

# PENERAPAN TEXT-MINING UNTUK MENGIDENTIFIKASI PENGGUNA TWITTER TERHADAP FENOMENA PERAN DPR RI

Herianto<sup>1</sup>

<sup>1</sup>Dosen Prodi Teknik Informatika Universitas Darma Persada

## Abstrak

Media sosial adalah media yang digunakan untuk berbagi teks, gambar, suara, dan termasuk yang paling populer di Indonesia adalah Twitter. Twitter digunakan penggunanya untuk menyampaikan pendapatnya ketika terdapat sebuah kejadian populer. Contoh peristiwa populer adalah fenomena politik dan tingkah laku pelakunya. Salah satu pelaku politik yang sering dibicarakan di twitter adalah anggota DPR RI yang sering menjadi berita trending topic. Penelitian ini bertujuan untuk menganalisa berbagai komentar di twitter yang berkaitan dengan anggota DPR terutama masalah keterpilihan anggota DPR baru periode 2019-2024. Pengumpulan data dilakukan berdasarkan kata kunci dan hastag dengan jumlah tertentu, setelah data terkumpul diolah menggunakan metoda text mining dan hasil olahannya diproses dengan algoritma Support Vector Machine (SVM) untuk mendapat klasifikasinya. Penelitian ini menghasilkan sebuah sistem otomatis yang dapat mengklasifikasikan (memilih) mana yang komentar negatif dan mana yang positif.

**Kata kunci:** Text-ming, twitter, DPR RI, SVM

## 1. PENDAHULUAN

### 1.1. Latar Belakang

Penggunaan internet di Indonesia saat ini sudah dapat dikatakan berkembang pesat. Adanya internet dapat dimanfaatkan sebagai sarana untuk menyalurkan hobi baik seperti jual beli secara online, menulis atau menyampaikan pendapat terhadap sesuatu melalui media sosial. Media sosial adalah media yang digunakan oleh konsumen untuk berbagi teks, gambar, suara, dan video informasi baik dengan orang lain maupun perusahaan dan vice versa (Philip Kotler dan Kevin Lane Keller 2016). Media sosial yang cukup populer di Indonesia salah satunya yaitu Twitter. Melalui Twitter pengguna menyampaikan pendapatnya secara bebas. Twitter juga menyediakan fitur trending topik untuk kawasan atau wilayah tertentu sesuai dengan preferensi pengguna masing-masing. Twitter berkembang dengan cepat dalam segi pengguna. Pada tahun 2013, terdapat lebih dari 500 juta pengguna terdaftar dan 200 juta diantaranya merupakan penggunaan aktif (Twitter, 2017).

Perkembangan yang sangat pesat pada Twitter ini sangat menarik untuk dianalisa cuitan (tweet) nya tentang berbagai opini dan fenomena. Penelitian bermaksud menganalisa kecendrungan pembicaraan tentang satu topik dalam hal ini adalah keanggotaan DPR RI.

### 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas maka ditemukan rumusan masalah pada penelitian ini adalah sebagai berikut:

- 1 Bagaimana merancang sebuah sistem Identifikasi berdasarkan Analisis Sentiment pada Twitter?
- 2 Bagaimana penerapan metode Support Vector Mechine (SVM) untuk mengetahui yang pro dan kontra terhadap anggota DPR RI yang baru ?

### 1.3. Batasan Masalah

Berdasarkan rumusan masalah yang telah ditetapkan, maka batasan masalah penelitian ini adalah sebagai berikut:

1. Media social yang digunakan adalah Twitter
2. Pengambilan data dilakukan berdasarkan kata kunci berupa komentar dan hastag dengan jumlah tertentu dan pengambilan sampel di tanggal tertentu.

### 1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengklasifikasi data pro dan kontra terhadap anggota DPR berdasarkan data dari twitter
2. menerapkan metode *Support Vector Mechine* (SVM ) untuk analisa sentimen pada sistem yang dibangun.

### 1.5. Manfaat Penelitian

Manfaat dari penelitian diatas adalah

1. Penelitian ini diharapkan agar anggota DPR RI baru mengetahui tanggapan awal masyarakat terhadap mereka apakah optimis (pro) atau pesimis (kontra).
2. Hasil penelitian ini diharapkan dapat menghasilkan sistem yang dapat melakukan Analisa sentiment/komentar masyarakat di social media twitter.
3. Diharapkan hasil penulisan Laporan Tugas Akhir ini dapat menjadi tambahan referensi untuk penulisan dan penelitian selanjutnya.

## 2. TINJAUAN PUSTAKA

### 2.1. Tinjauan Terhadap Penelitian Terkait

Berikut beberapa penelitian terkait yang menjadi acuan atau referensi pada penelitian ini:

Adhi Viky Sudiantoro, Eri Zuliarso (2018) melakukan penelitian dengan judul "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Support Vector Machine". Analisis sentimen atau bisa disebut juga opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif. Penelitian ini bertujuan untuk mengklasifikasi data tweet menjadi 2 sentimen yaitu positif dan negatif. Dalam penelitian ini menggunakan teks Bahasa Indonesia yang terdapat di sosial media twitter berupa tweet. Opini masyarakat yang terdapat di tweet tersebut dapat dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui masyarakat terhadap pilkada jawa barat apakah positif atau negatif. Data yang digunakan terdiri 300 data tweet yang dibagi menjadi 2 yaitu untuk data latih sebanyak 200 data dan 100 data untuk data uji. Pengklasifikasian data tweet menggunakan text mining dengan Naïve Bayes Classifier. Sebelum klasifikasi, dilakukan beberapa tahap pemrosesan teks seperti case folding, normalisasi, tokenisasi dan stopwords removal. Hasil dari 100 data uji yang klasifikasi menghasilkan 32 data bersentimen positif dan sebanyak 68 data sentimen negative. Dapat diartikan bahwa 100 data uji yang

diklasifikasi masuk dalam kategori bersentimen negatif dikarenakan data positif lebih kecil daripada data yang bersentimen negatif. Adapun nilai akurasi algoritma Naïve Bayes Classifier memberikan nilai akurasi sebesar 84%.

Berdasarkan penelitian diatas maka penulis tertarik membuat penelitian penerapan analisa text-mining terhadap komentar di twitter tentang fenomena anggota DPR RI Baru 2019 menggunakan Metode Support Vector Machine.

## **2.2. Twitter**

Twitter adalah sebuah media sosial dan layanan microblogging yang mengijinkan penggunaanya untuk mengirimkan pesan realtime. Pesan ini populer dengan sebutan tweet. Tweet adalah sebuah pesan pendek dengan panjang karakter yang dibatasi hanya sampai 140 karakter. Dikarenakan keterbatasan karakter yang bisa dituliskan, sebuah tweet seringkali mengandung singkatan, bahasa slang maupun kesalahan pengejaan (Agarwal et al., 2014). Sejak awal, Twitter memang dibuat sebagai layanan berbasis mobile yang didesain sesuai dengan batasan karakter pada sebuah pesan teks (SMS), dan sampai hari ini, Twitter masih bisa digunakan pada sembarang telepon genggam yang memiliki kemampuan untuk mengirim dan menerima pesan teks (The Twitter Government and Election Team, 2014).

Twitter diciptakan untuk menjadi tempat saling berbagi pengalaman antar sesama penggunaanya tanpa adanya sekat penghalang. Dengan menggunakannya, pengguna akan mudah untuk mengikuti tren, cerita, informasi dan berita dari seluruh penjuru dunia. Selain itu, Twitter juga membantu penggunaanya untuk selalu terhubung dengan orang-orang terdekatnya. Ketika penggunaanya mengirimkan tweet, pesan tersebut bersifat publik dan bisa diakses oleh siapapun, dimanapun dan kapanpun. Bahkan, bagi orang-orang yang mengikuti (follow) akun Twitter tersebut, tweet tersebut akan secara otomatis muncul di lini masa orang tersebut.

## **2.3. Klasifikasi Teks dan Text-mining**

Klasifikasi teks merupakan proses menemukan pola baru yang belum terungkap sebelumnya. Klasifikasi teks dilakukan dengan memproses dan menganalisa data dalam jumlah besar. Dalam prosesnya, klasifikasi teks melibatkan struktur yang mungkin terdapat pada teks dan mengekstraks informasi yang relevan pada teks. Dalam menganalisis sebagian atau keseluruhan teks yang tidak terstruktur, klasifikasi teks mencoba meng-asosiasikan sebagian atau keseluruhan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu (Miller, 2015). Tantangan dari klasifikasi teks adalah sifat data yang tidak terstruktur dan sulit untuk menangani, sehingga diperlukan proses text mining. Diharapkan melalui proses text mining, informasi yang ada dapat dikeluarkan secara jelas di dalam teks tersebut dan dapat dipergunakan dalam proses analisis menggunakan alat bantu komputer (Witten dkk, 2016).

## **2.4. Web Scrapping**

WebScraping Turland (2014) adalah proses pengambilan sebuah dokumensemi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain. Web scraping memiliki sejumlah langkah, sebagai berikut:



Gambar 2.1 Langkah Web Scraping

## 2.5. Support Vector Machine (SVM)

Menurut Santosa (2015), Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Support Vector Machine masuk kelas supervised learning, dimana dalam implementasinya perlu adanya tahap pelatihan menggunakan sequential training SVM dan disusul tahap pengujian. Konsep klasifikasi dengan Support Vector Machine adalah mencari hyperplane terbaik yang berfungsi sebagai pemisah dua kelas data. Support Vector Machine mampu bekerja pada dataset yang berdimensi tinggi dengan menggunakan kernel trik. Support Vector Machine hanya menggunakan beberapa titik data terpilih yang berkontribusi (support vector) untuk membentuk model yang akan digunakan dalam proses klasifikasi.

Persamaan Support Vector Machine

$$f(x) = w \cdot x + b$$

Keterangan:

- w: parameter hyperplane yang dicari (garis yang tegak lurus antara garis hyperplane dan titik support vector)
- x: titik data masukan Support Vector Machine
- b: parameter hyperplane yang dicari (nilai bias)

Untuk penelitian ini menggunakan kernel polynomial. Persamaannya:

$$K(x,y) = (x \cdot y + c)^d$$

Keterangan:

- (x,y): Nilai kernel dari data x dan data y
- x: fitur data 1
- y: fitur data 2
- c: nilai konstanta
- d: nilai degree

## 2.6. Unified Modeling Language (UML)

*Unified Modeling Language* (UML) merupakan sistem arsitektur yang bekerja dalam OOAD (*Object-Oriented Analysis/Design*) dengan satu bahasa yang konsisten untuk menentukan, visualisasi, mengkonstruksi, dan mendokumentasikan artifact (sepotong informasi yang digunakan atau dihasilkan dalam suatu proses rekayasa software, dapat berupa model, deskripsi, atau software) yang terdapat dalam sistem software (Martin, 2015: 63). UML merupakan bahasa pemodelan yang paling sukses dari tiga metode OO yang telah ada sebelumnya, yaitu Booch, OMT (*Object Modeling Technique*), dan OOSE (*Object-Oriented Software Engineering*) (Munawar, 2015:37). UML merupakan kesatuan dari ketiga pemodelan tersebut dan ditambah kemampuan lebih karena mengandung metode tambahan untuk mengatasi masalah pemodelan yang tidak dapat ditangani ketiga metode tersebut. UML dikeluarkan oleh OMG (*Object*

*Management Group, Inc*) yaitu organisasi internasional yang dibentuk pada 1989, terdiri dari perusahaan sistem informasi, software developer, dan para user sistem komputer.

Dengan adanya UML, diharapkan dapat mengurangi kekacauan dalam bahasa pemodelan yang selama ini terjadi dalam lingkungan industri. UML diharapkan juga dapat menjawab masalah penotasian dan mekanisme tukar menukar model yang terjadi selama ini. Nugroho (2015 : 87). Tujuan UML diantaranya adalah:

1. Memberikan model yang siap pakai, bahasa pemodelan visual yang ekspresif untuk mengembangkan dan saling menukar model dengan mudah dan dimengerti secara umum.

2. Memberikan bahasa pemodelan yang bebas dari berbagai Bahasa pemrograman dan proses rekayasa.

3. Menyatukan praktek-praktek terbaik yang terdapat dalam pemodelan.

## 2.7. PHP

Menurut Kadir (2014) "PHP adalah pemrograman interpreter yaitu proses penerjemahan baris kode sumber menjadi kode mesin yang dimengerti komputer secara langsung pada saat baris kode dijalankan". PHP disebut juga pemrograman Server Side Programming, hal ini dikarenakan seluruh prosesnya dijalankan pada server. PHP adalah suatu bahasa dengan hak cipta terbuka atau yang juga dikenal dengan open source yaitu pengguna data mengembangkan kode-kode fungsi sesuai kebutuhannya.

Seluruh aplikasi berbasis web dapat dibuat dengan PHP. Namun kekuatan yang paling utama PHP adalah pada konektivitasnya dengan system database di dalam web. Sistem database yang dapat didukung oleh PHP adalah Oracle MySQL, Ms. Access, Sybase PostgreSQL, dan lainnya. PHP dapat berjalan di berbagai system operasi seperti Windows 98/NT, UNIX/LINUX, Solaris maupun Macintosh (Rudianto, 2012: 44).

## 2.8. Hyper Text Markup Language (HTML)

*Hypertext Markup Language* (HTML) adalah bahasa yang digunakan untuk menulis halaman web (Rudianto, 2012:10). HTML merupakan pengembangan dari standar pemformatan dokumen teks yaitu Standard Generalized Markup Language (SGML) (Kadir, 2012:58). HTML sebenarnya adalah dokumen ASCII atau teks biasa, yang dirancang untuk tidak tergantung pada suatu sistem operasi tertentu. Mendesain HTML dapat dilakukan dengan cara menggunakan HTML Editor, seperti Microsoft FrontPage, Adobe Dreamweaver, dan lain-lain.

Dalam pemrograman web, kode HTML wajib digunakan untuk melakukan desain tampilan sistem yang dibuat. Dengan menggunakan HTML Editor moderen, programmer tidak perlu lagi melakukan pengetikan kode HTML, melainkan langsung melakukan desain tampilan sistem secara drag and drop.

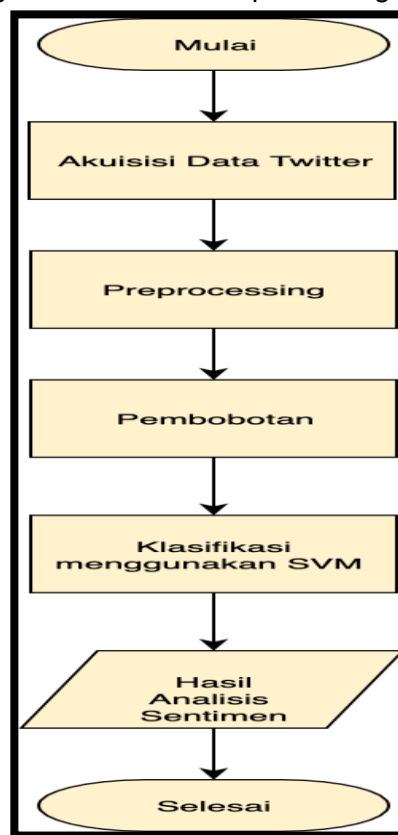
## 2.9. Mysql

Sistem manajemen database bersifat open source yang digunakan untuk menyimpan data. Menurut Anhar (2010: 21) MySQL (My Structure Query Language) adalah sebuah perangkat lunak sistem manajemen basis data SQL (Database Management System) atau DBMS dari sekian banyak DBMS, seperti Oracle, MS SQL, Postagre SQL, dan lain-lain. Menurut kadir (2008 : 2) MySQL merupakan software yang tergolong sebagai DBMS (Database Management System) yang bersifat open source, sedangkan Menurut Arief (2011 : 151) MySQL adalah salah satu jenis database server yang sangat terkenal dan banyak digunakan untuk membangun aplikasi web yang menggunakan database sebagai sumber dan pengelolaan datanya.

Berdasarkan pendapat yang dikemukakan di atas dapat disimpulkan bahwa, MySQL adalah sebuah konsep pengoperasian database, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis. Keandalan suatu sistem database (DBMS) dapat diketahui dari cara kerja optimizer-nya dalam melakukan proses perintah-perintah SQL, yang dibuat oleh user maupun program-program aplikasinya. MySQL biasanya digunakan atau diinstall bersamaan dengan XAMPP sehingga untuk melihat isi tabel bisa menggunakan PHPmyAdmin.

### 3. Metodologi Perancangan Sistem

Pada penelitian ini perancangan sistem menggambarkan kerangka atau pola dasar sistem yang akan dibangun. Berikut adalah perancangan sistemnya:



Gambar 3.1 Perancangan Sistem

Pada perancangan sistem di atas dijelaskan proses sentiment analysis pada topik Anggota DPR RI Baru 2019 sebagai berikut:

#### 1. Tahapan Akuisisi Data Twitter

Pada tahapan awal ini data yang di akuisisi atau dikumpulkan dari Twitter yang terhubung langsung melalui API (Application Programming Interface) dan menambahkan proses deteksi bahasa untuk mendapatkan data atau dokumen yang berbahasa Indonesia.

#### 2. Tahapan Preprocessing

Tahapan Preprocessing yaitu tahapan memproses data atau dokumen. Dalam tahapan preprocessing terdapat beberapa bagian yaitu cleaning, case folding,

parsing/tekonizing, filtering kemudian tahapan stemming untuk mendapatkan kata dasar yang akan diklasifikasikan.

**3. Tahapan Pembobotan**

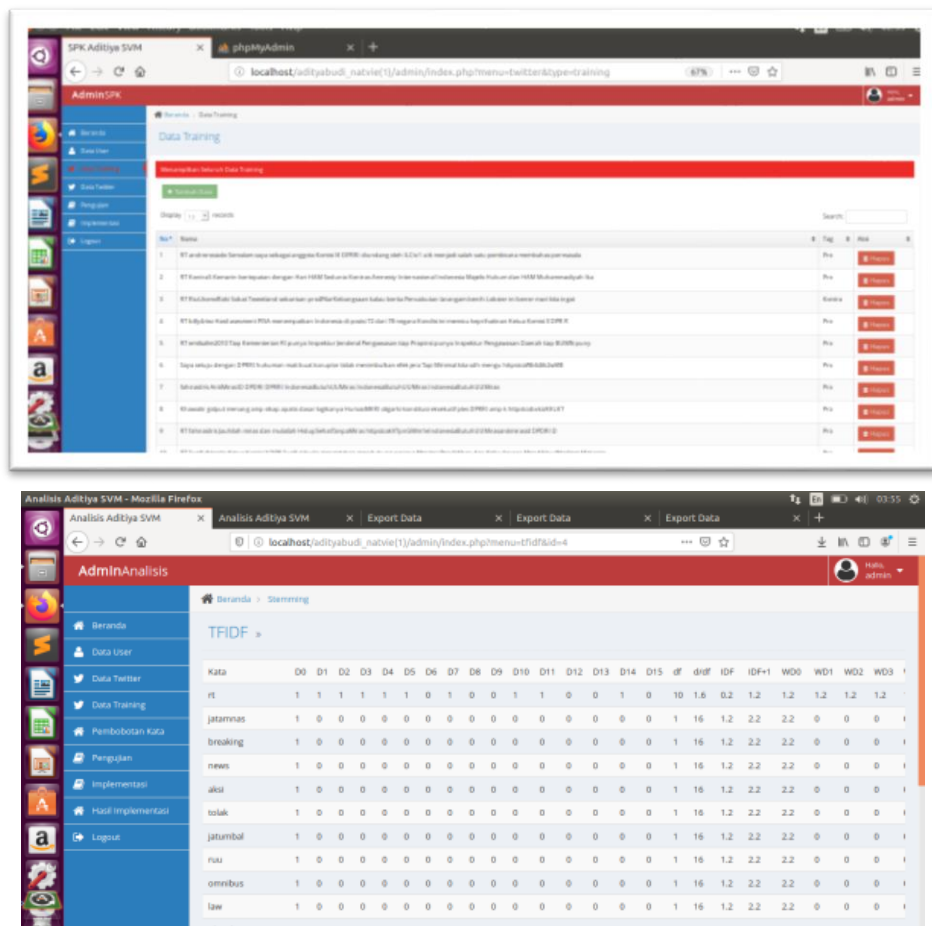
Pada tahapan pembobotan ini menggunakan metode termpresense(TP). term frequency(TF) dan term frequency-inverse document frequency (TF-IDF),

**4. Tahapan Klasifikasi SVM**

Pada tahapan ini merupakan tahapan proses klasifikasi menggunakan SVM.

**4. PEMBAHASAN**

Dalam pengolahan data ini dari mulai input sampai dengan output yang dihasilkan adalah melalui proses pengambilan data di twitter mengenai isu politik anggota DPR RI baru 2019 . Lalu di lakukan teks Processing setelah itu melakukan TF-IDF , *Support Vector Machine* dan menghasilkan sebuah output berupa sentimen pro, kontra .



Gambar. 4.1 Data Training

Pada halaman ini menampilkan data training yang akan di uji pada sistem . Data ini di peroleh melalui menu data twitter dengan mencari keyword dengan menggunakan hastag(#)DPRRI lalu meng *crawl* data tersebut sesuai banyak nya data yang kita cari.



Gambar. 4.2 Hasil Stemming

Pada halaman ini menampilkan hasil dari sebuah komentar yang telah di pilih dari data training . Hasil nya berupa penghilangan kata imbuhan, kata-kata tidsk penting, angka dan emoticon selain huruf abjad, dan menjadikan nya kata dasar.

No	Kata	Dokumen				DF	IDF	TF*IDF			
		D1	D2	D3	D4			D1	D2	D3	D4
1	anggota	1	0	0	1	2	0.301	0.301	0	0	0.301
2	bagai	1	0	0	0	1	0.602	0.602	0	0	0
3	bahas	1	0	0	0	1	0.602	0.602	0	0	0
4	bicara	1	0	0	0	1	0.602	0.602	0	0	0
5	buat	0	0	1	0	1	0.602	0	0	0.602	0
6	DPR	0	1	0	0	1	0.602	0	0.602	0	0
7	DPRRI	1	0	1	2	4	0	0	0	0	0
8	efek	0	0	1	0	1	0.602	0	0	0.602	0
9	hasil	0	1	0	0	1	0.602	0	0.602	0	0
10	hukuman	0	0	1	0	1	0.602	0	0	0.602	0
11	indonesia	0	1	0	0	1	0.602	0	0.602	0	0

Gambar. 4.3 Hasil TF-IDF

Hasil ini menampilkan proses selanjutnya setelah melewati tahap text processing data akan dilakukan pembobotan kata (TF-IDF) setiap kata yang dihitung bobotnya dengan *term* atau kata yang muncul dalam setiap dokumen dari data latih dan data yang akan di analisa.

## 5. PENUTUP

### 5.1. Kesimpulan

Berdasarkan hasil akuisisi dan percobaan data text yang di olah menjadi sentimen analysis menggunakan Support Vector Machine (SVM) dapat di ambil beberapa kesimpulan yaitu :

1. Klasifikasi tweet bersentimen lebih akurat jika data latih yang di gunakan semakin banyak dalam data pengetahuan.
2. Analisis tidak berjalan maksimal terhadap bahasa asing dan bahasa daerah.
3. Semakin banyak data latih dan mempunyai jumlah data latih dengan jumlah kategori yang sama antar kategori maka nilai sama antar kategori maka nilai sentimen yang didapat akan semakin bagus sentiment yang dihasilkan.



4. Dalam penelitian ini hasil analisis sentimen pada twitter adalah opini dari pengguna terhadap Dua Kubu Pro dan Kontra terhadap isu anggota DPR RI. Untuk mengambil data opini tersebut memanfaatkan web crawling untuk mengambil data dari twitter. Metode yang digunakan adalah SVM (Support Vector Machine), hasil analisa berupa diagram chart line untuk mengetahui tinggi akurasi.

## 5.2. Saran

Berdasarkan hasil evaluasi terhadap perangkat lunak yang dikembangkan berikut adalah saran-saran yang dipandang perlu :

1. Melakukan deteksi pada tweet spam sehingga mengurangi dupilkasi tweet.
2. Pemilihan data latih dilakukan lebih selektif sehingga fitur pada masing-masing kelas yang berbeda dapat terekstraksi dengan lebih baik.
3. Menentukan topik pada bagian analisis sehingga prediksi dapat dilakukan dengan lebih tepat.

## DAFTAR PUSTAKA

1. Khan, A. Z. H.,2015. **Sentiment Analysis Using Support Vector Machine**. [Online] Tersedia Di: <[Http://Www.Ijarcse.Com/Docs/Papers/Special\\_Issue/Itsd2015/25.Pdf](http://www.ijarcse.com/docs/papers/special_issue/itsd2015/25.pdf)> [Diakses 2 Februari 2016].
2. Maulana, A. (2016). **Twitter Rahasiakan Jumlah Pengguna di Indonesia**, Website: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-penggunadi-indonesia>, diakses 20 April 2018.
3. Faradhillah, N. Y., Kusumawardani, R. P., & Hafidz, I. 2016. **Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin**. Seminar Nasional Sistem Informasi Indonesia, 1(3), pp.16-24.
4. Yamatoto, Y., & Kumatato, T. 2015. **Multidimensional sentiment calculation method for Twitter based on emoticons**. International Journal of Pervasive Computing and Communications , 11 (2), 212-232.
5. Tiara, Sabariah, M.K. and Effendy, V., 2015. **Sentiment analysis on Twitter using the combination of lexicon-based and support vector machine for assessing the performance of a television program**. 2015 3rd International Conference on Information and Communication Technology, IColCT 2015, pp.386–390.
6. Windasari, I.P., Uzzi, F.N. and Satoto, K.I.,2017. **Sentiment Analysis on Twitter Posts : An analysis of Positive or Negative Opinion on GoJek**. pp.266–269.
7. Desai, M. and Mehta, M.A., 2017. **Techniques for sentiment analysis of Twitter data: A comprehensive survey. Proceeding- IEEE International Conference on Computing, Communication and Automation, ICCCA 2016, (March), pp.149–154.**
8. Rofiqoh, U., Perdana, R.S. and Fauzi, M.A., 2017. **Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature**. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.