JURNAL BORNEO SAINTEK Available online at www.ojs.borneo.ac.id
Volume 1, Nomor 3, Oktober 2018 Halaman 23-32
e-ISSN 2599-3313
p-ISSN 2615-434X

PERBANDINGAN IMPLEMENTASI KLASIFIKASI TEMA LITERASI DOSEN ILMIAH DAN MAHASISWA STMIK WIDYA CIPTA DHARMA DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MECHINE DAN K-MEANS CLUSTERING DENGAN DUKUNGAN VECTOR MECHINE

COMPARISON IMPLEMENTATION THEME LITERATION OF SCIENTIFIC LECTURERS AND STUDENTS OF STMIK WIDYA CIPTA DHARMA USING SUPPORT VECTOR MECHINE *(SVM) METHOD AND K-MEANS CLUSTERING WITH SUPPORT VECTOR MECHINE*

Andi Yusika Rangan[1], Mohammad Irwan Ukkas[2]

[1,2] STMIK Widya Cipta Dharma
Email: Andi@wicia.ac.id [1], irwanukkas212@gmail.com [2]

ABSTRAK

Klasifikasi tema karya tulis ilmiah adalah merupakan hal penting untuk mengetahui tren topik karya ilmiah yang dihasilkan. tren topik ini dapat digunakan untuk pengembangan kurikulum dan roadmap penelitian institusi. penelitian ini membandingkan metode clasifikasi pada teks mining yaitu using Support Vector Mechine (SVM) dan metode K-means Clustering dengan Dukungan Vector Mechine pada proses textmining. Dari hasil penelitian menyimpulkan tingkat akurasi dari Support Vector Mechine (SVM) sebesar 93.33% sedangkan metode K-means Clustering dengan Dukungan Vector Mechine (SVM) tingka akurasi yang dihasilkan 99,33%

Kata Kunci: SVM, K-means clustering, textmining

ABSTRACT

Classification of the theme of writing is an important thing to know the trends of the topic of scientific work produced. This topic trend can be used for curriculum and research roadmap. This research uses clasification method in mining text that is using Support Vector Mechine (SVM) and K-means Clustering method with Vector Mechine Support for clasification on textmining. The results of the final level of Vector Mechine Support (SVM) were 93.33% while the K-means Clustering method with Vector Mechine Support (SVM) resulted in 99.33% accuracy rate.

Keywords*: *SVM, K-means clustering, textmining

PENDAHULUAN

Pengetahuan mengenai tren karya ilmiah yang dihasilkan oleh perguruan tinggi memberikan manfaat bagi pengembangan kurikulum dan roadmap penelitian bagi institusi. Berbagai karya ilmiah dari sivitas akademika seperti skripsi, laporan penelitian, laporan penulisan ilmiah, laporan kuliah kerja nyata yang tersimpan secara digital namun, pada umumnya fenomena ini tidak disertakan pengetahuan yang disarikan dari dokumen-dokumen elektronika tersebut (Gupta, 2011). Metode *text mining* merupakan pengembangan dari *data mining* yang

diterapkan untuk mengatasi masalah tersebut. Algoritma-algoritma dalam text mining di buat untuk mengenali data yang sifatnya semi terstuktur misalnya sinopsis, abstrak maupun isi dari dokumen-dokumen (Gupta & Lehal, 2009). Kategorisasi teks dapat digunakan untuk melakuakn pengalian opini (*opinion mining)* dan analisa sentiment. Algoritma katagorisasi teks saat in banyak berkembang antara lain *support Vector Machine (*SVM), *Naïve Bayes*, C4.5, K-Nearest Neighbours (K-NN) dan lain-lain.

K-means clustering merupakan metode yang popular digunakan untuk mendapatkan deskripsi dari sekumpulan data dengan cara mengungkapkan kecenderungan setiap individu data untuk berkelompok dengan individu data lainya. Pada penelitian ini K-means clustering digunaan untuk memperbaiki proses kalsifikasi data teks yaitu melakuan klasterisasi data agar tingkat akurasi model yang diusulkan menjadi lebih baik.

Algoritma *Support Vector Machine* (SVM) digunkan untuk klasifikasi penentuan jenis tema karya ilmiah dosen dan mahasiswa. SVM adalah metode yang banyak digunakan untuk klasifikasi data berupa teks dengan tingkat akurasi yang baik.

Penelitian penerapan algoritma K-means Clustering untuk Optimasi Klasifikasi Tema karya ilmiah dosen danMahasiswa mengunakan *Support Vector Machine (SVM)* Pada STMIK Widya Cipta Dharma bertujuan untuk proses penemuan pola pengelompokan berbagai topik tugas akhir mahasiswa yang bermanfaat menghasilkan informasi tren penelitian perguruan tinggi dari tahun ke tahun.

METODE PENELITIAN

Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan.Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning.*

Text mining

Text mining adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen (Han & Kamber, 2006). Prosedur utama dalam metode ini terkait dengan menemukan kata-kata yang dapat mewakili isi dari dokumen untuk selanjutnya dilakukan analisis keterhubungan antar dokumen dengan menggunakan metode statistik tertentu seperti analisis kelompok, klasifikasi dan asosiasi. Tahapan dalam *text mining* secara umum adalah *tokenizing, filtering, stemming, tagging,* dan *analyzing* (Michael, 2004)

Tokenizing merupakan tahapan untuk memisah-misahkan setiap kata (*token*) pada dokumen *input*. *Filtering* merupakan proses seleksi terhadap kata-kata yang dihasilkan dari proses *tokenizing*, dapat dilakukan dengan algoritma *stop list* maupun *word list*. Algoritma *stop list* akan membuang kata-kata yang tidak penting seperti kata ganti, kata keterangan, kata sambung, kata depan dan kata sandang. Sebaliknya, algoritma *word list* akan menyimpan kata-kata yang penting.

Proses *stemming* kemudian dilakukan untuk mencari kata dasar dari setiap kata yang telah lolos proses *filtering*. Terdapat 4 varian algoritma untuk proses *stemming* ini, yaitu: (1) *Table lookup, s*eluruh kata

dasar disimpan dalam memori untuk selanjutnya dijadikan acuan dalam pemeriksaan dokumen *input*. Kelemahan metode ini adalah membutuhkan ruang penyimpanan yang besar; (2) *Successor variety,* setiap kata dalam dokumen *input* yang akan diperiksa dipecah secara bertahap menjadi awalan-awalan (prefiks). Untuk setiap awalan kemudian dicari kemungkinan bentuk lainnya (variasinya) didalam *corpus*, pencarian dihentikan jika jumlah temuan telah melampaui nilai batas tertentu; (3) *N-gram*, pemeriksaan setiap kata dalam dokumen *input* dilakukan dengan menerapkan konsep *clustering*. Setiap kata dicari nilai kedekatannya dengan kata-kata yang lain dan disimpan dalam sebuah matriks. Matriks tersebut kemudian dijadikan acuan untuk melakukan pengelompokan kata-kata; (4) *Affix removal*, untuk setiap kata pada dokumen *input* dihilangkan awalan dan akhirannya dengan mengacu kepada *action rules*.

Proses *tagging* dilakukan untuk mencari bentuk awal dari setiap kata lampau. Setelah semua kata penting berhasil dikoleksi dari rangkaian proses tersebut, maka tahap berikutnya adalah *analyzing* yaitu menentukan keterhubungan antar dokumen dengan mengamati frekuensi kemunculan tiap kata yang ada pada tiap dokumen.

K-means Clustring

K-Means Clustering merupakan metode yang popular digunakan untuk mendapatkan dekripsi sekumpulan data dengan cara mengungkapakan kecenderungan setiap data untuk berkelompok dengan individu-individu data lainya. Kecenderungan penegelompokan tersebut didasarkan pada kemiripan karakteristik individu-individu data yang ada. Ide dasar dari teknik ini adalah menemukan pusat dari setiap kelompok data yang mungkin ada untuk kemudian mengelompokkan setiap data individul kedalam salah satu dari kelompok-kelompok tersebut berdasarkan jaraknya (Turban dkk, 2005).

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode klasifikasi yang bekerja dengan cara mencari *hyperplane* dengan margin terbesar *Hyperplane* adalah garis batas pemisah data antar-kelas. Margin adalah jarak antara *hyperplane* dengan data terdekat pada masing-masing kelas. Adapun data terdekat dengan *hyperplane* pada masing-masing kelas inilah yang disebut *support vector* (J. Yunliang, et al., 2010). Pada dasarnya, SVM merupakan metode yang digunakan untuk klasifikasi dua kelas (*binary classification*). Pada perkembangannya, beberapa metode diusulkan agar SVM bisa digunakan untuk klasifikasi *multi-class* dengan cara mengombinasikan beberapa *binary classifier* (J.Z.Liang, 2004).

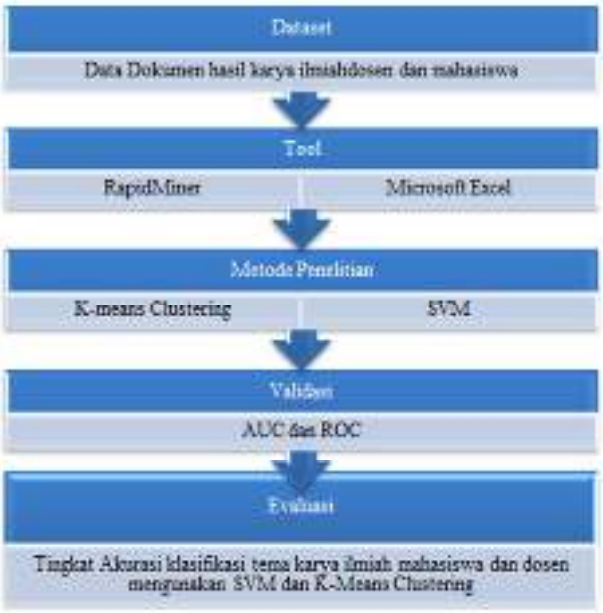
Lokasi Penelitian

Penelitian dilakukan di STMIK Widya Cipta Dharma, Subjek dalam penelitian ini tema dari karya ilmiah di STMIK Widya Cipta Dharma. Sedangkan objek dalam penelitian ini adalah Penerapan Metode K-Means Clustering untuk Optimasi Klasifikasi Tema Karya Ilmiah Dosen dan Mahasiswa mengunakan Support Vector Machine (SVM) Pada STMIK Widya Cipta Dharma.

Tahapan Penelitian

Penelitian ini akan dilaksanaman melalui beberapa tahapan yaitu :

1. Menentukan dataset sebagai sumber data penelitian yaitu data tema karya ilmiah dosen dan mahasiswa tiga tahun dari tahun 2014-2016, sebanyak 150 Karya yang terdiri dari berbagai tema hasil karya ilmiah. Tool (Software) yang akan digunakan dalam penelitian ini adalah RapidManer dan sebagai pendukung pengolahan data mengunakan Microsoft Excel 2010.
2. **Praprocessing Data**
   Tahap awal sebelum melakukan proses pengelompokan dokumen adalah mempersiapkan teks yang ada didalam dokumen. Pada tahap praproses ini dilakukan beberapa subproses agar dokumen dapat dipakai untuk melakukan proses pengelompokan. Subproses diantaranya yaitu:
   a. *Tokenizer*, yakni proses yang bertujuan untuk memisah teks menjadi beberapa *token* berdasarkan pembatas berupa spasi atau tanda baca.
   b. Proses selanjutnya adalah menghilangkan teks yang bersesuaian dengan teks yang terdapat pada daftar *stopword*, karena teks tersebut dianggap tidak dapat mewakili konten dokumen.
   c. Kemudian pada teks yang masih tersisa dilakukan proses *stemming*, yaitu proses pengubahan teks menjadi bentuk dasarnya.
   d. Selanjutnya, setiap kata tersebut disebut sebagai *term*. Nantinya setiap *term* akan didaftar dan diberi bobot.
   e. Pembobotan masing-masing term dilakukan dengan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). TF-IDF merupakan metode pembobotan *term* dengan menggunakan *termfrequency* (jumlah *term* yang terdapat pada tiap dokumen) serta *inverse document frequency* (*invers* jumlah dokumen yang memuat suatu *term*).
3. **Pengelompokan Dokumen**
   Dari *k* model klasifikasi yang telah ada, maka dapat dilakukan klasifikasi dokumen baru. Pengujian dilakukan dengan mengelompokkan dokumen baru kedalam kelompok yang ada menggunakan tetangga terdekat dari *centroid* pada masing-masing kelompok. Setelah didapatkan kelompok yang sesuai maka dilakukan proses klasifikasi dokumen baru dengan model *SVM* pada kelompok yang bersangkutan.
4. **Penentuan data Training dan data testing**
   Data *training* dan *testing* dalam penelitian ini diambil dari judul tugas akhir mahasiswa program studi S1 Sistem Informasi dan Teknik Informatika serta D3 Manajemen Informatika STMIK Widya Cipta Dharma, dimana setelah dijumlahkan akan di split menjadi 70% data *training* dan 30% data *testing*
5. **Eksperimen dan Pengujian**
   Dari *k* model klasifikasi yang telah ada, maka dapat dilakukan klasifikasi dokumen baru. Pengujian dilakukan dengan mengelompokkan dokumen baru kedalam kelompok yang ada menggunakan tetangga terdekat dari *centroid* pada masing-masing kelompok. Setelah didapatkan kelompok yang sesuai maka dilakukan proses klasifikasi dokumen baru dengan model *SVM* pada kelompok yang bersangkutan
6. **Evaluasi dan Validasi**
   Pada penelitian ini sebagai evaluasi dari model yang diusulkan, yaitu dengan menggunakan metode *Cross validations* untuk mencari nilai akurasi yang kemudian hasil dari akurasi tersebut dievaluasi dengan cara membandingkan tingkat akurasi yang dihasilkan oleh model SVM dengan menggunakan *K-Means* dan dengan model SVM tanpa *K-Means* seperti pada gambar 1 Alur Kerja Penelitian



Gambar 1. Alur Kerja Penelitian

HASIL DAN PEMBAHASAN

Hasil penelitian ini menguji 150 judul karya ilmiah dosen dan mahasiswa STMIK Widya Cipta Dharma dengan tahun kegiatan 2014-2016. Dengan menerapakan metode SVM (Support Vector Mechine) serta optimasi dengan mengabungkan metode K-Means Clustering dan SVM. Dari 150 data tersebut dibuatlah data training sebannyak 70 % (105) data sebagai data training dan 30 % (45) data sebagai data training.

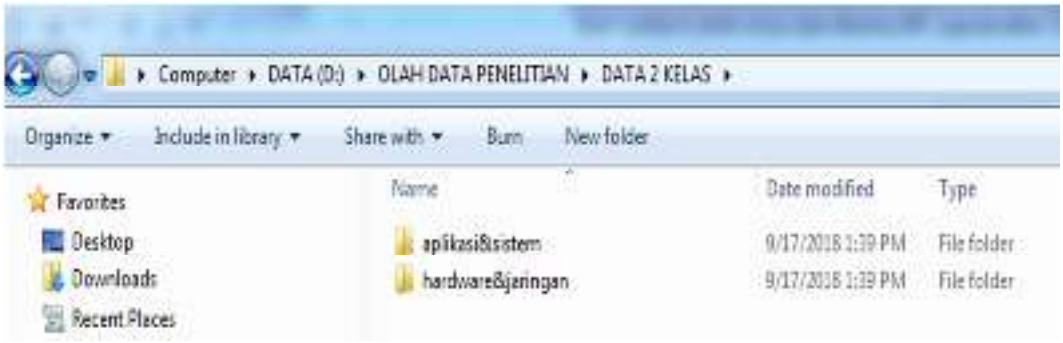
Tabel 1. Daftar Judul Penelitian

No	Tahun	Judul	Kategori
1	2016	application library on sman 1 sebulu	Aplikasi & Sistem
2	2016	application registration of new learners on smp al azhar syifa budi samarinda	Aplikasi & Sistem
3	2016	implementation of augmented reality home marketing pt. rika sakti brothers use the method of marker based tracking on the brochure housing	Aplikasi & Sistem
4	2016	learning to build augmented reality space with the android based marker based tracking method	Aplikasi & Sistem
5	2016	implementation of markerless augmented reality in learning of android based hijaiyah letters	Aplikasi & Sistem
6	2016	augmented reality-based solar system learning with the method of marker based tracking	Aplikasi & Sistem
7	2016	e-learning web-based design of case studies on brotherhood introspective Indonesia samarinda	Aplikasi & Sistem
8	2016	build edugame introduction of human digestive system with the logic of soa game randomization and development of game agent based on finite state	Aplikasi & Sistem

		machine	
9	2016	build edugame "baby zoo puzzle" based on android with game agent implementation finite state machine	Aplikasi & Sistem
10	2016	data security techniques with the stegangrafi method of end of file (eof) and vernam cipher cryptography	hardware&jaringan
…..	……	……	……
150	2016	implementation of home light control using arduino uno based	hardware&jaringan

a. Pra Procesing data

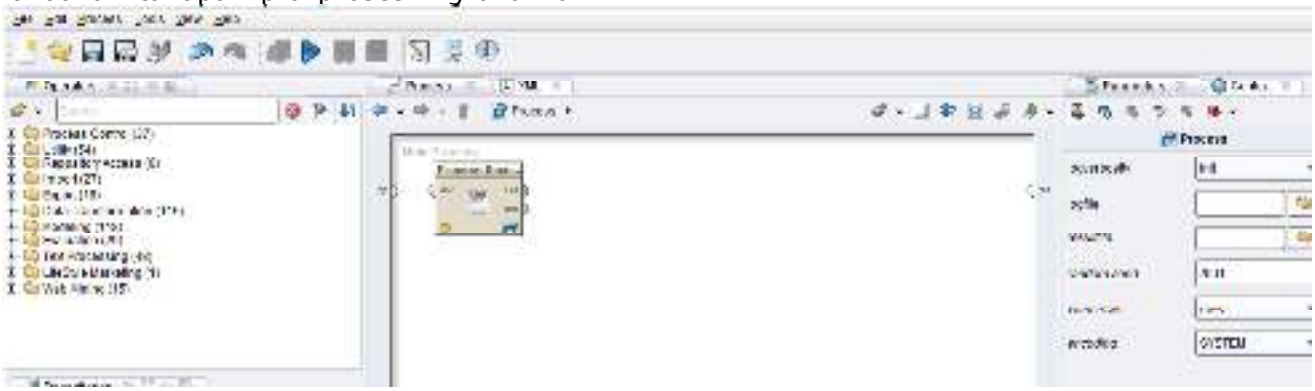
sebelum data set diolah dan diujicoba dega mengunakan metode yang dipilih terlebih dahulu dataset dirubah kedalam bentuk file bertipe .txt. Hal ini akan membantu memudahkan tools Rapid Miner dalam membaca dan mengolah data. Dataset yang sudah dirubah kedalam bentuk file .txt, dikelompokan penyimpanannya kedalam folder-folder yang penamaanya disesuaikan denga kategori masing-masing judul.



Gambar 2. Pra Procesing Data

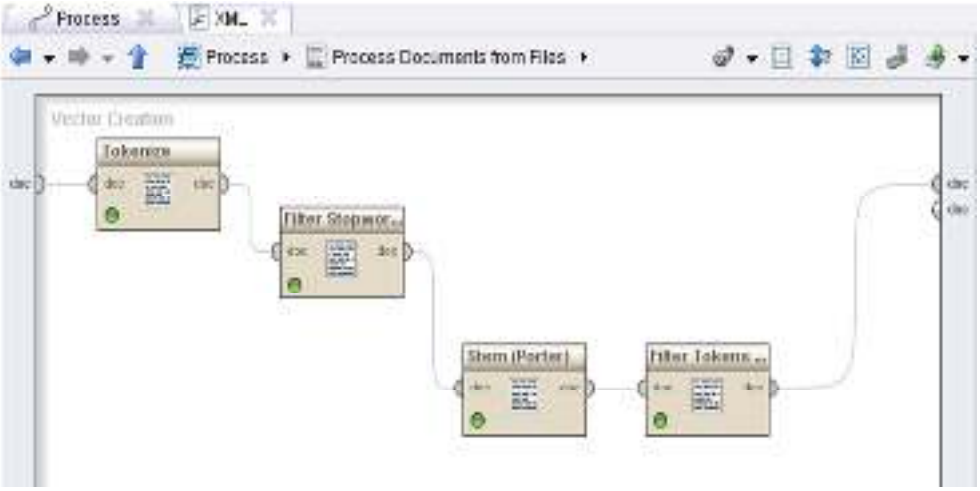
b. Procesing Data

Pada tahapan ini dataset yang sudah di lakuakan tahapan pra prosesning akan di load pada tool rapi miner, hal ini terlihat pada gambar 3 Loading data set.



Gambar 3. Loading Dataset

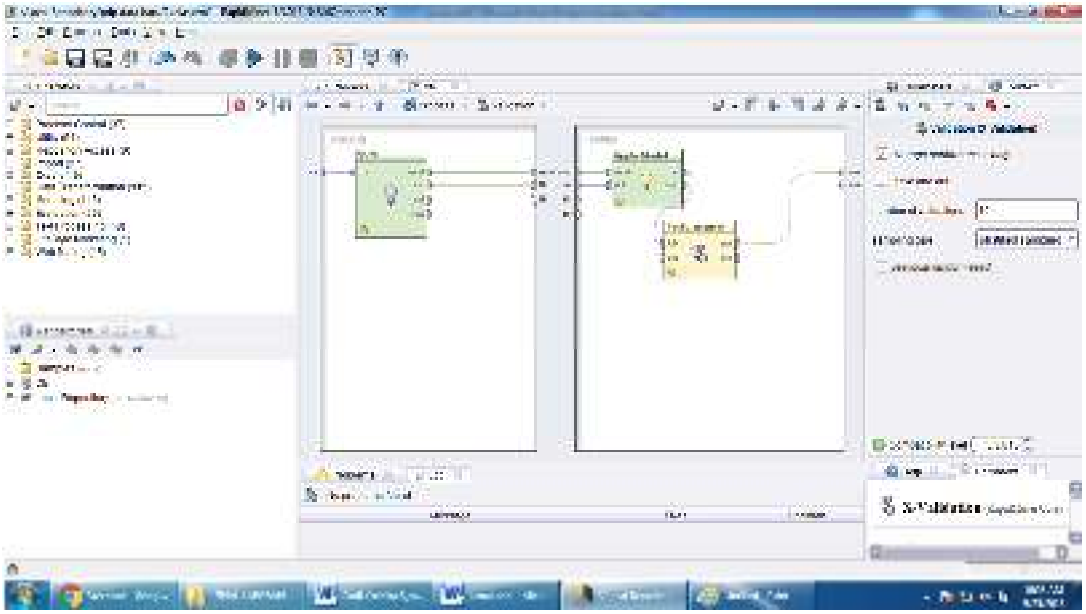
Kemudian data yang sudah diloading dilakuakn tahapan text processing seperti token, steeming filter yang terlihat pada gambar 4 text processing.



Gambar 4. text Processing

c. Clastering data dengan metode SVM

Selajutnya mengolah data denga *tool* Rapid Miner 5 dengan mengubungkan data yang sudah diolah dengan operator *cross validation* sehingga didapat hasil perhitungan model dengan tingkat akurasinya seperti pada gambar 5 Metode SVM pada rapid Miner.

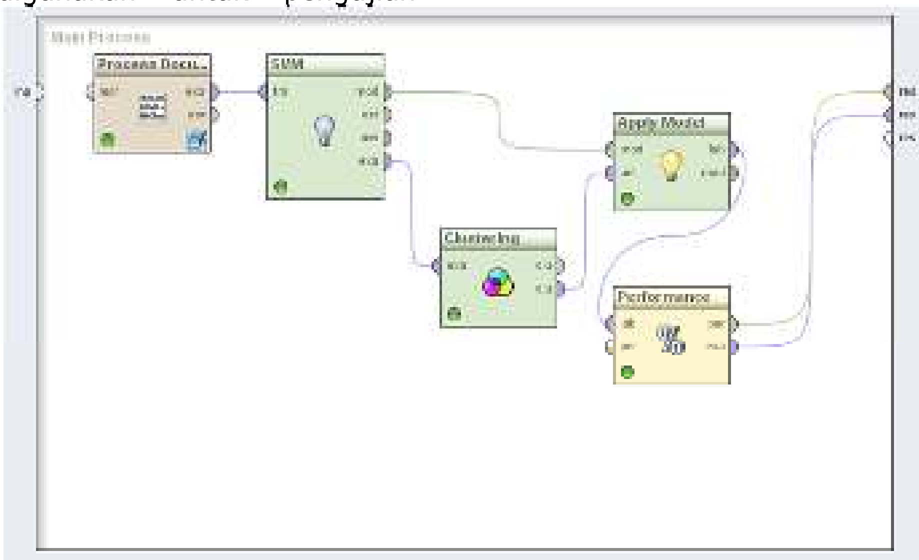


Gambar 5. Metode SVM Pada Rapid Miner

d. Clastring data K-Means dan SVM

Pengolahan data dilakukan dengan adanya tambahan proses yaitu pengelompokan data sebelumnya K-means Clustering kemudian dilanjutkan dengan pengolahan data

mengunakan metode SVM. Metode yang digunakan untuk pengujian seperi pada gambar dibawah ini.



Gambar 6. Pengolah Data Dengan Metode K-Means Cluster dan SVM

e. Hasil Akurasi

Diantara mekanisme yang dapat dilakukan untuk mengukur validitas hasil klasifikasi adalah dengan menghitung nilai prediction dan recall. Perhitungan nilai precision akan mengukur tingkat kepastian (exactness) atau jumlah data testing yang diklasifikasikan dengan benar oleh model klasifikasi yang dibangun Perhitungan recall merupakan kebalikan dari precision. Recall mengukur sensitifitas atau rasio dari data untuk setiap label yang diklasifikasikan dengan benar terhadap data yang salah diklasifikasikan ke label lainnya (missclassified). Pada masing-masing hasil evaluasi dapat dilihat pada gambar berikut:

Multiclass Classification Performance / Annotations

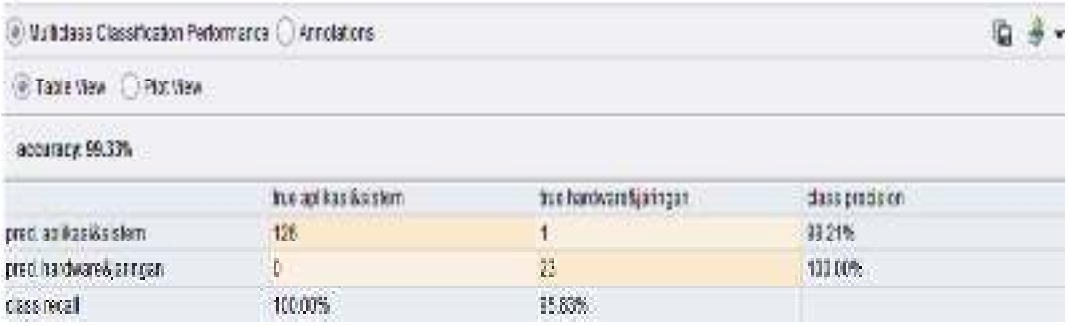
Table View / Plot View

accuracy: 93.33% +/- 5.18% (mikro: 93.33%)

	true aplikasi&sistem	true hardware&jaringan	class precision
pred. aplikasi&sistem	126	10	92.65%
pred. hardware&jaringan	0	14	100.00%
class recall	100.00%	58.33%	

Gambar 7. Hasil Evaluasi Metode SVM

Pada Gambar 7 Hasil Evaluasi menunjukan hasil akurasi model adalah 93.33% dengan class Prediksi untuk Hardware&Jaringan yang mencapai 100% tetapi class recallnya 58.33%. Class prediksi untuk aplikasi&Sistem mempunyai nilai 92,65% dengan class recall sebesar 100%.



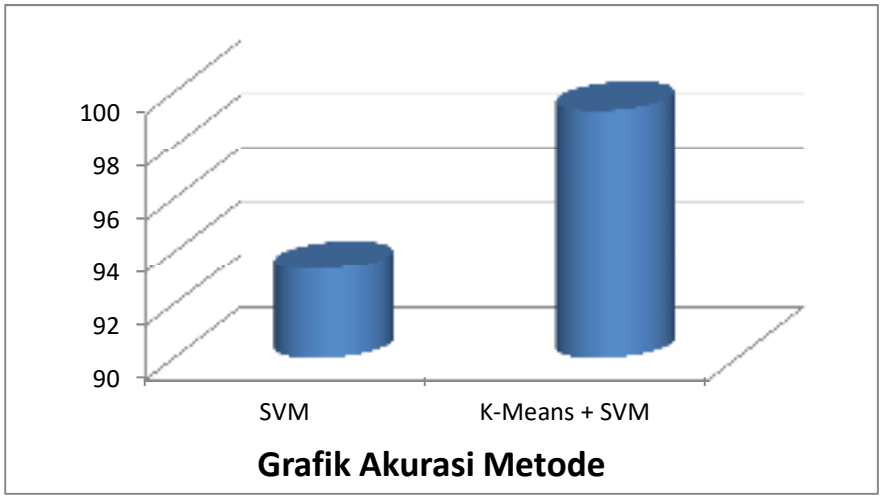
Gambar 8. Hasil Evaluasi K-Means Clusting dan SVM

Pada gambar 5.8 hasil evaluasi menunjukan hasil akurasinya sebesar 99,33% dengan rincian class prediksi untuk hardware&jaringan sebesar 100%. dan untuk aplikasi&system sebesar 99,21%. Class recall untuk hardware&jaringan sebesar 95,83% dan untuk apliksi dan jaringan sebesar 100%.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan hasil evaluasinya dapat digambarkan dalam bentuk grafik, hasil akurasi metode SVM menghasilkan tingkat akurasi 93,33 % dan hasil akurasi k-means clustering dan SVM sebesar 99,33%.

Grafik 1. Akurasi metode yang digunakan



UCAPAN TERIMA KASIH

Terimakasih peneliti ucapkan atas hibah penelitian yang diberikan dari KemenristekDikti anggaran tahun 2018

DAFTAR PUSTAKA

Prilianti,R. dan Wijaya,H. 2014, *Aplikasi Text Mining Untuk Autmasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering*, jurnal cybermatika vol 2 no 1 juni 2014 artikel 1.

Andini,S.2013,*Klasifikasi dokumen teks mengunakan algoritma naive bayes dengan bahasa pemrograma java,* jurnal teknologi informasi & pendidikan vol 6 no 2 september 2103 issn 2086-4981 hal 141-147.

Somantri,O. Wiyono,S. Dairoh., 2014, *Metode K-Means Untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Mengunakan Supprot Vector Machine (SVM)*. scientific journal of informatic vol 3 no 1 mei 2016 Hal 34-45.

Hidayatullah, F. Ma'arif, M.R.2016. *Penerapan Text Mining Mining Dalan Klasifikasi Judul Skripsi*. Seminar Nasional Aplikassi Teknologi Informasi (SNATI) 6 agustus 2016 a33-a36.

Indranandita, A. Santoso,B. Rahmat, A. 2008, .*Sistem Klasisfikasi DanPencarian Jurnaldengan Metode Naive Bayes Dan Vector Space Model*, jurnal informatika volume 4 nomor 2.

Ariadi, Fithriansari, 2015, *Klasifikasi Berita Indonesia Mengunakan Metode Naïve Bayesias Classification dan Support Vector Machine dengan Confix Stripping Stemmer* Jurnal Sains Dan Seni ITS Vol. 4, No.2 Halaman D248 – D253.