

Penerapan Metode *Support Vector Machine (SVM)* Menggunakan *Kernel Radial Basis Function (RBF)* Pada Klasifikasi *Tweet*

Imelda A.Muis¹, Muhammad Affandes, M.T²

^{1,2} Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
Email: imeldaamuis@uin-suska.ac.id, affandes@uin-suska.ac.id

(Received: 8 April 2015; Revised: 4 Juni 2015; Accepted: 25 Juni 2015)

ABSTRAK

Twitter merupakan wadah atau tempat untuk berbagi informasi dan juga dapat digunakan untuk berkampanye dan berpromosi barang atau jasa, juga sering disebut dengan bisnis. Twitter merupakan salah satu media sosial yang dapat digunakan untuk melakukan hal tersebut. Pada penelitian ini akan dilakukan pengklasifikasian data *tweet* menggunakan metode *support vector machine (SVM)* tersebut agar *tweet* yang ada tidak bercampur antara iklan dan tidak iklan. SVM salah satu metode yang dapat melakukan pengklasifikasian data dengan baik, karena proses yang akan dilakukan bersifat *non linear* maka akan menggunakan kernel RBF (*Radial Basis Function*) dimana parameter yang akan digunakan adalah nilai C dan γ . Dari hasil uji coba, aplikasi menunjukkan akurasi stabil pada rentang nilai $0 \leq C \leq 3$ dan $0.01 \leq \gamma \leq 10$ pada data yang belum dilakukan pemilihan fitur dan akurasi stabil pada rentang nilai $0 \leq C \leq 300$ dan $0.01 \leq \gamma \leq 10$. Dengan pencapaian nilai akurasi yang baik maka, hasil ini dapat diterapkan untuk membantu pengguna *twitter* untuk melakukan filter terhadap *tweet* iklan yang terdapat pada akun Twitter mereka.

Kata Kunci: Klasifikasi, *Radial Basis Function*, *Support Vector Machine*, *Tweet*, Twitter

ABSTRACT

Twitter is one of container or place to share information and also can be used for campaign and promote things or service called business. Twitter is one of social media do so. This research will classify data of tweet using support vector machine (SVM) method so that tweet don't mingled between a advertisement and not advertisement. SVM is one of method that can classify data well. Because the process will do a non linear, so will use kernel RBF (radial basis function) where parameter will be used is C and γ . And the result, application shows a stable accuration between range $0 \leq C \leq 3$ and $0.01 \leq \gamma \leq 10$ on data that have not done feature selection and stable accuration on between range $0 \leq C \leq 300$ and $0.01 \leq \gamma \leq 10$, by reaching a good accuration grade, the result can help twitter users to filter toward advertisement tweet that belong to them.

Keywords: Classifier, *Radial Basis Function*, *Support Vector Machine*, *Tweet*, Twitter.

Corresponding Author

Imelda A.Muis
Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau,
Email: imeldaamuis@uin-suska.ac.id

Pendahuluan

Iklan merupakan media komunikasi massa yang sangat efektif dalam menyampaikan pesan kepada khayalak. Pada penelitian ini media sosial yang akan dibahas adalah twitter. Media sosial twitter juga banya digunakan oleh para produsen barang atau jasa sebagai alat pemasaran. Para produsen cukup hanya dengan mengirimkan tweet yang berhubungan dengan usaha yang dikembangkan kepada pengguna twitter lainnya.

Iklan menurut [3] adalah segala bentuk pesan yang berbentuk barang atau produk dan jasa yang

disalurkan melalui media yang dibiayai oleh perusahaan yang bersangkutan yang ditujukan kepada sebagian atau seluruh masyarakat

Beberapa perusahaan besar telah melakukan iklan melalui Twitter karena melakukan iklan pada Twitter sangat memberi kemudahan kepada produsen karena mereka hanya cukup membagikan *tweet* promo kepada *follower*. Produsen yang telah melakukan promo melalui Twitter adalah Zalora dan Zalada. Zalora adalah salah satu *e-commerce fashion* terbesar di Indonesia, pernah mendapatkan penghargaan dari Museum Rekor Dunia Indonesia (MURI) dengan rata-rata pengunjung 150.000.

Zalora Indonesia menggunakan Facebook untuk beriklan. Berdasarkan penjelasan diatas banyak pengguna sosial media yang diuntungkan dengan adanya iklan pada soaial media, namun ada juga pengguna sosial media yang merasa terganggu dengan adanya iklan pada akun sosial miliknya.

Banyak penelitian – peneltian sebelumnya yang dikembangkan pada area klasifikasi khususnya teks untuk analisa *sentiment*, yaitu: klasifikasi *tweet entity* pada Twitter. Beberapa penelitian klasifikasi teks yang pernah dilakukan (Nur dan Santika, 2011 dimana mereka mengambil *tweet* untuk dataset dan SVM untuk metode klasifikasi untuk memperoleh akurasi *sentiment* terhadap merek telepon seluler) yang dikutip oleh [8] penelitian [4], Berdasarkan penelitian sebelumnya maka pada penelitian ini akan melakukan pengklasifikasian, berbeda dengan penelitian sebelumnya pada penelitian sekarang, penulis akan mencoba melakukan pengklasifikasian untuk “PENERAPAN METODE *SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN KERNEL RADIAL BASIS FUNCTION (RBF) PADA KLASIFIKASI TWEET.*”

Penelitian ini ditujukan untuk mengelompokkan *tweet* sesuai dengan kategorinya, pada penelitian ini penulis menggunakan dua kategori yaitu, kategori iklan dan kategori bukan iklan. Pemilihan metode SVM dikarenakan menurut (Joachims, 1998) dikutip oleh [8] menyatakan bahwa berdasarkan penelitian yang telah dia lakukan berpendapat bahwa SVM adalah klasifikasi yang paling tepat untuk melakukan pengklasifikasian teks. Dengan kata lain, SVM merupakan metode yang dapat diterapkan untuk mengklasifikasi teks pada *tweet entity* dengan tingkat keakurasian relatif lebih baik dibanding metode lain

Metode Penelitian

Text Mining

Text mining merupakan penemuan kembali data yang telah lama tersimpan di dalam teks, data tersebut tidak harus baru (Hearst dan Marti A,1999) dikutip oleh [6]. Pengertian *text mining* sudah banyak disampaikan oleh banyak peneliti. Dalam memberikan solusi *text mining* mengadopsi atau mengembangkan banyak teknik dan solusi dari bidang lain: Informasi Retrieval(IR), Data Mining, Statistik dan Ilmu Matematika, *Machine learning*, *Linguistik*, *Natural Language Processing*, dan *Visualization*. *Text mining* dapat digunakan untuk *riset* atau penelitian yang berkaitan dengan, *Searching*, Ekstraksi informasi, *Clustering*, *Categorization*, *Summarization*, Informasi Monitoring, *Question and answare* yang dikutip oleh [6]

Preprocessing

Text preprocessing merupakan tahapan sangat penting dalam melakukan proses klasifikasi data teks. Tujuan dilakukannya *text preprocessing* yaitu untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata. Berikut tahapan didalam *preprocessing* data teks.

1. *Tokening*
2. *Transform Case*
3. *Filtering*
4. *Stemming*
5. *Filter Token*
6. *Feture Selecton*

Pelabelan Data

Didalam permasalahan klasifikasi hal yang paling utama adalah melakukan pelabelan dengan benar agar data yang terkumpul atau terklasifikasi sesuai dengan kelas yang seharusnya. Pada penelitian ini akan dilakukan pelabelan *tweet* iklan dan *tweet* tidak iklan. Untuk melakukan pelabelan ini Menurut (Peni Adji, belajar iklan) yang dikutip dari [2] iklan yang bagus itu memenuhi kriteria rumus AIDCA, yaitu:

1. *Attencion* (perhatian)
2. *Interst* (minat)
3. *Desire* (kebutuhan)
4. *Conviction* (keinginan)
5. *Action* (tindakan)

Support Vector Machine (SVM)

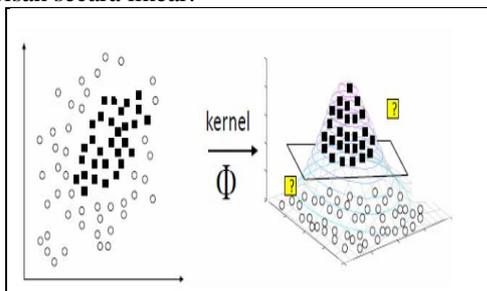
Menurut [5]SVM pertama kali diperkenalkan oleh Boser, Guyon, Vepnik, yang dipresentasikan untuk pertama kalinya pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. konsep dari svm merupakan kombinasi harmonis dari konsep komputasi yang sudah ada puluhan tahun sebelumnya, seperti *hyperplane* (Duda dan Hart, 1973, cover 1965, Vapnik,1964). Kernel diperkenalkan oleh Aronszajn,1950) dan demikian dengan konsep-konsep lainnya.

Prinsipnya SVM bekerja secara *linear*, dan dikembangkan untuk dapat diterapkan pada masalah *non-linear*. Dengan menggunakan metode *kernel trick* yang mencari *hyperplane* dengan cara mentransformasi *dataset* ke ruang vektor yang berdimensi lebih tinggi (*feature space*), kemudian proses klasifikasi dilakukan pada *feature space* tersebut. Penentuan fungsi kernel yang digunakan akan sangat berpengaruh terhadap hasil prediksi. Misalkan $\{x_1, \dots, x_n\}$ 1 adalah dataset dan $y_i \in \{+1, -1\}$ adalah label kelas dari data x_i . Pada gambar 1 dapat dilihat berbagai alternatif bidang pemisah yang pemisah terbaik tidak hanya dapat memisahkan data tetapi juga memiliki *margin*

paling besar. Data yang berada tepat pada bidang pemisah disebut sebagai *support vector* (SV). Berdasarkan data yang digunakan berupa data *text* maka proses svm yang digunakan akan proses *non linear*.

SVM Non Linier

Berikut metode yang dapat digunakan untuk mengklasifikasikan data yang tidak dapat diklasifikasikan secara linier adalah dengan cara mentransformasikan data kedalam ruang fitur (*feature space*) berdimensi tinggi sehingga dapat dipisah secara linier pada fitur space. *Feature space* dalam prosesnya biasanya memiliki dimensi yang lebih tinggi dari vektor *input* (*input space*). Hal ini akan mengakibatkan komputasi pada *feature space* akan menjadi sangat besar, karena ada kemungkinan *feature space* akan memiliki jumlah *feature* yang tidak terhingga. Selain itu, untuk mengetahui fungsi transformasi yang tepat juga sangat sulit. Untuk itu pada SVM menggunakan “*kernel trick*”. Syarat sebuah fungsi untuk menjadi fungsi kernel adalah memenuhi Teorema Mercer yang menyatakan bahwa matriks kernel yang dihasilkan harus bersifat *positif semi-definite*. Berikut gambar data yang tidak dapat dipisah secara linear:



Gambar 1. Data non linear

Feature space biasanya memiliki dimensi yang lebih tinggi dari vektor *input* (*input space*). Hal ini mengakibatkan komputasi pada *feature space* sangat besar, karena ada kemungkinan *feature space* memiliki jumlah *feature* yang tidak terhingga dan juga sulit mengetahui fungsi transformasi yang tepat. Untuk mengatasi masalah tersebut, pada SVM menggunakan *kernel trick*. Dari persamaan (2.23) terdapat *dot product* $\phi(x_i)\phi(x_d)$. Jika terdapat fungsi kernel K, sehingga $K(x_i, x_d) = \phi(x_i) \cdot \phi(x_d)$, maka fungsi transformasi $\phi(x_k)$ tidak perlu diketahui secara pasti. Fungsi yang dihasilkan dari pelatihan yaitu:

$$f(x_d) = \sum_{i=1}^{ns} a_i y_i K(x_i, x_d) + b \quad (x_i \text{ support vector}) \quad (1)$$

Sebuah fungsi bisa menjadi fungsi kernel jika memenuhi Teorema Mercer, yang menyatakan bahwa matriks kernel yang dihasilkan harus

bersifat semi *positive* semi *definite*. Menurut Hsu, dkk (2010), yang diikuti oleh [8] berikut ini adalah beberapa fungsi kernel yang umum digunakan yaitu:

- a. Kernel linier
 $K(x_i, x) = x_i^T x$
- b. Polynomial
 $K(x_i, x) = (\gamma \cdot x_i^T x + r)^\rho, \gamma > 0$
- c. Radial basis function
 $K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0$
- d. Sigmoid kernel
 $K(x_i, x) = \tanh(\gamma x_i^T + r)$

Pada penelitian ini akan diterapkan kernel *Radial Basis Function* (RBF) dengan parameter C dan Gamma.

Evaluasi Model Hasil Training

Validation model dapat diukur dengan menghitung keakurasi data. Akurasi dapat dirumuskan sebagai berikut :

$$\text{akurasi} = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah dokumen ujicoba}} \times 100\% \quad (2)$$

Optimasi Hyperplane

Optimasi *Hyperparameter* ini juga dikenal disebut seleksi model, atau pencarian parameter (*parameter search*). Proses ini dilakukan untuk mencari hyperparameter terbaik, biasanya dikenal dengan sebutan *grid search*. Algoritma *grid search* ini biasanya menggunakan fungsi *k-fold cross validation*. pencarian parameter terbaik akan dilakukan dengan membagi data menggunakan *k-fold cross validation* yaitu pada penelitian akan menggunakan *10-fold cross validation*. Setelah itu satu dari sepuluh akan menjadi data uji dan yang lainnya akan menjadi data latih, proses ini akan terus berjalan sampai seluruh data berkesempatan untuk menjadi data uji dan data latih. Berdasarkan kernel yang digunakan maka pada penelitian ini nilai c dan gamma akan ditetapkan dari rentang $0 \leq C \leq 1000$ dan $0.01 \leq \gamma \leq 10$.

Rapid Miner

Rapid Miner adalah salah satu aplikasi *opensource* yang dapat digunakan untuk melakukan proses data mining. Salah satu metode data mining adalah menggunakan regresi linier. Regresi linier merupakan metode statistik yang digunakan untuk melakukan estimasi atau perkiraan berdasarkan data yang ada [1] Rapid Miner menyediakan prosedur data mining dan machine learning termasuk : ETL (*Extraction, Transformation, Loading*), data *preprocessing*, *visualisasi*, *modelling*, dan evaluasi.

Bagian – bagian pada tampilan Rapid Miner
 A. Tipe Nilai

Pada Tools RapidMiner ada beberapa tipe nilai yang digunakan yaitu:

1. Nominal
Nominal adalah tipe nilai yang digunakan berdasarkan nilai secara kategori.
2. Numeric
Nilai numerik secara umum
3. Integer
Tipe nilai yang digunakan untuk bilangan bulat
4. Real
Tipe nilai yang digunakan untuk bilangan yang nyata
5. Text
Tipe nilai yang digunakan untuk teks bebas tanpa struktur.
6. Binomial
Tipe nilai yang digunakan untuk nilai yang terdiri dari dua nilai
7. Polynomial
Digunakan untuk nominal lebih dari dua nilai.
8. Date_Time
Digunakan untuk tanggal dan waktu

B. Prespektif Dan View

Sebuah prespektif berisikan pilihan elemen-elemen GUI, yang disebut dengan View, yang dapat dikonfigurasi secara bebas. Berikut perspective yang terdapat pada tools RapidMiner:

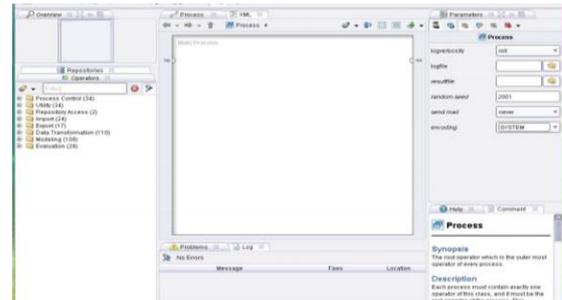
1. Perspektifif Selamat Datang (Welcome prespective)
2. Perspektif desain (Desain prespective)
3. Perspektif hasil (Result Prespective)

Berikut tampilan gambar prespektif pada tampilan tools RapidMiner



Gambar 2. Tampilan prespektif pada tools RapidMiner

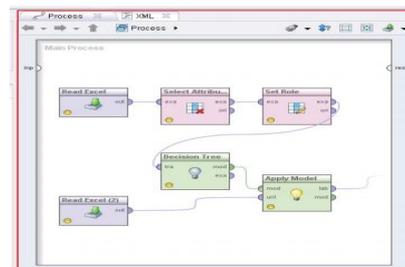
Perspektif pada tahapan desain merupakan tempat dilakukan proses utama. Berikut tampilan dari perspektif desain:



Gambar 3. Tampilan prespektif desain

Berikut penjelasan dari gambar diatas:

1. View Operator
 - a. Proses Control
Untuk mengontrol aliran proses, seperti *loop* atau *conditional branch*
 - b. Utility
Untuk mengelompokkan *subprocess*, juga *macro* dan *logger*.
 - c. Repository Acces
Untuk membaca dan menulis Repositori
 - d. Import
Untuk membaca data dari berbagai format eksternal
 - e. Export
Untuk menulis data ke berbagai format eksternal
 - f. Data Transformation
Untuk transformasi data dan metadata
 - g. Modelling
Untuk proses data mining yang sesungguhnya seperti klasifikasi, regresi, clustering, aturan asosiasi dan lain-lain.
 - h. Preprocessing
Untuk proses text mining seperti *tokenizing*, *stemming*, *stopword*.
 - i. Evaluation
Untuk menghitung kualitas dari modelling.
2. Main Proses
Operator – operator diatas yang akan dijalankan di main proses dengan cara mendrag operator yang akan digunakan. Berikut tampilan dari penggunaan operator diatas pada main proses.



Gambar 4. Tampilan penjalan operator pada main proses

Operator yang dijalankan pada Main Proses akan memerlukan parameter agar dapat berfungsi. Setelah operator dipilih dan dijalankan pada main proses, parameter dari operator tersebut akan ditampilkan pada view parameter.

3. Parameter View

Parameter View berfungsi sebagai tempat untuk menampilkan parameter yang terdapat pada operator yang dijalankan pada main proses. Berikut tampilan dari parameter view.

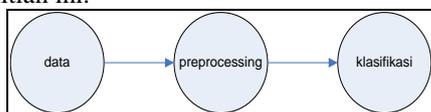


Gambar 5. Tampilan parameter view

Hasil dan Pembahasan

Gambaran Secara Umum

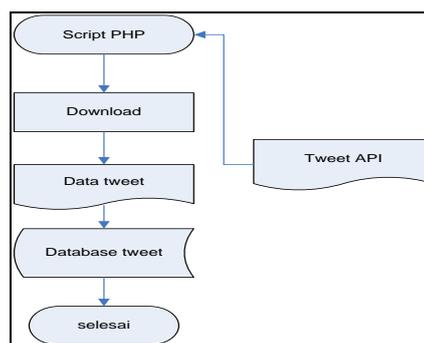
Secara gambaran umum proses pada penelitian ini melalui tiga tahapan proses yaitu: proses pengambilan data (*download tweet*), *preprocessing*, dan yang terakhir klasifikasi. Berikut gambaran secara umum tahapan dalam penelitian ini.



Gambar 6. Gambar umum proses

Analisa Sumber Data

Pada penelitian ini data yang akan digunakan bersumber dari Twitter. Pada Twitter tersedia *Application Programming Interface (API)*. Menggunakan Twitter API ini data yang dibutuhkan pada penelitian ini diperoleh. Proses untuk mendapatkan data dengan API *twitter* tidak cukup dengan *script PHP* saja, pertama kita harus memiliki akun *twitter* terlebih dahulu agar dapat melakukan registrasi ke Twitter API untuk memperoleh serial token yang digunakan untuk dapat mengakses Twitter API. Serial token ini yang akan digunakan pada *script PHP* agar dapat melakukan *download* data yang akan digunakan pada penelitian ini. Data yang disediakan API berupa nama *user*, isi *tweet*, tanggal. Alamat dari *tweet* API sebagai berikut <https://api.twitter.com/1.1/>. Berikut diagram proses pengambilan atau *download* data:



Gambar 7. Flowchart download data

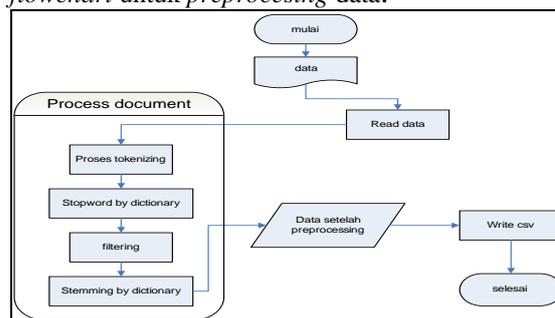
Analisa Pelabelan

Pada penelitian ini proses klasifikasi akan menggunakan dua kelas yaitu: kelas iklan dan kelas tidak iklan.

1. Kelas iklan, mengandung kata promo, jual, harga, beli dan mampu membuat masyarakat terfokus untuk melihat dan mendengarkan iklan tersebut. Sehingga membuat masyarakat ingin membeli produk yang di iklankan.
2. Kelas tidak iklan, unsur tidak iklan semua kata yang tidak terdapat pada unsur iklan. Pelabelan pada penelitian ini akan dilakukan secara manual, dilakukan langsung pada *database*.

Analisa Preprocessing

Preprocessing data pada penelitian ini menggunakan *tools*. *Tools* yang akan digunakan pada penelitian ini yaitu Rapid Miner. Pada tahapan ini *preprocessing* data akan melalui beberapa tahapan yaitu: *tokenizing*, *stopword*, *filtering* dan *stemming*. Pada RapidMiner keempat process tersebut akan dilakukan pada tahapan yang disebut dengan *process documnet*. Berikut *flowchart* untuk *preprocessing* data.



Gambar 8. Flowchart preprocessing data

Berikut penjelasan dari gambat diatas:

1. Proses *Read* data

Pada proses ini data yang tersimpan didalam *database my sql* akan dibaca terlebih dahulu kedalam Rapid Miner untuk dapat dilakukan *process document*.

2. Process document

Process document data yang sudah berada didalam RapidMiner akan melalui tahapan *process document*. Pada proses ini akan dilakukan beberapa sub proses sebelum dilakukan penyimpanan data dalam bentuk format CSV. Berikut penjelasan dari sub proses dari *process document*:

a. Proses *tokenizing*

Pada tahapan ini data yang sudah diinputkan kedalam *process document* akan dilakukan pemecahan kalimat sehingga menjadi kata-kata yang tunggal. Pada penelitian ini proses akan diterapkan pada *tools* RapidMiner. Hasil dari proses ini akan dilakukan proses *stopword*.

b. Proses *stopword*

Pada proses *stopword* data yang digunakan berasal dari proses *tokenizing*. Tahapan ini akan dilakukan penghilangan kata-kata yang tidak digunakan dalam proses klasifikasi seperti kata: di, yang, dengan dan lain-lain. Proses ini akan diterapkan pada RapidMiner dengan menggunakan kamus yang dibuat sendiri.

c. *Filtering*

Data yang berasal dari *stopword* akan dilakukan *filtering* untuk membatasi ukuran kata yang akan dilakukan proses selanjutnya. Pembatasan ini dimulai dari ukuran minimal dan maksimal.

d. *Stemming*

Setelah data memasuki proses *filtering* maka akan dilakukan proses terakhir dalam *process document*. Proses ini akan menentukan kata dasar dari suatu kata yang memiliki imbuhan, akhiran, awalan, sisipan, awalan dan akhiran. Berikut contoh kata yang memiliki imbuhan dengan kata dasarnya:

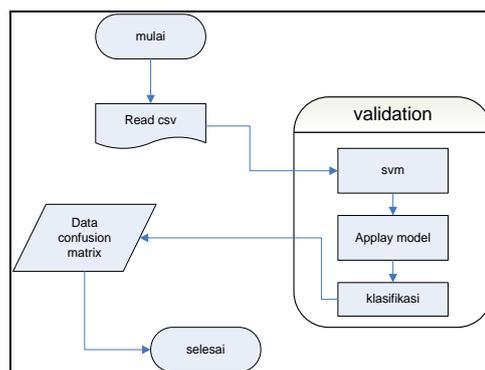
Memakan = makan
 Menyapu = sapu

3. Write CSV

Data yang sudah melalui seluruh *process document* kemudian akan di simpan dalam bentuk format CSV. Data yang dalam format CSV ini kemudian akan dilakukan proses klasifikasi yang akan diterapkan pada *tools* RapidMiner.

Klasifikasi Data

Pada penelitian ini proses klasifikasi akan menggunakan *tools* RapidMiner sebagai *tools* pembantu. Sebelum dilakukan klasifikasi data, sebelumnya akan dilakukan proses pembelajaran dengan menggunakan data latih terlebih dahulu. Pada pembelajaran ini akan menghasilkan model. Model hasil dari pembelajaran ini kemudian akan dilakukan pengujian dengan beberapa data yang akan digunakan sebagai data uji. Berikut *flowchart* untuk klasifikasi.



Gambar 9 Flowchart tahapan klasifikasi

Berikut penjelasan dari proses *validation*:

Cross validation atau pada *tools* RapidMiner juga disebut *number of validation*. *Cross validation* digunakan untuk menemukan nilai akurasi terbaik dengan cara melakukan pelatihan dan pengujian terhadap data yang sebelumnya sudah dibagi menjadi beberapa bagian. Data tersebut akan dilakukan proses *iterasi* dimana satu data akan dijadikan sebagai data uji dan selebihnya akan dijadikan data latih. Proses ini akan dilakukan sampai seluruh data dapat menjadi data uji dan yang lain juga dapat menjadi data latih. Pada penelitian ini *cross validation* akan dibagi menjadi 10 *fold cross validation*. Dari 1020 data yang ada maka 1 bagian data akan terdiri dari 102 data dengan 10 bagian.

a. Analisa Penerapan SVM

Algoritma SVM sebagai metode yang akan digunakan untuk menyelesaikan kasus klasifikasi pada penelitian ini. Pada proses klasifikasi akan menggunakan *function* LibSVM yang dapat digunakan untuk klasifikasi SVM. Pada *function* LibSVM formula yang akan digunakan C-SUPPORT VECTOR CLASSIFIER (C-SVC) dengan kernel *Radial Basis Function* (RBF).

Pada penelitian ini data yang digunakan berbentuk teks proses klasifikasi tidak dapat di asumsikan secara *linear*, oleh karena itu akan digunakan kernel sebagai parameter yang dapat melakukan klasifikasi terhadap data yang bersifat tidak *linear*. Seperti yang sudah disebutkan sebelumnya pada bab 2 kernel yang akan digunakan kernel RBF dengan parameter c dan γ . Penerapan SVM pada penelitian ini akan dilakukan pada *tools* RapidMiner.

Berdasarkan dari kernel yang digunakan pada penelitian ini, maka parameter c dan γ akan digunakan sebagai parameter pendukung. Pada penelitian ini rentang nilai yang akan digunakan ada beberapa bagian rentang nilai.

- $100 \leq c \leq 1000$ dan $100 \leq \gamma \leq 1000$
- $10 \leq c \leq 100$ dan $10 \leq \gamma \leq 100$
- $1 \leq c \leq 10$ dan $1 \leq \gamma \leq 10$
- $0 \leq c \leq 1$ dan $0.01 \leq \gamma \leq 1$
- $0 \leq c \leq 3$ dan $0.01 \leq \gamma \leq 10$

b. *Apply Model*

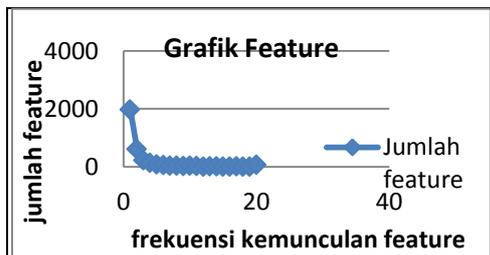
Apply model digunakan untuk dapat membaca data yang akan diestimasi berdasarkan data yang sudah dipelajari sebelumnya.

c. *Klasifikasi*

Tahapan klasifikasi untuk melakukan pengklasifikasian terhadap data yang akan diinputkan. Proses klasifikasi ini akan menggunakan model yang dihasilkan dari proses pembelajaran yang disimpan dalam proses *apply* model.

Analisa Akurasi

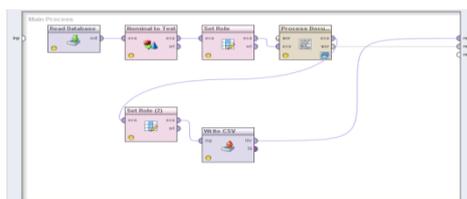
Penerapan paramater c dan γ dengan cara memasang nilai c dan γ dengan rentang nilai yang sudah disebutkan sebelumnya. Pada pengujian akurasi ini akan dilakukan pengujian dengan menggunakan dua model data yaitu: data yang belum dilakukan pemilihan *feature* dan data yang sudah dilakukan pemilihan *feature*. Jumlah *feature* pada data yang belum dilakukan pemilihan *feature* ± 2000 *feature* sedangkan untuk data yang sudah dilakukan pemilihan *feature* ± 200 *feature*. Pemilihan *feature* ini dilakukan dengan cara meranking nilai kemunculan kata. Berikut gambar grafik ekstraksi *feature*



Gambar 10. Grafik ekstraksi *feature*

Implementasi preprocessing data

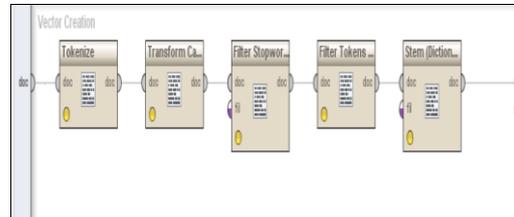
Pada tahapan ini, *tweet* yang sudah terkumpul akan dilakukan *preprocessing* data dengan menggunakan *tools* RapidMiner. Tahapan *preprocessing* data ini meliputi tahapan membaca data dari *database*, *nominal to text*, *set role*, *process document*, *set role* dan terakhir menyimpan data dalam bentuk format CSV atau dalam bentuk format lainnya. Berikut tampilan pada tahapan *preprocessing* data.



Gambar 11. Tampilan *preprocessing* data

Process Document

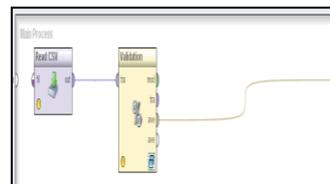
Tahapan ini merupakan tahapan pembersihan data. Pada penelitian ini *process document* meliputi proses *tokenize*, *transform case*, *filter stopwords*, *filter tokenize by light* dan terakhir proses *stemming*. Berikut tampilan tahapan *process document*.



Gambar 12. *Process document*

Klasifikasi

Tidak jauh berbeda dengan proses data sebelumnya, pada tahapan ini juga akan menggunakan *tools* RapidMiner untuk melakukan klasifikasi pada data. Berikut tampilan untuk proses klasifikasi.



Gambar 13. Proses klasifikasi

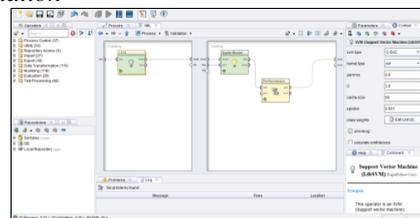
Berikut penjelasan untuk proses klasifikasi data:

1. *Read CSV*

Proses dimana data yang memiliki bobot nilai dipanggil untuk dilakukan proses *validation*.

2. *Validation*

Proses *validation* adalah inti dari proses pengklasifikasian data. Pada proses *validation* ini semua fungsi, pengujian dan testing akan dilakukan. Fungsi yang digunakan pada penelitian ini yaitu fungsi LibSVM dengan kernel RBF dengan fungsi nilai (γ, c), dan menggunakan *number of validation*. Berikut tampilan dari proses *validation*



Gambar 14. Proses *validation*

Pengujian

Tujuan dari dilakukannya pengujian dalam penelitian ini adalah untuk melihat tingkat akurasi klasifikasi data dalam menentukan kelas suatu data. Berdasarkan hasil dari model pada tahapan *training*.

γ c	0.01	0.03	0.06	0.09	0.1	0.3	0.6	0.9	1	3	6	9	10
0.0	77.55 %	77.16 %	77.16 %	86.57 %	91.08 %	95.20 %	95.69 %	95.20 %	95.00 %	88.04 %	80.39 %	73.53 %	72.55 %
0.1	77.55 %	77.35 %	77.06 %	76.76 %	76.76 %	76.76 %	72.16 %	70.76 %	69.71 %	60.00 %	58.82 %	56.18 %	55.98 %
0.3	77.55 %	77.35 %	77.06 %	76.76 %	76.76 %	86.08 %	88.73 %	88.24 %	87.55 %	73.43 %	61.37 %	57.45 %	56.67 %
0.6	77.55 %	77.35 %	77.75 %	84.31 %	85.49 %	92.75 %	93.73 %	92.94 %	92.65 %	82.75 %	74.31 %	68.73 %	68.53 %
0.9	77.55 %	77.35 %	77.55 %	89.80 %	90.59 %	94.51 %	95.39 %	94.80 %	94.41 %	87.06 %	79.02 %	70.98 %	70.49 %
1	77.55 %	77.16 %	86.57 %	90.78 %	91.08 %	95.20 %	95.69 %	95.20 %	95.00 %	88.04 %	80.39 %	73.53 %	72.55 %
1.5	77.56 %	82.16 %	90.98 %	93.24 %	93.43 %	96.57 %	96.96 %	96.18 %	95.88 %	88.73 %	82.65 %	76.57 %	73.63 %
2	77.56 %	87.35 %	92.94 %	94.31 %	94.90 %	97.25 %	97.06 %	96.18 %	95.88 %	88.73 %	82.65 %	76.57 %	73.63 %
2.5	77.55 %	90.10 %	94.02 %	95.39 %	95.78 %	97.44 %	97.06 %	96.18 %	95.88 %	88.73 %	82.65 %	76.57 %	73.63 %
3	77.45 %	91.18 %	94.90 %	96.08 %	96.27 %	97.44 %	97.06 %	96.18 %	95.88 %	88.73 %	82.65 %	76.57 %	73.63%

Akurasi diukur berdasarkan perbandingan antara data yang telah dilabel sebelumnya dengan hasil klasifikasi yang menggunakan model pembelajaran pada SVM.

Dari seluruh *tweet* yang diunduh yaitu 1020 dataset, maka akan dilakukan pelatihan dengan menggunakan *tools* pada RapidMiner. Model yang akan digunakan pada pengujian adalah model pembelajaran hasil dari *training* data latih. Karena klasifikasi pada penelitian ini bersifat *non linear* maka penelitian ini akan menggunakan kernel RBF dengan nilai parameter sebagai berikut:

- $C = 1 \leq C \leq 1000$
- $\gamma = 0.0001 \leq \gamma \leq 1000$
- validation = 10

pada penelitian ini akan dilakukan pengujian dengan data yang belum dilakukan pemilihan fitur dengan data yang sudah dilakukan pemilihan fitur. Pada data yang belum dilakukan pemilihan fitur nilai akurasi terbaik terdapat pada $c=2.5, \gamma=0.3$ dan $c=3, \gamma=0.3$ dengan nilai akurasi 97.54%. Untuk data yang sudah dilakukan pemilihan fitur nilai akurasi terbaik terdapat pada titik $c=20, \gamma=6$ dengan nilai akurasi 99.12%.

Kesimpulan

Penelitian ini berhasil dengan mendapatkan nilai akurasi yang tinggi dalam melakukan klasifikasi, sehingga dapat diterapkan agar memberikan bantuan kepada pengguna dalam mengelola *tweet*, terutama *tweet* iklan. Keberhasilan dengan mendapatkan nilai akurasi tertinggi 97,54% untuk data yang belum dilakukan pemilihan *feature*, sedangkan untuk data yang sudah dilakukan pemilihan terhadap *feature* mencapai nilai akurasi tertinggi 99.12%. Akurasi tertinggi

pada penelitian ini terdapat pada titik $c=2.5, \gamma=0.3$ dan $c=3, \gamma=0.3$ pada data yang belum dilakukan pemilihan *feature* sedangkan nilai akurasi tertinggi pada data yang sudah dilakukan pemilihan *feature* terdapat pada titik $c=20, \gamma=6$. Pengujian pada penelitian ini terjadi kesetabilan nilai akurasi pada rentang nilai $1 \leq c \leq 3$ dan $0.01 \leq \gamma \leq 10$ pada data yang belum dilakukan pemilihan *feature* sedangkan untuk data yang sudah dilakukan pemilihan *feature* terjadi kesetabilan nilai akurasi pada rentang nilai $1 \leq c \leq 300$ dan $0.01 \leq \gamma \leq 10$.

Ucapan terimakasih

Peneliti mengucapkan terima kasih yang tidak terhingga kepada para responden dilingkungan UIN Suska baik ditingkat rektorat, fakultas, jurusan maupun lembaga yang telah banyak membantu pelaksanaan penelitian ini. Masukan dari para reviewer jurnal SITEKIN juga sangat membantu memperbaiki kualitas jurnal ini. Terima kasih banyak.

Daftar Pustaka

- [1] Dennis Aprilla C, Donny Aji Baskoro, I Wayan Simri Wicaksana, Lia Ambarwati, Belajar Data Mining Dengan Rapid Miner: Jakarta. 2013.
- [2] Herlian, M. Machine Learning Teks Categorization. University Of Texas at Austin. 2006. (diakses pada tanggal 18/12/2013)
- [3] Kasali. Manajemen Periklanan dan Aplikasinya di Indonesia. 1992.
- [4] Liliana, Dewi Y, dkk. Indonesian News Classification Using Support Vector Machine. Word Academy Of Science, Engineering And Technology. 2011.

- [5] Nugroho, dkk. *Support Vector Machine – Teori Dan Aplikasinya Dalam Bioinformatika*. 2003.
- [6] Sulistyio Wiwin . Klasifikasi Dokumen Berbahasa Inggris Berdasarkan *Weighted-Trem*. *Jurnal Teknologi Informasi* . 2008. . (diakses pada tanggal 15/04/2014)
- [7] Wibisono, Yudi. Klasifikasi Berita Berbahasa Indonesia Menggunakan Metode *Naive Bayes Classifier*. *In Proceeding Of Seminar Nasional Matematika UPI Bandung*, 2005.
- [8] Yusra. Klasifikasi Minat Pengguna *Twitter* Berdasarkan *Tweet*, Menggunakan *Support Vector Machine*. Institut Teknologi Bandung. 2013.