

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Twitter

Twitter adalah salah satu dari media sosial yang saat ini tengah banyak digunakan oleh masyarakat. Twitter merupakan layanan media sosial yang mendunia yang tentu penggunaannya tidak hanya pada satu negara saja tapi melainkan dari berbagai negara. Menurut (Hanifan, 2013) Twitter memiliki kemampuan untuk pengguna dapat mengirim serta membaca pesan berbentuk teks hingga mencapai 140 karakter. Selain itu, *tweet* dapat dilihat oleh publik serta pengguna dapat membatasi pesan *tweet* tersebut untuk dapat dilihat oleh *followers*nya. Twitter tersebut menyediakan sebuah aplikasi yaitu *Application Programming Interface* (API), dimana untuk pengguna API tersebut tidak akan berbayar, pengguna hanya cukup melakukan registrasi, dikutip dari (Muis, 2015).

Menurut (Widiastuti, 2014) adapun beberapa fitur yang ada didalam *twitter* yaitu seperti berikut :

1. *Following* dan *Followers*

Following dan *Followers* merupakan salah satu fitur yang terdapat didalam *twitter*. *Following* (mengikuti) merupakan akun yang diikuti akun yang lain. misalnya, akun kita yang mengikuti seluruh kegiatan, aktivitas dan postingan dari akun lainnya. Sedangkan *Followers* (Pengikut) merupakan akun yang mengikuti aku lain. Misalnya, akun orang lain yang mengikuti seluruh aktivitas dan postingan dari akun kita. Dengan kata lain, bagi pengguna *twitter* yang telah mem *Following* (mengikuti) pengguna *twitter* lainnya, akan secara otomatis menerima *update* dari pengguna *twitter* tersebut.

2. *Direct Message*

Direct Message merupakan salah satu fitur yang terdapat didalam *twitter*. Fitur *Direct Message* adalah sebuah fitur yang dapat mengirim pesan pribadi kepada pengguna *twitter* lainnya yang telah mengikuti pengguna

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tersebut. Tetapi, jika pengguna lain tidak mengikuti pengguna tersebut, mereka tetap bisa mengirimkan pesan secara pribadi kepada pengguna tersebut. Dengan kata lain, pengguna twitter yang tidak saling mengikuti dapat saling mengirimkan pesan secara pribadi.

3. *Twitter Search*

Twitter Search adalah salah satu fitur yang terdapat didalam twitter. Fitur *twitter Search* ini memiliki fungsi yaitu untuk dapat memberikan kemudahan kepada para pengguna twitter dalam mencari pengguna twitter lainnya, kata kunci, subjek maupun tempat-tempat.

4. *Trending Topics*

Trending Topics merupakan salah satu fitur lainnya yang terdapat didalam twitter. *Trending Topics* adalah sebuah topik yang paling sering disebut atau dibicarakan didalam Twitter pada saat waktu tertentu.

5. *Latest News* (berita terbaru)

Menurut (Widiastuti, 2014) twitter memungkinkan para penggunanya dengan cepat melihat berita terbaru saat ini. Dengan kata lain, jika salah satu pengguna twitter mendapatkan berita terbaru, maka pengguna tersebut dapat mempostingkan informasi atau berita tersebut di Twitter, maka dalam beberapa detik secara otomatis informasi atau berita tersebut akan muncul di internet.

2.2 Bahasa

Menurut (Watie, 2011) Bahasa merupakan sebuah ungkapan yang mempunyai maksud dengan tujuan untuk dapat menyampaikan sesuatu hal kepada orang lain. Dengan kata lain, antara pembicara dan pendengar dapat saling berkomunikasi satu sama lain, yaitu pembicara dapat dipahami oleh pendengar melalui bahasa yang diungkapkan oleh pembicara. Menurut (Setyonegoro, 2015), bahasa merupakan rangkaian bunyi yang membentuk arti tertentu. Selain itu, menurut (Setyonegoro, 2015) fungsi bahasa yaitu sebagai cara untuk dapat menyampaikan informasi-informasi. Dengan bahasa, manusia dapat menyampaikan perasaan dan pikiran kepada orang lain (Waridah, 2015).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3 Bahasa Melayu Indonesia

Menurut (Nugroho, 2015) bahasa Indonesia berasal dari bahasa Melayu yang masuk kedalam rumpun bahasa *Austronesia* yang telah digunakan di Nusantara pada abad-abad awal penanggalan modern. Bahasa Indonesia juga berfungsi sebagai bahasa Negara (Nugroho, 2015). Bahasa Indonesia merupakan bahasa yang digunakan sehari-hari yang mudah dimengerti. Dengan berkembangnya zaman, bahasa Indonesia perlahan mengalami perkembangan, sehingga bahasa Indonesia telah banyak berubah menjadi bahasa modern, yang banyak akan kosakatanya. Adapun menurut (Nugroho, 2015) fungsi bahasa Indonesia yaitu, sebagai berikut :

1. Bahasa resmi negara
2. Lambang identitas nasional
3. Lambang kebanggaan bangsa
4. Alat perhubungan antar warga, antar daerah, dan antar budaya
5. Antar pemersatu suku budaya dan bahasanya

2.4 Bahasa Melayu Malaysia

Bahasa Malaysia berasal dari bahasa Melayu yang masuk ke dalam rumpun *Austronesia* yang sama seperti bahasa Indonesia. Menurut (Takari, 2015) Jika di negara Indonesia, bahasa melayu disebut dengan bahasa Indonesia, maka di Malaysia, bahasa Melayu disebut dengan bahasa Melayu atau bahasa Malaysia. Menurut (Takari, 2015) bahasa Melayu Malaysia merupakan bahasa kebangsaan dan telah menjadi bahasa resmi di negara Malaysia. Di Negara Malaysia, bahasa melayu telah mengalami perubahan beberapa kali. Pada tahun 1970, bahasa Melayu dinamakan dengan bahasa Malaysia atas sebab politik.

2.5 *Text Mining*

Menurut (Aisyah & Fithriasari, 2016) *Text mining* adalah salah satu bagian ilmu data mining yang bertugas untuk menganalisis data seperti dokumen teks.. Menurut (Aisyah & Fithriasari, 2016) *text mining* merupakan salah satu cara dalam menganalisis teks secara otomatis yang dilakukan oleh komputer, dengan tujuan untuk dapat menggali informasi yang dibutuhkan dari sekumpulan teks.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Menurut (Susilowati, Sabariah, & Gozali, 2015) *text mining* merupakan penambangan teks yang dilakukan oleh komputer dengan tujuan untuk mendapatkan sesuatu yang baru dalam bentuk informasi. Sesuatu tersebut sebelumnya tidak diketahui, yang kemudian diekstrak dari sumber-sumber data teks yang berbeda-beda. Penambangan data berupa teks dimana sumber datanya berasal dari dokumen.

Menurut (Aisyah & Fithriasari, 2016) *text mining* saat ini telah banyak dikembangkan dalam berbagai bidang seperti, pemasaran, keamanan, aplikasi dan perangkat lunak, media *online*, biomedis dan akademik. Proses pada *text mining* ada beberapa kategori teks yaitu antara lain teks *clustering*, ekstraksi konsep, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas.

2.6 Klasifikasi Teks

Klasifikasi merupakan suatu proses untuk dapat menemukan fungsi yang menjelaskan serta membedakan kelas data. Klasifikasi bertujuan untuk menemukan serta memperkirakan kelas mana yang tidak diketahui dari suatu objek tersebut. Pada tahap pengklasifikasian data terdapat dua proses yang harus dilakukan yaitu :

1. Proses *training*

Proses *training* ini dilakukan untuk melatih set yang sebelumnya telah diketahui label-labelnya untuk dapat membangun sebuah model.

2. Proses *testing*

Proses *testing* ini bertujuan untuk dapat mengetahui bagaimana keakuratan model yang telah dibangun sebelumnya pada proses *training* tadi.

Menurut (Februariyanti, 2012) klasifikasi merupakan salah satu metode data mining dalam mengelompokkan kelas dari suatu objek yang sebelumnya belum diketahui. Dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan kata lain, orang-orang yang melakukan pencarian informasi dapat mudah melewatkan kategori yang tidak relevan dengan informasi yang dicari atau yang tidak menarik perhatian.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.7 Text Preprocessing

Text preprocessing merupakan tahapan penting dalam melakukan proses klasifikasi teks. Dalam melakukan *text mining*, sebelumnya dokument tersebut harus dipersiap dahulu baru dapat digunakan dalam proses utama klasifikasi teks. *Text preprocessing* bertujuan untuk mengubah data teks yang sebelumnya tidak terstruktur menjadi data yang terstruktur. *Text preprocessing* bertujuan untuk dapat menghilangkan *noise*, serta dapat menyeragamkan bentuk kata dan dapat mengurangi volume kata.

Menurut (Leman & Andesa, 2015) ada beberapa tahapan dalam *Text preprocessing*, yaitu sebagai berikut :

1. Case folding

Case folding merupakan suatu tahapan dalam menyeragamkan bentuk huruf, dari huruf besar menjadi huruf kecil, selain itu pada tahapan ini dilakukan penghilangan tanda baca dan unsur angka. *Case folding* dilakukan dengan tujuan mempermudah pencarian dokumen (Leman & Andesa, 2015)

2. Cleaning

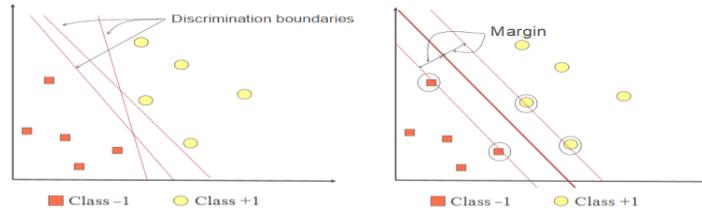
Cleaning merupakan suatu tahap pembersihan *text* dari kata-kata yang tidak dibutuhkan dengan tujuan yaitu untuk mengoptimalkan serta mengurangi terjadinya *noise* pada tahapan klasifikasi, seperti *hashtag* (#), *mention* (@), *emoticon*, *retweet* (RT), dan link url.

3. Tokenizing

Tokenizing merupakan sebuah proses untuk memecah teks menjadi sebuah kata tunggal. Kata tunggal tersebut terdiri dari satu kata atau lebih dari satu kata. Menurut Raymond J. Mooney yang dikutip oleh (Muis, 2015) semua tergantung aturan dan juga metode yang digunakan dalam proses memecah teks tersebut.

4. Filtering

Filtering merupakan tahapan dimana pengambilan kata-kata penting saja dan menghilangkan kata-kata yang tidak berpengaruh banyak didalam isi dokumen, seperti kata-kata, “dan, adalah, di, yang”.



Gambar 2.1 Konsep *hyperlane* dari SVM (Nugroho, Witarto, Handoko, 2003)

Menurut (Aisyah & Fithriasari, 2016) Pada gambar diatas, terdapat suatu bidang pembatas yang disebut *support vector*. bidang pembatas merupakan suatu pembatas yang sejajar bertujuan untuk memisahkan dua kelas. Pada bidang pembatas pertama bertujuan untuk memisahkan atau membatasi kelas pertama, dan pada bidang pembatas kedua bertujuan untuk memisahkan atau membatasi kelas kedua. Berdasarkan penjelasan diatas diperoleh :

$$x_i w + b \geq 1, y_i = +1 \tag{2.3}$$

$$x_i w + b \leq -1, y_i = -1$$

Pada rumus diatas, diperoleh bahwa :

w = normal bidang

b = posisi bidang terhadap pusat koordinat

Menurut (Aisyah & Fithriasari, 2016) untuk dapat menghitung nilai margin antara bidang pembatasnya didasarkan dari rumus jarak garis ke titik pusatnya, yaitu :

$$\frac{1-b-(-1-b)}{\|w\|} = \frac{2}{\|w\|} \tag{2.4}$$

Pada rumus diatas, didapat bahwa :

w = normal bidang

b = posisi bidang terhadap pusat koordinat

Adapun menurut (Nugroho, Witarto, & Handoko, 2003) beberapa karakteristik dari *Support Vector Machine* (SVM) yaitu sebagai berikut :

1. Prinsip kerja dari *Support Vector Machine* (SVM) adalah pada dasarnya hanya mampu menangani klasifikasi dua kelas saja, namun telah dikembangkan untuk klasifikasi lebih dari dua kelas dengan adanya *pattern recognition*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. *Pattern recognition* dilakukan dengan mengtransformasikan data pada ruang input kedalam ruang yang berdimensi lebih tinggi dan untuk optimisasi yang dilakukan pada ruang vector yang baru tersebut.

3. Dilihat dari prinsipnya metode *Support Vector Machine* (SVM) merupakan linear *Classifier*.

Menurut sumber (Nugroho, Witarto, & Handoko, 2003) yang mengangkat judul tentang *Support Vector Machine* Teori dan Aplikasinya dalam Bioinformatika, bahwa walaupun dalam hal ini metode *Support Vector Machine* (SVM) merupakan metode yang lebih baik dibandingkan dengan metode lain, tetapi metode *Support Vector Machine* (SVM) juga memiliki berbagai kelebihan serta kelemahan.

Adapun kelebihan dari metode *Support Vector Machine* (SVM) yaitu sebagai berikut :

1. Generalisasi

Generalisasi merupakan suatu kemampuan metode dalam mengklasifikasi suatu pola, yang tidak merupakan data yang digunakan dalam tahap pembelajaran dalam metode tersebut. Jika generalisasi *error*, maka hal tersebut dipengaruhi oleh faktor *error* pada training set. Pada umumnya, *learning machine* berfokus pada usaha untuk dapat meminimalkan *error* yang terdapat pada training set yang disebut dengan *Empirical Risk Minimization* (ERM).

2. *Curse of dimensionality*

Curse of dimensionality merupakan suatu masalah yang dihadapi suatu metode dalam mengestimasi parameter karena jumlah sampel datanya yang lebih sedikit daripada dimensional ruang vektornya. Dengan kata lain, metode *Support Vector Machine* (SVM) ini memungkinkan untuk mengestimasi parameter yang jumlah sampel datanya lebih sedikit daripada dimensional pada ruang vektor tersebut. Oleh karena itu, ini membuktikan *Support Vector Machine* (SVM) adalah metode yang tepat digunakan dalam memecahkan masalah yaitu terbatasnya sampel data yang diperoleh pada dimensi yang tinggi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.

Landasan teori

Landasan teori artinya yaitu metode yang berbasiskan statistik. Dengan kata lain, metode *Support Vector Machine* (SVM) ini mempunyai landasan teori yang dapat dianalisa dengan sangat jelas.

4.

Feasibility (kemungkinan)

Menurut (Nugroho, Witarto, & Handoko, 2003) metode *Support Vector Machine* (SVM) memungkinkan pengimplementasian menjadi relatif mudah karena pada proses penentuannya, *Support Vector Machine* (SVM) dapat dirumuskan dalam *Quadratic Progaming (QP) problem*.

Menurut (Nugroho, Witarto, & Handoko, 2003)Adapun kelemahan pada metode *Support Vector Machine* (SVM) ini yaitu sebagai berikut :

1. Pada umumnya metode *Support Vector Machine* (SVM) digunakan untuk permasalahan pada pengklasifikasian dua kelas. Tetapi metode *Support Vector Machine* (SVM) telah dikembangkan agar bisa menyelesaikan permasalahan pada pengklasifikasian yang lebih dari dua kelas.
2. Tidak dapat digunakan dalam permasalahan berskala besar. Dalam hal ini, berskala besar artinya jumlah sampel datanya.

2.10 Confusion Matrix

Menurut (Indriani, 2014) *Confusion Matrix* merupakan suatu tabel yang berfungsi menyatakan jumlah data yang benar diklasifikasikan dan jumlah data yang salah diklasifikasikan. Adapun rumus *Confusion Matrix* yaitu, sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah seluruh prediksi}} \times 100 \% \quad (2.5)$$

2.11 Penelitian Terkait

Berikut Tabel 2.1 adalah beberapa penelitian terkait sistem yang akan dibangun oleh penulis.

Tabel 2.1 Penelitian Terkait

No.	Penulis	Judul	Tahun	Metode	Jumlah Data	Tingkat Akurasi
1.	Imelda A.Muis	Penerapan Metode <i>Support Vector Machine</i> (SVM) menggunakan <i>Kernel Radial Basis Function</i> (RBF) pada Klasifikasi Tweet	2015	<i>Support Vector Machine</i> (SVM)	1020	99,12%
2.	Defri Kurniawan, dan Catur Supriyanto	Optimasi algoritma <i>Support Vector Machine</i> (SVM) menggunakan Adaboost untuk penilaian risiko kredit	2013	<i>Support Vector Machine</i> (SVM)	690	86,09%
3.	Moh. Yamin Darsyah	Klasifikasi Tuberkulosis dengan Pendekatan Metode <i>Support Vector Machine</i> (SVM)	2014	<i>Support Vector Machine</i> (SVM)	700	98%
4.	Puspita Anna Octaviani, Yuciana	Penerapan Metode Klasifikasi <i>Support Vector Machine</i> (SVM)	2014	<i>Support Vector Machine</i> (SVM)	419	93,902%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	Wilandari, Dwi Ispriyanti	pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang				
5.	Lestari Handayani, Fitriandini	Prediksi kebangkrutan perusahaan menggunakan <i>Support Vector Machine (SVM)</i>	2010	<i>Support Vector Machine (SVM)</i>	50	90,78%
6.	Siti Nurhayati, Kusrini, Emha Taufiq Luthfi	Prediksi Mahasiswa Drop Out Menggunakan Metode <i>Support Vector Machine (SVM)</i>	2015	<i>Support Vector Machine (SVM)</i>	310	98,06%
7.	Endina Putri, Diah Puspitaningr um, Andre Mirfen	Identifikasi Tanda Tangan dengan Pendekatan <i>Support Vector Machine (SVM)</i>	2015	<i>Support Vector Machine (SVM)</i>	800	97,33%
8.	Neneng, Kusworo Adi, R. Rizal Isnanto	<i>Support Vector Machine</i> untuk Klasifikasi Citra Jenis Daging berdasarkan Tekstur Menggunakan Ekstraksi ciri <i>Gray</i>	2016	<i>Support Vector Machine (SVM)</i>	600	87,5%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		<i>Level Co-Occurrence Matrices (GLCM)</i>				
9.	Enda Esyudha Pratama, Bambang Riyanto Trilaksono	Klasifikasi Keluhan Pelanggan berdasarkan Tweet dengan menggunakan Metode <i>Support Vector Machine</i> (SVM)	2014	<i>Support Vector Machine</i> (SVM)	600	89,76%
10.	Elly Susilowati, Mira Kania Sabariah, Alfian Akbar Gozali	Implementasi Metode <i>Support Vector Machine</i> untuk Melakukan Klasifikasi Kemacetan Lalu Lintas pada Twitter	2013	<i>Support Vector Machine</i> (SVM)	934	90%