**DEPARTEMEN INFORMATIKA**

**FAKULTAS TEKNOLOGI**

**INFORMASI DAN KOMUNIKASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# IDENTITAS PENGUSUL

**NAMA : Rakhma Rufaida Hanum**

**NRP : 05111540000161**

**DOSEN WALI : Prof. Ir. Supeno Djanali, M.Sc., Ph.D.**

**DOSEN PEMBIMBING : 1. Abdul Munif, S.Kom., M.Sc.Eng.  
 2. Nurul Fajrin Ariyani, S.Kom., M.Sc.**

# JUDUL TUGAS AKHIR

“Merepresentasikan makna kata untuk metode klasifikasi *Naïve Bayes, Random Forest,* dan *Support Vector Machine* dalam studi kasus kemacetan di Surabaya”

# LATAR BELAKANG

Kemacetan adalah kondisi dimana arus lalu lintas yang lewat pada ruas jalan yang ditinjau melebihi kapasitas rencana jalan tersebut mengakibatkan kecepatan bebas ruas jalan tersebut mendekati 0 km/jam atau bahkan menjadi 0 km/jam sehingga mengakibatkan terjadinya antrian (MKJI,1997). Kota Surabaya merupakan kota terbesar kedua di Indonesia setelah Jakarta, Ibu Kota Jawa Timur ini berada di posisi 8 dalam daftar kota termacet di Tanah Air [1]

Twitter merupakan media sosial berbasis *microblogging* yang memungkinkan pengguna untuk *memposting* tulisan pendek yang dikenal dengan istilah *tweet*. Pada *Twitter*, untuk membuat sebuah kabar menjadi viral atau *trending topic*, pengguna dapat memberi tanda *hashtag* di depan kata kunci yang diinginkan [2]. Pada *Twitter* terdapat fitur *follow*, yaitu pengguna dapat mengikuti halaman akun pengguna lain. Dengan fitur *follow* tersebut, pengguna dapat melihat *tweet* dari pengguna lain yang di follow. Pada *tweet* hanya dibatasi 280 karakter saja. Pengguna dapat menuliskan kejadian di sekitar lingkungan mereka, contohnya adalah informasi kemacetan. Dengan mudahnya pengarsipan ucapan dan ekspresi manusia, maka volume data teks akan bertambah seiring waktu [3]. Dengan banyaknya data *tweet*, maka diperlukan sebuah model klasifikasi untuk mengklasifikasikan informasi dari *tweet*. Pada referensi-referensi Tugas Akhir sebelumnya mengenai klasifikasi maupun dataset yang menggunakan data *tweet*, dapat disimpulkan kelebihan dari Tugas Akhir yang sudah ada yaitu hasil uji yang dilakukan menghasilkan tingkat akurasi yang tinggi, contohnya pada judul “Aplikasi Deteksi Kejadian di Jalan Raya berdasarkan Data Twitter Menggunakan Metode *Support Vector Machine*“ oleh Vessa Rizky Oktavia. Pada Tugas Akhir tersebut evaluasi berupa akurasi paling baik yang didapatkan sebesar 96.25%, tetapi metode klasifikasi yang digunakan hanya menggunakan *Support Vector Machine* (SVM). Dataset yang digunakan menggunakan *tweet* dari *Twitter* [4]. Kedua, tugas akhir dengan judul “Deteksi Gempa Berdasarkan Data Twitter Menggunakan *Decision Tree, Random Forest*, dan SVM” oleh Rendra Dwi Lingga. Pada Tugas Akhir tersebut melakukan klasifikasi deteksi gempa dengan dataset *tweet* dengan menggunakan beberapa *classifier* yaitu *Decision Tree*, *Random Forest* dan SVM. Dengan menggunakan beberapa metode maka evaluasi hasil akhirnya adalah membandingkan evaluasi setiap metode dan menyimpulkan metode mana yang memiliki *recall* yang tertinggi untuk studi kasus deteksi gempa.

Oleh karena itu dalam tugas akhir ini yang akan dilakukan yaitu mengimplementasikan *word vector representation* untuk menggambarkan makna dari suatu kata dan memberikan hasil uji pada metode klasifikasi *Naïve Bayes*, *Random Forest* dan *Support Vector Machine* (SVM) berupa akurasi. Dataset untuk klasifikasi berupa *tweet* dari media sosial *Twitter*. Model klasifikasi merupakan metode *data mining* yang mengelompokkan data sesuai label kelas yang sudah ditentukan. Klasifikasi yang akan dilakukan adalah klasifikasi *tweet* dengan *dataset* yang diambil dengan hashtag #lalinsby yang diharapkan akan mengelompokkan *tweet* yang memberikan informasi tentang keadaan lalu lintas di Surabaya. Representasi kata-kata terdistribusi dalam ruang vektor diharapkan dapat membantu *learning algortihms* untuk mendapatkan hasil yang lebih baik dalam *natural language processing* dengan mengelompokkan kata-kata yang serupa. Salah satu *word vector representation* adalah *Word2vec. Word2vec* salah satu *word embedding* yang dipublikasikan pertama kali oleh Mikolov et al [5]. Vektor yang dihasilkan diharapkan bisa mewakili makna dari sebuah kata. Tugas akhir ini akan memanfaatkan *word vector representation* yaitu *Word2vec*. Selain itu untuk klasifikasi menggunakan beberapa metode *classifier*. Setelah melakukan klasifikasi data tersebut, maka akan didapatkan informasi mengenai klasifikasi *tweet* kemacetan.

# RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana cara mendapatkan representasi makna kata yang baik?
2. Apakah word2vec dapat memberikan hasil yang lebih baik untuk metode klasifikasi *Naïve Bayes, Random Forest,* dan *Support Vector Machine* pada studi kasus kemacetan?

# BATASAN MASALAH

Permasalahan pada tugas akhir ini memiliki beberapa batasan antara lain:

1. Bahasa yang digunakan pada *dataset* adalah bahasa Indonesia
2. Tidak bisa menangani kata-kata yang tidak ada dalam *corpus*
3. Dataset yang digunakan berasal dari data *tweet* dari media sosial *Twitter*

# TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan dari pembuatan tugas akhir ini adalah untuk merepresentasikan makna kata untuk metode klasifikasi dalam studi kasus kemacetan lalu lintas di Surabaya. Dimana dengan mempresentasikan makna kata dimungkinkan akan memberikan hasil uji yang lebih baik dalam kasus klasifikasi.

# MANFAAT TUGAS AKHIR

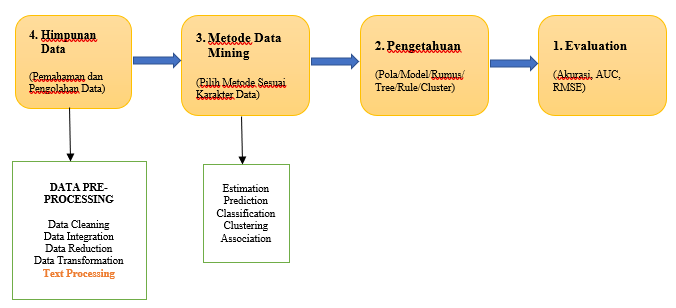
Manfaat dari pembuatan tugas akhir ini antara lain:

1. Mengetahui salah satu metode untuk merepresentasikan makna dari suatu kata
2. Mengetahui hasil dari metode klasifikasi dengan menggunakan *Word2vec*

# TINJAUAN PUSTAKA

* + 1. ***Text Mining***

*Text Mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanta didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. *Text Mining* merupakan penerapan konsep dan teknik *data mining* untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna menemukan informasi yang bermanfaat untuk tujuan tertentu [6]. Gambar dibawah ini merupakan proses *data mining*:



Gambar ‎8‑1 Proses *Data Mining* [18]

* + 1. ***Word2vec***

*Word2vec* merupakan salah satu teknik *word embedding* yang dipublikasikan pertama kali oleh Mikolov et al [5]. Sebagai gambaran bahwa vektor dari Word2vec bisa mewakili makna dari sebuah kata dan dapat mengukur beberapa vektor sebagai perbandingan [7]. Gagasan utama word2vec adalah memprediksi antara setiap kata dan kata konteksnya. Word2vec menggunakan 2 algoritma yaitu algoritma skip-grams dan *Continuous Bag of Word (CBOW)* [8]. Algoritma skip-gram untuk memprediksi kata konteks yang diberikan target (position independent) dan memaksimalkan rata-rata *log probability*, persamaannya yaitu:

(1)

dimana ‘w’ adalah word, ‘c’ adalah *size of the training context* (yang dapat menjadi fungsi dari *center word* wt). Lebih besar c maka menghasilkan lebih banyak *training example* dan dengan demikian dapat menghasilkan akurasi yang lebih tinggi dengan mengorbankan *training time* [9]. Formulasi dasar skip-gram mendefinisikan

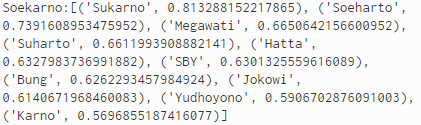
(2)

menggunakan *softmax function* untuk membuat probabilitas yang dihitung akan berada dikisaran 0 sampai 1 dan jumlah probabilitas sama dengan 1:

(3)

dimana ‘ν`w’ dan ‘ν`w ‘ adalah *input* dan *output* representasi vektor dari ‘w’, ‘W’ adalah jumlah kata dalam kosakata.

Sebagai gambaran, vektor dari word2vec bisa mewakili makna dari sebuah kata, kita bisa mengukur beberapa vektor sebagai perbandingan. Apabila kita mengukur jarak antara kata “France” dengan “Paris” maka akan ditemukan bahwa jarak yang akan muncul pada angka yang berdekatan. Berikut adalah contoh jika *word2vec* sudah dimodelkan dan kita ingin melihat kesamaan pada kata “Soekarno”, hasil yang muncul adalah :

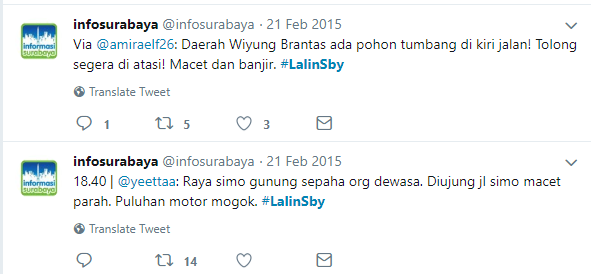


Gambar ‎8‑2 Contoh *Word2vec* [19]

Vektor-vektor yang dihasilkan merupakan output dari proses *natural language processing*.

* + 1. ***Dataset***

*Dataset* merupakan kumpulan item-item yang berhubungan dan terpisah dari data yang dapat diakses secara individual atau dalam kombinasi atau dikelola sebagai satu kesatuan utuh [10] . *Dataset* yang digunakan pada tugas akhir ini diambil dari *tweet* pada *Twitter*. *Tweet* merupakan catatan singkat dimana pengguna dapat tuliskan tentang apa yang terjadi di lingkungan sekitarnya. Berikut adalah contoh tweet tentang kemacetan:



Gambar ‎8‑3 Contoh *tweet* kemacetan

* + 1. ***Text Preprocessing***

*Text Preprocessing* merupakan sebuah tahap dimana sistem melakukan seleksi data yang akan diproses pada dokumen. Pada proses *preprocessing* ini meliputi *case folding, tokenizing, filtering, stemming* [11]. Preprocessing merupakan tahap awal yang bertujuan untuk mempersiapkan agar data teks dapat diubah menjadi lebih terstruktur dan memastikan data yang akan diolah di *data mining* adalah data yang terstruktur dan baik.

* + 1. ***Naïve Bayes***

*Naïve Bayes* merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistic. Algoritma *Naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes [12]. Ciri utama dari *Naïve Bayes* adalah asumsi yang sangat kuat akan indepedensi dari masing-masing kondisi atau kejadian.Teorema bayes memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada *database* yang besar [13]. Persamaan 1 merupakan model dari *Naïve Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi

(1)

Persamaan 2 merupakan penjabaran dari persamaan 1

(2)

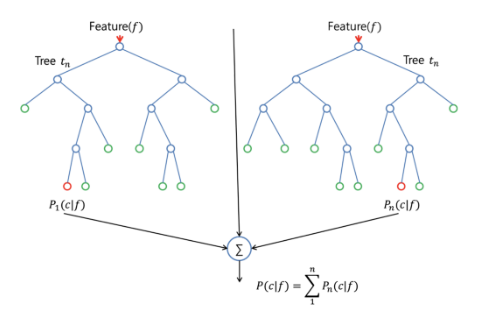
dimana ‘x’ adalah data dengan class yang belum diketahui, ‘c’ adalah hipotesis data yang merupakan suatu *class* spesifik, ‘P(C|X)’ adalah probabilitas hipotesis berdasarkan kondisi, ‘P(c) adalah probabilitas hipotesis, ‘P(x|c)’ adalah probabilitas berdasarkan kondisi pada hipotesis dan ‘P(x)’ adalah probabilitas ‘c’. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

(3)

Dimana ‘P’ adalah peluang, ‘Xi’ adalah atribut ke-I, ‘xi’ adalah nilai atribut ke-I, ‘Y’ adalah kelas yang dicari, ‘yj’ adalah sub-kelas Y yang dicari, ‘u’ adalah *mean* yang menyatakan rata-rata dari seluruh atribut dan ‘σ’ adalah standar deviasi yang menyatakan varian dari seluruh atribut. Naïve Bayes memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada database yang besar [13]. Selain itu Naïve Bayes merupakan salah satu algoritma yang sederhana tetapi memiliki kemampuan dan akurasi tinggi. Jadi, dapat dimungkinkan bahwa dengan mengklasifikasikan data kemacetan yang besar dapat menghasilkan akurasi yang baik.

* + 1. ***Random Forest***

*Random Forest*  merupakan *supervised learning algorithm*, ‘forest’ yang dibangun adalah *ensemble* *decision tree*. *Random Forest* membangun banyak *decision tree* dan menggabungkannya untuk mendapatkan prediksi yang lebih akurat dan stabil [13]. *Random Forest* bekerja dengan cara membangun lebih dari satu *Decision Tree* secara random saat training. Hasil yang diberikan oleh *Random Forest* untuk klasifikasi adalah modus dari *decision tree* nya. Sementara nilai yang diberikan untuk regresi adalah mean . Di bawah ini merupakan bagaimana tampilan *Random Forest* dengan dua *tree*:



Gambar ‎8‑1 *Random Forest* dengan dua *tree* [16]

*Random Forest*  mampu menjadi sebuah metode klasifikasi yang baik karena beberapa *Decision Tree* yang ikut dibuat saat konstruksi memiliki kemampuan prediksi yang baik. Selain itu, *Random Forest* dikembangkan untuk memperbaiki metode *Decision Tree* yang rawan overfitting dan sekarang *Random Forest* merupakan salah satu metode paling populer di *bidang machine learning* karena dapat mencapai akurasi yang tinggi tanpa perlu melakukan banyak parameter *tuning*.

* + 1. ***Support* *Vector* *Machine* (SVM)**

*Support Vector Machine* (SVM) adalah salah satu algoritma *supervised* *machine learning* yang dapat digunakan baik itu dalam masalah klasifikasi dan regresi. Dalam SVM terdapat istilah *support vector* yang merupakan titik terdekat dengan *hyperplane. Hyperplane* dapat dimisalkan sebuah garis yang memisahkan dan mengklasifikasikan secara linier satu set data [14]. Tujuan SVM adalah untuk mendapatkan *optimal hyperplane (boundary)* yang memisahkan dua kelas yang berbeda. Persamaan 1 digunakan sebagai fungsi diskriminasi untuk model diskriminatif SVM.

(1)

dimana ‘w’ adalah berat vektor, ‘b’ adalah bias dan ‘’ adalah ruang input pada dimensi fitur tingi dengan pemetaan *non-linear*. Parameter ini dipelajari otomatis dengan *maximized margin principle* pada *training dataset*

(2)

(3)

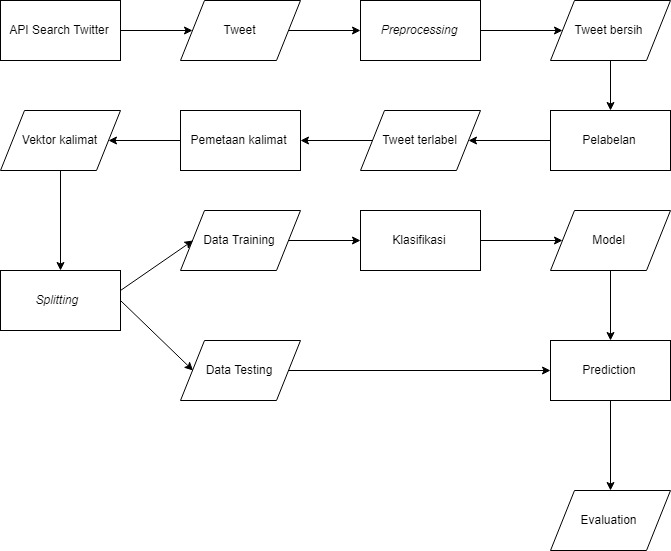
dimana ‘’ dan ‘C’ adalah *slack variable* dan koefisien penalti. Dengan diperkenalkannya fungsi *kernel*, fungsi diskriminan didefinisikan sebagai persamaan 4

(4)

SVM digunakan untuk menemukan fungsi pemisah yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda.

# RINGKASAN ISI TUGAS AKHIR

Berikut adalah penjelasan terkait tahapan tugas akhir:



Gambar 9‑1 Diagram alir tugas akhir

* + 1. **API *Search Twitter***

Pada tugas akhir ini untuk mendapatkan dataset *tweet* maka akan menggunakan *Spark Streaming*, yaitu data dapat terus masuk secara *streaming* dan data yang masuk dibatasi dengan hashtag lalinsby (#lalinsby) .

* + 1. ***Preprocessing* Data**

Pada tugas akhir ini *prepocessing data* untuk *tweet* meliputi proses *tokenizing,* mengubah kata tidak baku menjadi kata baku, *remove stopword, question word, URLs,* dan *special character*. Tahap *remove stopword, question word, URLs* dan *special character* merupakan bagian-bagian dari tahap *filtering*. Tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya, contoh penggunaan tokenizing:

manajemen

pengetahuan

adalah

sebuah

konsep

baru

di

dunia

bisnis

manajemen pengetahuan adalah sebuah konsep baru di dunia bisnis

Teks input

Teks output

Gambar 9‑2 *Tokenizing* [11]

Selanjutnya,tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token, contoh dari bagian *filtering* adalah *remove stopword*. *Remove stopword* adalah tahap dimana membuang kata-kata yang kurang penting. Contoh dari *stopword* adalah “yang”, “di”, “dari”, dan sebagainya. Berikut adalah contoh penggunaan *remove stopword*:

manajemen

pengetahuan

adalah

sebuah

konsep

baru

di

dunia

bisnis

manajemen

pengetahuan

sebuah

konsep

baru

dunia

bisnis

Hasil *filter stopword*

Hasil token

Gambar 9‑2 *Remove stopword* [11]

* + 1. **PelabelanData**

Pada tugas akhir ini pelabelan pada data *tweet* yang sudah bersih dilakukan secara manual. Pembagian label terdiri dari 2 kelas yaitu macet dan tidak macet.

* + 1. ***Pemetaan Kalimat***

Pada dataset *tweet* yang sudah bersih, kalimat dari *tweet* tersebut akan dilakukan pemetaan vektornya. Kalimat tersebut dicari vektornya dari hasil vektor perkata yang telah didapatkan dari proses *Word2vec*.

* + 1. **Metode Klasifikasi**

Pada tugas akhir ini metode klasifikasi yang digunakan adalah *Naïve Bayes, Random Forest,* dan *Support Vector Machine*. Naïve Bayes memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada database yang besar [13]. Untuk *Random Forest*, mampu menjadi sebuah metode klasifikasi yang baik karena beberapa *Decision Tree* yang ikut dibuat saat konstruksi memiliki kemampuan prediksi yang baik. Sedangkan untuk SVM, digunakan untuk menemukan fungsi pemisah yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda.

* + 1. **Evaluasi**

Pada tugas akhir ini evaluasi atau pengujian dari hasil model klasifikasi tersebut berupa nilai akurasi dari setiap metode klasifikasi. Saat hasil akurasi sudah didapatkan maka akan dilakukan evaluasi apakah model yang dibuat memiliki akurasi yang baik atau tidak baik jika menerapkan *word representation vector*.

* + 1. **Referensi**

Pada tugas akhir sebelum-sebelumnya dan referensi lainnya yaitu yang pertama dengan judul “Aplikasi Deteksi Kejadian di Jalan Raya berdasarkan Data Twitter Menggunakan Metode Support Vector Machine” oleh Vessa Rizki Oktavia. Pada tugas akhir tersebut melakukan klasifikasi kejadian yaitu kecelakaan, kemacetan dan lain-lain. Dataset *tweet* yang diambil dari akun DISHUB SURABAYA (sits\_dishubsby), infosurabaya (infosurabaya), dan Radio Suara Surabaya (e100ss), serta beberapa akun pengguna *twitter* lainnya. Tahap-tahap yang dilakukan pada tugas akhir tersebut dataset *tweet* dilakukan praproses, setelah itu dilakukan pelabelan dan ekstraksi fitur, selanjutnya dilakukan klasifikasi. Metode klasifikasi yang digunakan adalah metode *Support Vector Machine* dari sisi kernel dan parameter regularisasi variable nu.

Kedua, dengan judul tugas akhir “Deteksi Gempa Berdasarkan Data Twitter Menggunakan *Decision Tree, Random Forest*, dan SVM” oleh Rendra Dwi Lingga P. Pada tugas akhir tersebut melakukan klasifikasi deteksi gempa berdasarkan data *twitter* dengan menggunakan beberapa metode klasifikasi. Tahap-tahap yang dilakukan pada tugas akhir tersebut data *tweet* diberi label dan setelah itu dataset dilakukan *preprocessing*, setelah itu dilakukan ekstraksi fitur dan kemudian dilakukan klasifikasi menggunakan 3 *classifier* dan memberikan hasil perbandingan *weight average* pada setiap *classifier* [15].

# METODOLOGI

## Penyusunan proposal tugas akhir

Proposal tugas akhir ini berisi tentang pendahuluan tugas akhir yang akan dibuat. Pendahuluan tugas akhir berisi deskripsi tugas akhir yang meliputi latar belakang dalam pembuatan tugas akhir, rumusan masalah yang akan diangkat dalam tugas akhir, batasan masalah dalam pembuatan tugas akhir, tujuan dari dibuatnya tugas akhir, dan manfaat dari hasil pembuatan tugas akhir. Selain yang telah disebutkan, pendahuluan tugas akhir ini juga menjabarkan tinjauan pustaka yang dijadikan referensi atau pendukung dalam pembuatan tugas akhir.

## Studi literatur

Pada studi literatur ini, akan dipelajari beberapa referensi yang diperlukan antara lain mengenai *Text Mining, Word2vec, text processing, preprocessing, classifier Naïve Bayes, classifier Random Forest* dan *classifier Support Vector Machine* (SVM).

## Analisis dan desain perangkat lunak

Pada implementasi dalam mempresentasikan makna dari suatu kata, maka akan dihasilkan suatu makna kata dalam bentuk vektor. Selain itu juga akan dijabarkan hasil uji dari penggunaan *Word2vec* untuk metode klasifikasi *Naïve Bayes, Random Forest* dan *Support Vector Machine*(SVM)*.*

## Implementasi perangkat lunak

Penerapan metode *Word2vec* ini akan dibangun menggunakan bahasa pemrograman Python dengan IDE Spyder, dengan berbagai macam library pendukungnya yang terdapat pada Python.

Pada tugas akhir ini, kode *Word2vec* tidak akan ditulis dari awal, melainkan menggunakan source code *Word2vec* yang dipublikasikan secara open source di alamat: <http://textminingonline.com/training-word2vec-model-on-english-wikipedia-by-gensim> dan [https://yudiwbs.wordpress.com/2018/03/31/word2vec-wikipedia-bahasa-indonesia-dengan-python-gensim](https://yudiwbs.wordpress.com/2018/03/31/word2vec-wikipedia-bahasa-indonesia-dengan-python-gensim/).

Masukan untuk metode *Word2vec* adalah hasil *dump* dari artikel-artikel dari Wikipedia Bahasa Indonesia yang dapat diakses melalui alamat: [https://dumps.wikimedia.org/idwiki/latest](https://dumps.wikimedia.org/idwiki/latest/) dan data yang diambil adalah yang terformat XML.

Dalam mengaplikasikan metode klasifikasi, preprocessing, *splitting* *data training* dan *testing*, *modelling* dan evaluasi menggunakan *RapidMiner Studio 8.1*.

## Pengujian dan evaluasi

Pengujian dan evaluasi pada tugas akhir ini dilakukan dengan menggunakan *RapidMiner Studio 8.1* dengan evaluasi hasil uji berupa akurasi dari hasil klasifikasi.

Akurasi =

Dimana TP, TN, FP, dan FN adalah *true positive, true negative, false positive, false negatives.*

## Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
   1. Latar Belakang
   2. Rumusan Masalah
   3. Batasan Tugas Akhir
   4. Tujuan
   5. Metodologi
   6. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

# JADWAL KEGIATAN

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tahapan | 2018 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mei | | | | Juni | | | | Agustus | | | | September | | | | Oktober | | | | November | | | | Desember | | | | |
| Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Perancangan Metode |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Pengujian dan Evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penyusunan Buku |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | A. Ramadhiani, “properti.kompas,” 2018. [Online]. Available: https://properti.kompas.com/read/2018/02/25/182046621/ini-10-kota-termacet-di-indonesia. [Diakses 23 5 2018]. |
| [2] | Seon, “Seon,” 2018. [Online]. Available: https://seon.co.id/pengertian-media-sosial-facebook-twitter-google-youtube-instagram/. [Diakses 23 5 2018]. |
| [3] | C. C. Aggarwal, Data Mining The Textbook, vol. 181, London: Springer International Publishing Switzerland, 2015, pp. 1138-1152. |
| [4] | V. R. Oktavia, Aplikasi Deteksi Kejadian di Jalan Raya berdasarkan Data Twitter Menggunakan Metode Support Vector Machine, Surabaya: Institut Teknologi Sepuluh Nopember, 2017. |
| [5] | T. Mikolov, K. Chen, G. Corrado dan J. Dean, “Efficient Estimation of Word Representations in,” 2013. |
| [6] | H. Nindito, “School of Information Systems,” Binus, 15 12 2016. [Online]. Available: https://sis.binus.ac.id/2016/12/15/teori-text-mining-dan-web-mining/. [Diakses 23 5 2018]. |
| [7] | D. Suhartono, “SOCS Binus,” 22 12 2016. [Online]. Available: https://socs.binus.ac.id/2016/12/22/word-vector-representation-word2vec-glove/. [Diakses 5 5 2018]. |
| [8] | S. U. S. o. Engineering, “youtube,” 3 4 2017. [Online]. Available: https://www.youtube.com/watch?v=ERibwqs9p38. [Diakses 5 5 2018]. |
| [9] | T. Mikolov, I. Sutskever, K. Chen, G. Corrado dan J. Dean, “Distributed Representations of Words and Phrases,” no. In NIPS Conference, 2013. |
| [10] | M. Rouse, “whatsis.techtarget,” 3 2016. [Online]. Available: https://whatis.techtarget.com/definition/data-set. [Diakses 5 5 2018]. |
| [11] | Informatikalogi, “Informatikalogi.com,” 27 12 2016. [Online]. Available: https://informatikalogi.com/text-preprocessing/. [Diakses 5 5 2018]. |
| [12] | Informatikalogi, “Informatikalogi.com,” 27 12 2016. [Online]. Available: ttps://informatikalogi.com/algoritma-naive-bayes/. [Diakses 5 5 2018]. |
| [13] | J. Han, M. Kamber dan J. Pei, Data Mining Concepts and Techniques 3rd edition, Elsevier Inc, 2012. |
| [14] | B. Pang dan L. Lee, Opinion Mining and Sentiment Analysis, now Publishers Inc, 2008. |
| [15] | R. D. Lengga P, “Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM,” 2017. |
| [16] | N. Donges, “Towards Data Science,” [Online]. Available: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd. [Diakses 5 5 2018]. |
| [17] | J. Brownlee, “Machinelearningmastery,” 6 10 2017. [Online]. Available: https://machinelearningmastery.com/develop-word-embeddings-python-gensim/. [Diakses 23 5 2018]. |
| [18] | R. S. Wahono, “romisatriawahono.net,” 11 2017. [Online]. Available: http://romisatriawahono.net/dm/. [Diakses 22 5 2018]. |
| [19] | Y. Wibisono, “yudiwbs,” 31 3 2018. [Online]. Available: https://yudiwbs.wordpress.com/2018/03/31/word2vec-wikipedia-bahasa-indonesia-dengan-python-gensim/. [Diakses 26 5 2018]. |