Positif
4.	Bus media peliputan Olimpiade Rio kecelakaan, 3 wartawan cedera	Negatif
5.	info dr tukang becak yg mangkal disana sudah sering terjadi kecelakaan disana namun blm ada tindakan dr pemda	Negatif

Tabel 4.1. Tabel contoh pengkategorian *Tweets* pada dataset

Pada Tabel 4.1 dapat dilihat beberapa hal, tweet nomor 2 diklasifikasikan sebagai negatif karena isi tweet hanya berupa himbauan dan tidak menginformasikan adanya kejadian kecelakaan lalu lintas. Sementara nomor 4 diklasifikasikan negatif karena menginformasikan kejadian kecelakaan yang terjadi bukan di Indonesia. Sedangkan nomor 5 adalah informasi umum yang cenderung merujuk pada opini dan tidak menginformasikan suatu kejadian spesifik sebagai fakta.

Pada tahap terakhir yaitu visualisasi, *tweet* yang berhasil diklasifikasikan positif akan diekstrak informasinya.

Informasi yang diekstrak adalah informasi lokasi kejadian kecelakaan, caranya adalah dengan mendeteksi munculnya kata ‘jl’, ‘jln’, ‘jalan’, atau ‘tol’ (yang biasanya sering terjadi kecelakaan), dan selanjutnya akan diambil 2 – 4 kata setelah kata ‘jl’, ‘jln’, ‘jalan’, dan ‘tol’ dideteksi. Selanjutnya akan dicari lokasi kejadian dengan menggunakan kata – kata tersebut dengan menggunakan *Google Maps API Geocode*, jika lokasi berhasil ditemukan maka akan ditampilkan dalam peta untuk menunjukkan lokasi kejadian kecelakaan lalu lintas menggunakan *Google Maps API* pula.

V. DESAIN DAN IMPLEMENTASI

Data yang diperoleh dari Twitter API adalah berupa sebanyak 600 *tweets* yang mengandung kata 'kecelakaan' di dalamnya, yang di antaranya berasal dari akun resmi portal berita, situs penyedia informasi kecelakaan, channel radio, *traffic management center*, dan akun-akun pribadi yang diambil secara acak. Data disimpan dan diberi kategori secara manual, yaitu positif jika berhubungan dengan kecelakaan lalu lintas dan negatif jika tidak berhubungan dengan kecelakaan lalu lintas. Dari 600 data, 224 di antaranya mendapat klasifikasi positif dan 376 sisanya mendapat klasifikasi negatif. Untuk lebih jelasnya bisa dilihat pada lampiran Gambar 5.1. Sedangkan untuk melihat dataset lebih lengkapnya, bisa dilihat pada *file* ‘combinetrain4.csv’ yang disertakan pula dengan dokumen ini.

Data yang telah ditandai sesuai kategorinya kemudian dimasukkan ke dalam tahap preprocessing untuk dibersihkan, dimulai dari 1) pembersihan dari tanda baca, mentions, dan tautan, 2) stemming, 3) stop words removal, 4) tokenisasi, dan 5) *bag of words*. Dataset yang sudah dibersihkan ini akan dibagi diuji dengan menggunakan 10 Fold Cross Validation, dengan kata lain, 600 tweets yang diperoleh tadi akan dibagi menjadi 10 bagian, dimana setiap bagian terdiri dari 30 tweets. Setelah dibagi menjadi 10 bagian, akan dilakukan pengujian 10 kali, sehingga setiap bagian memiliki kesempatan diuji sekali. Sebagai contoh, pada pengujian pertama, bagian ke-1 akan dijadikan testing set, dan 9 bagian lainnya (bagian 2,3,4,5,6,7,8,9,10) akan dijadikan training set.

Selanjutnya, pada pengujian berikutnya, bagian ke-2 akan dijadikan testing set, dan 9 bagian lainnya (bagian 1,3,4,5,6,7,8,9,10) akan dijadikan training set, dan begitu seterusnya hingga pengujian ke – 10. Dari setiap pengujian akan dihitung akurasi, dan akurasi akhir diperoleh dari rata – rata akurasi dari seluruh pengujian. Hasilnya bisa dilihat pada lampiran Gambar 5.2 (hasil yang ditunjukkan adalah akurasi tiap pengujian, tweets yang salah diklasifikasikan dan rata – rata akurasi dari setiap pengujian).

Dari hasil pengujian tersebut diperoleh total akurasi akhir adalah 95% dengan menggunakan 10 Fold Cross Validation.



Gambar 5.3. Visualisasi hasil pemetaan.

Sedangkan daftar lokasi yang berhasil diekstrak dari tweet dan dipetakan bisa dilihat di lampiran Gambar 5.4.

Dari 224 data positif berhubungan dengan kecelakaan lalu lintas, 130 data di antaranya berhasil diekstraksi untuk memperoleh titik lokasi yang dapat divisualisasikan pada Google Maps. Untuk lebih jelasnya mengenai hasil implementasi dan source code yang digunakan, bisa dilihat pada *file* ‘tautan_repository_source_code.txt’ (tercantum tautan *repository* penelitian ini).

VI. ANALISIS

Setelah selesai preprocessing, data selanjutnya melalui tahap klasifikasi. Tahap klasifikasi menggunakan algoritma Naive Bayes dengan 10 fold Cross Validation memberikan hasil akurasi sebesar 95% di mana dari 600 data, 30 di antaranya diklasifikasikan secara salah dan sisanya benar. Beberapa contoh pengklasifikasian yang salah adalah sebagai berikut.

no	Tweet	Kategori	Hasil klasifikasi	Analisis error
1	korban tewas celaka pesawat virginia enam orang	Negatif	Positif	Kemungkinan kesalahan klasifikasi menjadi positif diakibatkan penggunaan kata-kata seperti 'tewas' dan 'korban' yang banyak terdapat pada dataset, sedangkan tweet dengan kata 'pesawat' (yang menyebabkan data menjadi negatif) tidak banyak terdapat pada dataset.
2	mohon maaf jalan ganggu celaka truk ranji ciawi rtmcjabar	Positif	Negatif	Kesalahan klasifikasi kemungkinan disebabkan oleh penggunaan kata 'mohon' dan 'maaf' yang jarang terdapat pada data bernilai positif.
3	arus lalin wilayah kab malang pantau lancar	Negatif	Positif	Kemungkinan kesalahan klasifikasi menjadi negatif diakibatkan penggunaan

1	korban tewas celaka pesawat virginia enam orang	Negatif	Positif	Kemungkinan kesalahan klasifikasi menjadi positif diakibatkan penggunaan kata-kata seperti 'tewas' dan 'korban' yang banyak terdapat pada dataset, sedangkan tweet dengan kata 'pesawat' (yang menyebabkan data menjadi negatif) tidak banyak terdapat pada dataset.
2	mohon maaf jalan ganggu celaka truk ranji ciawi rtmcjabar	Positif	Negatif	Kesalahan klasifikasi kemungkinan disebabkan oleh penggunaan kata 'mohon' dan 'maaf' yang jarang terdapat pada data bernilai positif.
3	arus lalin wilayah kab malang pantau lancar	Negatif	Positif	Kemungkinan kesalahan klasifikasi menjadi negatif diakibatkan penggunaan

	apabila jumpa celaka penyebab macet hub pos polisi dekat			kata-kata seperti 'apabila' dan 'hubungi' yang cenderung merupakan himbauan dan banyak tweet pada dataset yang negatif jika berbentuk himbauan.
4	giat patroli r satlantas polres dumai antisipasi macet celaka berlalu lin tas	Positif	Negatif	Kesalahan klasifikasi kemungkinan diakibatkan penggunaan kata-kata 'satlantas', 'polres' yang umum muncul pada data yang positif. Selain itu, pada data ini juga terdapat kata yang gagal di-stemming, yaitu 'berlululintas'.
5	korban tipu mah ga sakit korban celaka biasa tpi kalo jdi korban asa kmu	Negatif	Positif	Kesalahan klasifikasi kemungkinan diakibatkan oleh kurangnya tweet pada dataset yang menggunakan kata-kata

	rasa sakit bgt			bahasa Indonesia non formal.
--	-------------------	--	--	------------------------------------

Tabel 6.1. Tabel analisis kesalahan klasifikasi pada data yang telah melalui preprocessing

Dari 224 data positif dengan akurasi klasifikasi sebesar 95%, ternyata tidak semua data yang terklasifikasi benar dan positif berhasil divisualisasikan dalam peta. Dari 224 data, 130 di antaranya berhasil diekstrak titik lokasinya pada peta. Hal ini karena tidak semua tweet positif yang terklasifikasi dengan benar memiliki informasi lokasi yang dapat diekstrak dan dipetakan ke Google Maps.

VII. KESIMPULAN

Pada penelitian ini penulis berhasil melakukan *event-detection* untuk kejadian kecelakaan lalu lintas di Indonesia berdasarkan data yang diambil dari Twitter dengan sampel data sebanyak 600 *tweets*. Penelitian dilakukan dengan menggunakan metode text mining dan algoritma klasifikasi Naive Bayes 10 fold cross validation. Hasil akurasi dari klasifikasi ini adalah sebesar 95%. *Tweets* yang diklasifikasi positif berhubungan dengan kecelakaan lalu lintas di Indonesia kemudian divisualisasikan pada peta sesuai lokasi kejadian kecelakaan. Hasilnya, diperoleh sebanyak 130 titik lokasi yang berhasil divisualisasikan pada peta.

REFERENSI

BHOSALE, S., KOKATE, S. (2014). TRAFFIC DETECTION USING TWEETS ON TWITTER SOCIAL NETWORK. *INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH (IJSR)*, (PP. 2037-2039).

GUTIÉRREZ, C., FIGUERIAS, P., OLIVEIRA, P., COSTA, R., & JARDIM-GONCALVES, R. (2015, JULY). TWITTER MINING FOR TRAFFIC EVENTS DETECTION. IN *SCIENCE AND INFORMATION CONFERENCE (SAI), 2015* (PP. 371-378). IEEE.

MAULANA, A., 23 MARET 2016. CNN INDONESIA: *TWITTER RAHASIAKAN JUMLAH PENGGUNA DI INDONESIA*.

<http://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia/>. Diakses pada 20 Agustus 2016 pukul 17.20 WIB.

SONUNE, P. S., SHAIKH, S. K., SONAVANE, R. L., & SHISHUPAL, R. (2016). TWITTER STREAM ANALYSIS FOR TRAFFIC DETECTION AND EARTHQUAKE. *JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER ENGINEERING*, VOL.– 04, ISSUE – 01, (PP. 705-709).

TELANG, H. (2016). REAL-TIME TRAFFIC DETECTION BY ANALYZING TWITTER'S TWEET. *INTERNATIONAL JOURNAL OF EMERGING TECHNOLOGY AND COMPUTER SCIENCE*, 1(2).

LAMPIRAN

,OLIMPIADE 2016 - Pelatih kano Jerman cedera serius akibat kecelakaan di taksi https://t.co/a6Y2qpCfkw,F
,Kepala Ombudsman tewas dalam kecelakaan usai investigasi https://t.co/ePhbNtLrr0,F
,Bus pemain Persib alami kecelakaan tol Jakarta-Cikampek https://t.co/KnoQy6VIif,T
,Start Tour Singkarak Etape V diundur akibat kecelakaan https://t.co/PV0fb14uaf,F
,Polres Madiun tangani kecelakaan tewaskan sopir truk https://t.co/6uMV8F2YWz,T
,OLIMPIADE 2016 - Van Vleuten tegaskan tidak ada masalah setelah kecelakaan https://t.co/CbZxemXPdX,F
,",OLIMPIADE 2016 - Pesepeda Italia kecelakaan, tulang selangkanya patah https://t.co/MS01EXsyjv",F
,Polisi identifikasi korban kecelakaan di tol Cirebon https://t.co/BP3StuQMUZ,T
,",Stop melanggar! ,karna kecelakaan berawal dari pelanggaran",F
,Melihat Kecelakaan Kendaraan Mogok Hub Info Tol amp Bantuan tol mms ,F
,Seorang pengemudi pesawat sangat ringan terpaksa berada di atas pohon selama jam tempat dia mengalami kecelakaan ,F
,",stop melanggar! ,karna kecelakaan berawal dari pelanggaran .",F
,",selalu berhati hati dan fokus dalam berkendara ,mampu mengurangi resiko terjadinya kecelakaan dlm berlalu lintas.",F
,",Jangan mengemudikan kendaraan dlm kondisi mabuk atau terpengaruh minuman keras/Narkoba, krn rawan kecelakaan.",F
,",06:10 wita Jangan mengemudikan kendaraan dlm kondisi mabuk atau terpengaruh minuman keras/Narkoba, krn rawan kecelakaan.",F
,",18:57 wita Gunakan jalur sesuai ketentuan, jangan melawan arus karena rawan kecelakaan. Kesadaran kita keselamatan bersama",F

Gambar 5.1. Dataset yang diperoleh dari Twitter API dan telah dikategorikan. 'T' berarti positif berhubungan dengan kecelakaan lalu lintas di Indonesia. 'F' berarti negatif.

Total data : 600 with data perfold : 60
korban tewas celaka pesawat virginia enam orang --- classification is T. It should be F
olimpiade sepeda italia celaka tulang selangka patah --- classification is T. It should be F
lihat celaka kendara mogok hub info tol amp bantu tol mms --- classification is T. It should be F
orang kemudi pesawat sangat ringan paksa atas pohon lama jam tempat alami celaka --- classification is T. It should be F
wt apabila temu celaka maupun macet segera hubungi polisi dekat lalu lintas --- classification is T. It should be F
wita ingat korban celaka banyak jln tetap hati waspada --- classification is T. It should be F
lihat celaka kendara mogok hub info tol amp bantu tol mms --- classification is T. It should be F
Test 0 : 53 is correct. The accuracy is 88.3333333333 percent.

arus lalin wilayah kab malang pantau lancar apabila jumpa celaka penyebab macet hub pos polisi dekat --- classification is T. It should be F
wil kab malang gujur hujan deras tetap jaga jarak ingat jalan licin rawan terjadi celaka --- classification is T. It should be F
hampir celaka telpon pake hp kendara headseat tepi kendara telepon --- classification is T. It should be F
satlantas polrestabes sby laku tindak hadap langgar bonceng yg akibat celaka lalu lintas --- classification is F. It should be T
tol cpk putih arah cawang sdh stuk mulai kelapa gading imbas celaka exit karawang via --- classification is T. It should be F
orang tewas akibat celaka jalan kolonel masturi celaka duga akibat rem fungsi net --- classification is F. It should be T
Test 1 : 54 is correct. The accuracy is 90 percent.

bus media liput olimpiade rio celaka wartawan cedera --- classification is T. It should be F
celaka maut --- classification is T. It should be F
video celaka maut --- classification is T. It should be F
Test 2 : 57 is correct. The accuracy is 95 percent.

korban tipu mah ga sakit korban celaka biasa tpi kalo jdi korban asa kmu rasa sakit bgt --- classification is T. It should be F
Test 3 : 59 is correct. The accuracy is 98.3333333333 percent.

ngebut d jalan faktor utama celaka tertib lah kendara harga sama jalan nya --- classification is T. It should be F
giat patroli r satlantas polres dumi antisipasi macet celaka berlalulintas --- classification is T. It should be F
orang tewas celaka bus mesir --- classification is T. It should be F
rtmc dilantas polda riau rtmc dilantas polda riau duri celaka tsb tangan tugas --- classification is T. It should be F
tingkat celaka dubai turun persen server global blackberry lumpuh survei the national --- classification is T. It should be F
kembar celaka pulang meruyaa yaalah pikir bom kabar motor aja ampe --- classification is T. It should be F
masyarakat yka yg tdk terlalu penting tdk keluar rumah utama jalan raya kabut debu tebal jarak pandang meter rawan celaka --- classification is T. It should be F
Test 4 : 53 is correct. The accuracy is 88.3333333333 percent.

aaa tadi malam gak nonton sampe selesai sih tidur depan tv nonton cuma sampe pas celaka doang kan tinggal cerita --- classification is T. It should be F
aku buat ngerapihin ga ngefek rahang habis celaka parah kalau mau bener jaw surgery dulu wkwkwkw --- classification is T. It should be F
Test 5 : 58 is correct. The accuracy is 96.6666666667 percent.

Test 6 : 60 is correct. The accuracy is 100 percent.

celaka bunuh ampo sekitarkita takut hub pin bb cdc --- classification is T. It should be F
Test 7 : 59 is correct. The accuracy is 98.3333333333 percent.

mohon maaf jalan ganggu celaka truk ranji ciawi rtmcjabar --- classification is F. It should be T
Test 8 : 59 is correct. The accuracy is 98.3333333333 percent.

lucu celaka paling ajaib panjang masa --- classification is T. It should be F
studi laku buah usaha asuransi inggris tutur pakai minim tingkat angka celaka --- classification is T. It should be F
Test 9 : 58 is correct. The accuracy is 96.6666666667 percent.

Total accuracy: 95.00

Gambar 5.2. Hasil pengujian dengan Naive Bayes 10 fold cross validation.

Total places : 130

Jl. Tol Jkt - Cikampek, Kalihurip, Cikampek, Kabupaten Karawang, Jawa Barat 41373, Indonesia
Jl. Tol Palimanan - Kanci, Argasunya, Harjamukti, Kota Cirebon, Jawa Barat 45145, Indonesia
Kb. Jeruk I, Kb. Jeruk, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia
Plaza Tol Bukit Jelutong, 40150 Shah Alam, Selangor, Malaysia
Kb. Jeruk I, Kb. Jeruk, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia
Jl. Tol Jagorawi, Indonesia
Jl. H. R. Koroh, Nusa Tenggara Tim., Indonesia
Surabaya, Surabaya City, East Java, Indonesia
Jl. Kedung Cowek, Kota SBY, Jawa Timur, Indonesia
Jl. Tol Cipularang, Jawa Barat, Indonesia
Jakarta, Special Capital Region of Jakarta, Indonesia
Jl. Tol Lingkar Luar, Cakung Bar., Cakung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13910, Indonesia
Mini, Lake Isabella, MI 48893, USA
K-M Rd, Lewistown, MT 59457, USA
Jakarta Inner Ring Road, Grogol, Grogol petamburan, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia
Jl. Tol Serpong, Pd. Jaya, Pd. Aren, Kota Tangerang Selatan, Banten 15220, Indonesia
Jl. Raya Sukabumi, Sukamaju, Kec. Cianjur, Kabupaten Cianjur, Jawa Barat 43215, Indonesia
Jl. Tol Jkt - Cikampek, Kalihurip, Cikampek, Kabupaten Karawang, Jawa Barat 41373, Indonesia
Gerbang Tol Cikunir 1, Jakasampurna, Bekasi Bar., Kota Bks, Jawa Barat, Indonesia
Jakarta Inner Ring Road, Senayan, Kby. Baru, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12190, Indonesia
Jakarta, Special Capital Region of Jakarta, Indonesia
Jl. MT. Haryono Dalam, Balikpapan Sel., Kota Balikpapan, Kalimantan Timur, Indonesia
Tol Tangerang, Parung Jaya, Central Karang, Tangerang City, Banten 15159, Indonesia
Jl. Lingkar, Kota Tangerang Selatan, Banten, Indonesia
Jl. Teuku Umar, Indonesia
Jl. Gatot Subroto, Banten, Indonesia
Gg. Jnana, Tonja, Denpasar Utara, Kota Denpasar, Bali, Indonesia
Jl. Imam Bonjol, Pemecutan Klod, Denpasar Bar., Kota Denpasar, Bali 80119, Indonesia
Jl. Anjatan II, Rangdu, Pusakajaya, Kabupaten Subang, Jawa Barat, Indonesia
Jl. Wr. Supratman, Bekasi, Jawa Barat, Indonesia
Jl. Tirtanadi I, Sanur Kauh, Denpasar Sel., Kota Denpasar, Bali, Indonesia
Jl. Mahendradata, Kabupaten Gianyar, Bali, Indonesia
Jl. Patimura, Kota Batu, Jawa Timur, Indonesia
Jl. Yudistira, Seminyak, Kuta, Kabupaten Badung, Bali 80361, Indonesia
Jl. Sedap Malam, Kota Denpasar, Bali, Indonesia
Jl. Melati, Kota Denpasar, Bali, Indonesia
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. Nangka Raya, Tj. Bar., Jagakarsa, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12530, Indonesia
Jl. Gatot .S, Pagesangan Tim., Kec. Mataram, Kota Mataram, Nusa Tenggara Bar. 83127, Indonesia
Jl. Tangkuban Perahu, Lembang, Kabupaten Bandung Barat, Jawa Barat 40391, Indonesia
Jl. Wr. Supratman, Bekasi, Jawa Barat, Indonesia
Jl. Lintas Tim., Kisaran Tim., Kabupaten Asahan, Sumatera Utara, Indonesia
Jl. Imam Bonjol, Pemecutan Klod, Denpasar Bar., Kota Denpasar, Bali 80119, Indonesia

Jl. Mahendradata, Kabupaten Gianyar, Bali, Indonesia
Jl. Cokroaminoto, Bali, Indonesia
Jl. Cok Agung Tresna, Denpasar Tim., Kota Denpasar, Bali, Indonesia
Jl. Mahendradata, Kabupaten Gianyar, Bali, Indonesia
Jl. Raya Pemogan, Pemogan, Denpasar Sel., Kota Denpasar, Bali, Indonesia
Jl. Gajah Mada, Indonesia
Gg. Kulma, Padangsambian, Denpasar Bar., Kota Denpasar, Bali 80119, Indonesia
Jl. Gn. Sopotan, Denpasar Bar., Kota Denpasar, Bali, Indonesia
Jl. Mahendradata, Kabupaten Gianyar, Bali, Indonesia
Calle Alonso de la Raya, 38450 Garachico, Santa Cruz de Tenerife, Spain
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. Sunset Road, Kabupaten Badung, Bali, Indonesia
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. Imam Bonjol, Pemecutan Klod, Denpasar Bar., Kota Denpasar, Bali 80119, Indonesia
Jl. A. Yani, Blimbing, Kota Malang, Jawa Timur 65126, Indonesia
Jl. Diponegoro, Indonesia
Jl. Teuku Umar, Indonesia
Jl. Hayam Wuruk, Indonesia
Jl. A. Yani, Blimbing, Kota Malang, Jawa Timur 65126, Indonesia
Jl. Sekar Tunjung, Kesiman Kertalangu, Denpasar Tim., Kota Denpasar, Bali 80237, Indonesia
Jl. Basuki Rahmat, Kota SBY, Jawa Timur, Indonesia
Jl. Waribang, Kesiman Petilan, Denpasar Tim., Kota Denpasar, Bali 80237, Indonesia
Jl. Sumatera, Kota Bandung, Jawa Barat, Indonesia
Jl. Kamboja, Kota Denpasar, Bali 80236, Indonesia
Jl. Batukaru, Kabupaten Tabanan, Bali, Indonesia
Jl. By Pass Ngurah Rai, Bali, Indonesia
Jl. Hayam Wuruk, Indonesia
Jl. Hayam Wuruk, Indonesia
Jl. Batur Sari, Sanur, Denpasar Sel., Kota Denpasar, Bali 80228, Indonesia
Jl. Hangtuah, Kota Pekanbaru, Riau, Indonesia
Jl. Mataram, Kota Semarang, Jawa Tengah, Indonesia
Jl. Soekarno - Hatta, Kota Bandung, Jawa Barat, Indonesia
Jl. Durian, Kota Pekanbaru, Riau, Indonesia
Jl. Lintas Tim., Pangkalan Kerinci, Kabupaten Pelalawan, Riau, Indonesia
Jl. Jend. Sudirman, Riau, Indonesia
Jl. Jend Subrantas, Dumai Tim., Kota Dumai, Riau 28826, Indonesia
Jl. Lintas Tim., Pangkalan Kerinci, Kabupaten Pelalawan, Riau, Indonesia
Jl. Kuras, Kp. Baru, Senapelan, Kota Pekanbaru, Riau 28291, Indonesia
Jl. Lintas Tim., Kesumbo Ampai, Mandau, Kabupaten Bengkalis, Riau 28983, Indonesia
Jalan Hang Tuah, Taman Skudai Baru, 81300 Skudai, Johor, Malaysia
Jl. Poros, Waseki Indah, Prafi, Kabupaten Manokwari, Papua Bar. 98356, Indonesia
Jl. Sriwijaya Raya, Sanga Desa, Kabupaten Musi Banyuasin, Sumatera Selatan 30759, Indonesia
Jl. Putri Tujuh, Sidomulyo Bar., Tampan, Kota Pekanbaru, Riau 28289, Indonesia

Jl. Lintas Tim., Kabupaten Langkat, Sumatera Utara, Indonesia
 Jl. Lintas Tim., Rengat Bar., Kabupaten Indragiri Hulu, Riau, Indonesia
 Delowe Stanton Acces Rd, East Point, GA 30344, USA
 Jl. Raya Padang-Painan, Sumatera Barat, Indonesia
 Plaza Tol Bukit Jelutong, 40150 Shah Alam, Selangor, Malaysia
 Jl. H. R. Koroh, Nusa Tenggara Tim., Indonesia
 Sendurwafa Toll Plaza, Dongargarh - Nagpur Rd, Maharashtra 441802, India
 Jl. Tol Lkr. Luar Bar., Cengkareng, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11750, Indonesia
 Jl. Alternati JORR Tol Jatiwarna, Jatiwarna, Pondokmelati, Kota Bks, Jawa Barat 17415, Indonesia
 Jl. Alternati JORR Tol Jatiwarna, Jatiwarna, Pondokmelati, Kota Bks, Jawa Barat 17415, Indonesia
 Jl. Alternatif Tol Merak Atas, Gerem, Grogol, Kota Cilegon, Banten 42438, Indonesia
 Jl. Akses Tol, Gandasari, Cikarang Bar., Bekasi, Jawa Barat 17530, Indonesia
 Tol Jagorawi, Kramat Jati, Kramatjati, East Jakarta City, Special Capital Region of Jakarta 13510, Indonesia
 Jl. Tegar Beriman, Tengah, Cibinong, Bogor, Jawa Barat 16914, Indonesia
 Jl. Raya Cigudeg, Bogor, Jawa Barat, Indonesia
 Tol Jagorawi, Kramat Jati, Kramatjati, East Jakarta City, Special Capital Region of Jakarta 13510, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Cikarang Bar., Bekasi, Jawa Barat 17530, Indonesia
 Jl. Tol Jkt - Cikampek, Tambun Sel., Bekasi, Jawa Barat 17510, Indonesia
 Jl. Tol. Cikarang Bar., Gandasari, Cikarang Bar., Bekasi, Jawa Barat 17530, Indonesia
 Jl. Tol Jkt - Cikampek, Cikarang Bar., Bekasi, Jawa Barat 17530, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Bekasi Tim., Kota Bks, Jawa Barat 17114, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Bekasi Tim., Kota Bks, Jawa Barat 17114, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol Jkt - Cikampek, Bekasi, Jawa Barat 17530, Indonesia
 Dekat Jl. Tol Gempol, Kb. Waris, Pandaan, Pasuruan, Jawa Timur 67156, Indonesia
 Jl. Tol Jkt - Cikampek, Tambun Sel., Bekasi, Jawa Barat 17510, Indonesia
 Jl. Tol Jkt - Cikampek, Tambun Sel., Bekasi, Jawa Barat 17510, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 Jl. Tol. Cikarang Bar., Gandasari, Cikarang Bar., Bekasi, Jawa Barat 17530, Indonesia
 Jl. Tol Jkt - Cikampek, Tambun Sel., Bekasi, Jawa Barat 17510, Indonesia
 NHAI Toll Plaza, Rajchandrapur, Bally, Howrah, West Bengal 711205, India
 Jl. Tol Jkt - Cikampek, Tambun Sel., Bekasi, Jawa Barat 17510, Indonesia
 Jl. Tol Jkt - Cikampek, Indonesia
 NHAI Toll Plaza, Rajchandrapur, Bally, Howrah, West Bengal 711205, India

Gambar 5.4. Hasil ekstraksi data lokasi dari *tweets* yang telah diklasifikasi