ANALISA DAN KLASIFIKASI SENTIMENT OPINI PENONTON PADA WEBSITE IMDB.COM DENGAN ALGORITMA SUPPORT VECTOR MACHINE

Cindy Claudia [1] dan Akhmad Budi [2]
[1] Alumni Program Studi Sistem Informasi
[2] Staf Pengajar Program Studi Teknik Informatika
Institut Bisnis dan Informatika Kwik Kian Gie
Jl. Yos Sudarso Kav.87 Sunter Jakarta Utara 14350
http://www.kwikkiangie.ac.id
akhmad.budi@kwikkiangie.ac.id

ABSTRACT

In its development, information technology brought about many changes and advances in the context of everyday life. One of the benefits of information technology to process and process large amounts of data is data mining and text mining. In this study the authors will analyze and discuss the sentiment (emotion) contained in a comment or post on reviews given by the audience on a movie and see whether it will be classified as positive sentiment or negative so that it can be used to classify the reviews on different movies at imdb.com in the future. This research will gain some classification models based on modeling with support vector machine algorithm to classify the opinion of the new movie and the level of accuracy of each models. Also there will be a wordlist data that contains details of each positive and negative comments in the classification process. Results of this study indicate that it is possible to create a classification model based on audience opinions to classify new opinion using support vector machine algorithms with a high degree of accuracy and produce the details of words in the wordlist database to find a word or phrase that affects most to determine the sentiment in an opinion.

Keywords : Classification, Opinion, Audience, Movie, Website, Data Mining, Text Mining, Sentiment Analysis, Support Vector Machine.

1. PENDAHULUAN

Perkembangan teknologi di zaman sekarang bisa dibilang sangat pesat dan dirasakan di berbagai bidang dalam kehidupan manusia. Dalam perkembangannya teknologi informasi membawa banyak perubahan dan kemajuan dalam konteks kehidupan sehari-hari. Setiap hari manusia melakukan berbagai kegiatan dan aktivitas yang beragam, kegiatan ini menghasilkan sebuah fakta-fakta yang bersangkutan akan seseorang atau tindakan tertentu yang dapat diolah kembali dan diterjemahkan dalam bentuk data. Data-data tersebut kemudian dapat diolah dalam sebuah sistem untuk membantu pengambilan keputusan dalam menyelesaikan masalah yang dihadapi. Kebutuhan data yang besar harus diimbangi dengan sistem yang mampu memproses data dalam jumlah yang besar sehingga informasi yang diperoleh dari data-data tersebut dapat dianalisa secara efektif untuk digunakan kembali.

Kebutuhan manusia dalam kehidupan sehari-hari tidak lepas dari hiburan baik itu secara langsung maupun tidak langsung. Hiburan yang didapat bisa bermacam-macam tetapi tidak dapat dipisahkan dari hiburan yang bersifat visual seperti menonton. Teknologi visual dan digital yang semakin berkembang diiringi dengan internet semakin membuat manusia mudah mendapatkan hiburan tontonan dari berbagai sumber seperi film layar lebar. film serial tv, maupun video-video pada situs web berbagi video gratis yang jumlahnya sangat banyak tersedia di dunia maya dan mudah untuk dilihat dari berbagai pemutar media baik pada *smartphone*, televisi, maupun media lainnya.

Setiap orang memiliki persepsi dan pandangan yang berbeda setiap kali menonton

sebuah video atau film, sehingga opini yang dihasilkan dari sebuah tontonan bisa berbeda. Opini tersebut merupakan sebuah penilaian dari individu akan sebuah tontonan baik itu bersifat positif maupun negatif. Opini-opini tersebut dapat dengan mudah ditemukan di berbagai situs berbagi video pada *comment box* yang tersedia ataupun pada situs-situs penyedia informasi dan *review* filmografi seperti imdb.com atau rottentomatoes.com. Opini dari tiap-tiap penonton dapat dilihat sebagai referensi akan kepuasan dalam menonton film tersebut, rujukan untuk calon penonton, maupun sebagai umpan balik kepada tim produksi dan pembuat film sebagai masukan untuk film selanjutnya. Karena hal tersebut maka data opini yang dihasilkan dari berbagai film atau tontonan memiliki jumlah sangat banyak dan beragam.

Dalam pengolahan data dalam jumlah yang besar sistem yang dimiliki juga dituntut untuk dapat beradaptasi dan mampu berbagai menyelesaikan masalah-masalah di masa depan. Salah satu manfaat teknologi informasi yang berguna untuk memproses dan mengolah data dalam jumlah besar adalah *data mining*. *Data mining* merupakan proses untuk mengekstraksi data dalam jumlah besar untuk mencari sebuah pola atau informasi yang berguna dari data tersebut untuk digunakan kembali dalam sistem atau untuk membantu pengambilan keputusan.

Dalam *data mining* ada sebuah teknik lanjutan yang disebut dengan *text mining* yaitu teknik yang dikhususkan untuk memproses data yang tidak terstruktur terutama data yang berbentuk *text.* Pada penerapannya *text mining* dikhususkan untuk memproses data dengan bentuk yang tidak terstruktur (*unstructurized data)* seperti dokumen atau teks yang tidak termasuk dalam tabel atau memiliki bentuk yang tersusun sehingga dalam pengaplikasiannya *text mining* memiliki modul tersendiri untuk mendukung proses ekstraksi data dalam *data mining*. Karena alasan tersebut maka penulis mencoba mengambil tema aplikasi *text mining* untuk menganalisa dan mengklasifikasi sentiment penonton terhadap komentar dan review penonton pada website imdb.com dengan algoritma *support vector machine*.

Dalam penelitian ini penulis akan menganalisa dan membahas *sentiment* (emosi) yang terdapat dalam komentar atau post tentang ulasan yang diberikan oleh para penonton akan sebuah film dan melihat apakah layak diklasifikasikan sebagai *sentiment* yang positif atau negatif sehingga dapat digunakan untuk mengklasifikasikan ulasan pada film yang berbeda pada imdb.com di masa depan. Data ulasan yang digunakan oleh penulis di peroleh dari http://ai.stanford.edu/~amaas/data/sentiment/ dan http://www.imdb.com yang berisikan komentar dan ulasan penonton terhadap berbagai film untuk dijadikan sampel dalam proses klasifikasi yang akan dilakukan pada penelitian ini.

Website imdb.com adalah situs yang berisi tentang informasi filmografi terbesar di dunia yang berisikan informasi dan ulasan tentang film, acara TV, dan berbagai bidang sinematografi lainnya lengkap dengan data-data pemain dan informasi pendukungnya. Pada kesempatan kali ini penulis menggunakan data dari website tersebut untuk melakukan penelitian tentang *sentiment* dan ulasan dari berbagai pihak untuk mengukur dan menentukan pola klasifikasi dari data penilaian tersebut dengan menggunakan algoritma *support vector machine*.

Penelitian yang dilakukan difokuskan kepada *sentiment analysis* (emosi) yang terdapat pada setiap komentar dalam suatu ulasan film. Jenis sentiment yang dianalisa antara lain adalah opini positif dan negatif yang terdapat dalam data yang diperoleh. Data opini yang diperoleh kemudian akan diolah dalam modul *text mining* untuk dicek keakuratan nya dan di klasifikasikan sebagai bentuk yang dapat diolah kembali dengan algoritma *support vector machine*. Algoritma tersebut diperuntukkan untuk mengukur keakurasian opini yang diklasifikasikan secara manual serta memperoleh pola atau metode hasil klasifikasi *sentiment* dari berbagai film yang diperoleh secara acak untuk digunakan kembali dalam mengukur *sentiment* opini dari film atau acara TV lainnya di masa depan dan dapat digunakan untuk metode penilaian lebih lanjut. Penelitian ini menggunakan konsep *sentiment analysis* atau biasa disebut dengan *opinion mining* yaitu konsep yang dilakukan dalam untuk mengambil dan mengolah data opini yang diperoleh menggunakan modul *text mining* pada rapidminer untuk menemukan pola data yang dibutuhkan serta hasilnya akan diterapkan dalam

aplikasi yang dibuat dalam microsoft access untuk digunakan oleh pembaca.

Pada prakteknya tentu saja dalam menentukan klasifikasi opini pada *sentiment analysis* ditemukan adanya beberapa kendala atau masalah yang muncul baik dalam pengolahan data maupun klasifikasi yang akan dilakukan, yaitu seperti sulitnya memproses data opini yang tidak terstruktur untuk *sentiment analysis* sehingga memakan banyak waktu jika dilakukan dengan metode pengolahan data biasa, ini disebabkan karena data yang dimiliki masih berbentuk dokumen yang tidak tersusun sehingga pengolahan data opini tidak mampu dilakukan tanpa software *data mining*. Selain itu tingkat akurasi klasifikasi opini yang dilakukan secara manual belum dapat dibuktikan karena masih tergantung *bias* (prasangka) individu yang bersangkutan. Karena klasifikasi masi menggunakan opini pribadi maka belum diterapkannya teknik klasifikasi yang sesuai untuk memproses opini penonton secara berkala di masa depan selain metode manual. Masalah tersebut timbul karena teknik klasifikasi yang digunakan masih belum dilakukan secara sistematis sehingga hasilnya dapat dengan mudah dimanipulasi kebenaranya karena tidak menggunakan algoritma klasifikasi sehingga data yang diperoleh masih berbentuk acak atau tidak terstruktur.

2. LANDASAN TEORI

2.1 Data

Data [1] adalah aliran fakta mentah yang merepresentasikan kejadian yang terjadi pada suatu organisasi atau lingkungan fisik sebelum diorganisir dan disusun menjadi bentuk yang dapat dimengerti dan digunakan oleh seseorang.

2.2. Informasi

Informasi [1] adalah kumpulan fakta terorganisir dan terolah sehingga mereka memiliki nilai tambahan di luar nilai fakta individu.

2.3 Sistem Informasi

Sistem informasi dapat didefinisikan secara teknis sebagai satu set komponen yang mengumpulkan (atau mengambil), memproses, menyimpan, dan mendistribusikan informasi untuk mendukung pengambilan keputusan dan pengendalian dalam suatu organisasi.

2.4. Opini

Opini [9] adalah sebuah sentimen, ekspresi, entitas dari suatu pendapat dan ekspresi terhadap sesuatu seperti produk, jasa, individu, atau organisasi.

2.5 *Database*

D*atabase* [3] adalah koleksi terpadu dari informasi yang secara logis terkait dan disimpan sedemikian rupa untuk Meminimalkan duplikasi dan memfasilitasi pengambilan secara cepat. [1]

2.6. *Data Warehouse*

D*ata warehouse* [15] adalah repositori data sentral yang berisi informasi yang diambil dari berbagai sumber yang dapat digunakan untuk analisis, pengumpulan intelijen, dan perencanaan strategis. Dalam teorinya data-data yang terdapat *data warehouse* dapat diperoleh dari 2 macam sumber yaitu :

1. Sumber Internal (*Internal Data Sources)* : Sumber data yang didapatkan dari dalam perusahaan atau organisasi. Sumber data ini dapat berupa data pelanggan, transaksi, inventori, aset dan hutang, SDM, dll.

2. Sumber External (External Data Sources) : Sumber data yang didapatkan dari informasi di luar perusahaan atau organisasi yang mampu membantu proses bisnis (memiliki nilai).

2.7. *Database Management Systems (DBMS)*

Sistem manajemen database (DBMS) [15] adalah sistem yang terdiri dari kumpulan data yang saling berhubungan, yang dikenal sebagai database, dan satu set program perangkat lunak untuk mengelola dan mengakses data. Program perangkat lunak ini menyediakan mekanisme untuk mendefinisikan struktur database dan penyimpanan data; untuk menentukan dan mengelola akses data bersama, bersamaan, atau

terdistribusi; dan untuk memastikan konsistensi dan keamanan informasi yang tersimpan meskipun sistem *crash* atau upaya akses yang tidak sah.

2.8. *Data Mining*

Data mining [3] adalah jenis pengumpulan intelijen yang menggunakan teknik statistik untuk mengeksplorasi set data yang besar, berburu pola tersembunyi dan hubungan yang tidak terdeteksi dalam laporan rutin

2.9. Tahap-Tahap *Data Mining*

Menurut para ahli ada beberapa tahapan dalam melakukan proses *data mining* yaitu antara lain [11] :

1. Data collection (Pengumpulan data) : Tahap dimana data yang dibutuhkan untuk penelitian dikumpulkan sesuai dengan kebutuhan. Pengumpulan data mungkin memerlukan penggunaan hardware khusus seperti jaringan sensor, tenaga kerja manual seperti koleksi survei pengguna, atau perangkat lunak seperti mesin dokumen perangakak web (*web crawling)* untuk mengumpulkan dokumen. Sementara tahap ini sangat-aplikasi spesifik dan sering di luar bidang analis *data mining*, menyebabkan tahap ini sangat penting karena pilihan yang baik pada tahap ini secara signifikan dapat mempengaruhi proses *data mining*. Setelah tahap pengumpulan, data sering disimpan dalam *database*, atau, lebih umum, sebuah gudang data untuk diproses.

Feature extraction and data cleaning (Ekstraksi fitur dan pembersihan data) : Tahap dimana data di ekstraksi, seleksi, dan dibersihkan untuk digunakan dalam pemodelan. Ketika data dikumpulkan, mereka sering tidak dalam bentuk yang cocok untuk diproses. Sebagai contoh, data dapat dikodekan dalam log kompleks atau dokumen-bentuk bebas. Dalam banyak kasus,

jenis data yang berbeda dapat sewenang-wenang dicampur bersama dalam dokumen-bentuk bebas. Untuk membuat data yang sesuai untuk pengolahan, penting untuk mengubah mereka menjadi format yang ramah untuk algoritma *data mining*, seperti multidimensi, *time series*, atau format semi terstruktur. Format multidimensi adalah salah satu yang paling umum, di mana berbagai bidang data sesuai dengan sifat yang diukur berbeda yang disebut sebagai fitur, atribut, atau dimensi. Hal ini penting untuk mengekstrak fitur yang relevan untuk proses penambangan. Tahap ekstraksi fitur sering dilakukan secara paralel pembersihan data, di mana data hilang dan bagian yang salah dari data yang baik diperkirakan atau diperbaiki. Dalam banyak kasus, data dapat diekstraksi dari berbagai sumber dan perlu diintegrasikan ke dalam format terpadu untuk diproses. Hasil akhir dari prosedur ini adalah kumpulan data terstruktur dengan baik, yang dapat secara efektif digunakan oleh program komputer. Setelah tahap ekstraksi fitur, data dapat lagi disimpan dalam *database* untuk diproses.

Analytical processing and algorithm : Proses analisa dan aplikasi algoritma pada data yang telah dibersihkan. Bagian akhir dari proses penambangan adalah untuk merancang metode analisis yang efektif dari data diproses. Dalam banyak kasus, mungkin tidak dapat langsung menggunakan masalah data mining standar, seperti empat "superproblems" yang dibahas sebelumnya, untuk aplikasi yang berada di tangan. Namun, empat masalah ini memiliki cakupan yang luas sehingga banyak aplikasi dapat dipecah menjadi komponen yang menggunakan blok-blok bangunan yang berbeda.

3. METOLOGI PENELITIAN

1.1 Teknik Pengumpulan Data

Penelitian ini menggunakan data sekunder yang berupa data dokumen yang berisi komentar atau review dari penonton yang diambil dari http://imdb.com dan http://ai.stanford.edu/~amaas/data/sentiment/ berdasarkan penelitian dan klasifikasi manual oleh pemilik website tersebut. Metode pengumpulan data yang dilakukan adalah metode kuantitatif. Tujuan penelitian kuantitatif adalah mengembangkan dan menggunakan model-model matematis, teori-teori dan/atau hipotesis yang berkaitan dengan fenomena alam. Proses pengukuran adalah bagian yang sentral dalam penelitian kuantitatif karena hal ini memberikan hubungan yang fundamental antara pengamatan

empiris dan ekspresi matematis dari hubungan-hubungan kuantitatif.

1.2 Teknik Analisa dan Pemrosesan data

Penelitian ini menggunakan teknik analisa dan proses data dengan metode (CRISP-DM) CRISP-DM (*CRoss-Industry Standard Process for Data Mining*) [11] yang terbagi dalam dalam 6 proses besar yaitu :

1. *Business Understanding* (pemahaman bisnis) : Pada tahap pertama ini, tim desain harus berpikir hati-hati tentang skenario penggunaan (*use case)*. Awalnya, sangat penting untuk memahami masalah yang akan dipecahkan. Hal ini mungkin tampak jelas, tetapi proyek bisnis jarang datang sebagai sesuatu yang bersifat "pra-dikemas" dan mudah untuk dipecahkan atau tidak ambigu. Bahkan butuh kajian ulang dan mendesain solusi secara berulang kali.

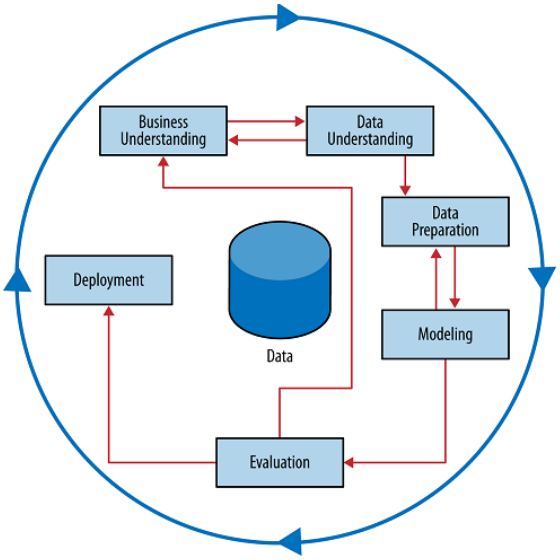
2. *Data Understanding* (pemahaman atau pengertian data) : Dalam tahap pemahaman data yang kita perlu lakukan adalah menggali di bawah permukaan untuk mengungkap struktur masalah bisnis dan data yang tersedia, dan kemudian mencocokkan mereka untuk satu atau lebih banyak dari data dimliki agar memiliki ilmu yang cukup besar dan teknologi untuk diterapkan.

3. *Data Preparation* (persiapan data) : Dalam teknologi analisa kita mengetahui bahwa ada beberapa persyaratan untuk bagaimana data tersebut dapat digunakan. Terkadang sering dibutuhkan data dalam bentuk yang berbeda dari yang disediakan secara alami, maka beberapa proses konversi akan diperlukan.

4. *Modeling* (pemodelan) : Tahap pemodelan adalah tahap utama di mana teknik data mining yang dipilih diterapkan untuk data.

5. *Evaluation* (evaluasi) : Tujuan dari tahap evaluasi adalah untuk menilai hasil data mining secara ketat dan untuk mendapatkan kepercayaan diri bahwa mereka adalah valid dan reliabel sebelum dipindahkan.

6. *Deployment* (penerapan) : Dalam tahap penerapan, hasil data mining dan teknik data mining sendiri mulai digunakan nyata dalam rangka mewujudkan beberapa pengembalian investasi dalam tahap penelitian sebelumnya.



Sumber : Foster Provost dan Tom Fawcett (2013:27)
Proses *data mining* (CRISP-DM)

Untuk persiapan data tekstual pada penelitian ini menggunakan teknik pengolahan *document prepraration* yaitu :

a. *Stop Word Removal* (Penghapusan kata pemberhenti)
b. *Stemming* (Pengelompokan)
c. Punctuation marks (Tanda baca)

3.3 Teknik Pengukuran

Penghitungan Classification Error

$$L = \frac{\sum_{j=1}^{n} w_j e_j}{\sum_{j=1}^{n} w_j}$$

n = Jumlah data.

j = Objek observasi.

$\mathbf{w_j}$ = Berat dari observasi j.

(Disini software akan menormalisasikan berat menjadi 1)

$\mathbf{e_j}$ = 1 jika kelas j yang diprediksi

berbeda dari observasi aslinya dan sisanya dianggap 0.

3.4 Teknik Perancangan Tampilan (GUI)

Perancangan GUI dilakukan dengan melakukan analisa data opini dari tahun 2011 yang diperoleh sebagai data *training* untuk menemukan metode klasifikasi yang paling akurat dengan algoritma s*upport vector machine* pada *rapidminer.* Lalu penulis akan melakukan implementasi hasil analisa tersebut kedalam data baru untuk melihat hasil klasifikasi opini dengan metode klasifikasi yang didapat. Kemudian melakukan perancangan GUI untuk menghasilkan laporan klasifikasi dan tampilan hasil penelitian untuk pembaca dengan *Microsoft Access 2013*.

4. HASIL DAN PEMBAHASAN

4.1. Rancangan Sistem

Dalam penelitian yang dilakukan penulis akan melakukan perancangan sistem yang berupa GUI (*Graphical User Interface*) untuk disajikan kepada para pembaca dan pengguna aplikasi agar mudah digunakan untuk melihat hasil penelitian klasifikasi yang dilakukan pada aplikasi *Rapidminer*. Perancangan sistem akan dilakukan dengan terlebih dahulu menemukan model klasifikasi yang akan digunkan dengan menggunakan metode *text mining* untuk menemukan hasil keakuratan metode klasifikasi yang akan diperoleh dengan menggunakan algoritma *Support Vector* Machine. Kemudia jika model klasifikasi sudah diperoleh akan dihitung menggunakan data *training* baru untuk mengklasifikasikan komentar atau opini baru yang akan ditentukan apakah terklasifikasi dalam positif atau negatif. Perancangan GUI akan dilakukan dan diterapkan menggunakan aplikasi *Microsoft Access 2013.*

4.2. Rancangan Basis Data

Akan dilakukan proses untuk menemukan model klasifikasi data yang paling akurat sehingga dapat diimplementasikan ke dalam data *training* baru dengan menggunakan metode *text mining* yang menerapkan algoritma klasifikasi *support vector machine.* Pada tahap ini proses akan dilakukan menggunakan aplikasi *rapidminer 5.3*. Proses tersebut meliputi:

1. Persiapan data:

a. Persiapan Data Opini
Sebelum proses dalam *rapidminer* dilakukan pada tahap ini penulis akan mempersiapkan data yang dibutuhkan yang diperoleh dari http://imdb.com dan http://ai.stanford.edu/~amaas/data/sentiment/ ke dalam folder pada windows untuk dipecah menjadi beberapa bagian sesuai dengan jumlah data yang dimiliki tiap folder. Total data yang diperoleh secara keseluruhan adalah sebesar 12.500 data pada masing-masing opini positif dan negatif. Data yang diperoleh berbentuk teks yang berisikan ulasan atau opini dari penonton yang meninggakan komentar pada imdb.com yang berasal dari berbagai macam film yang diambil secara acak.

b. Pembagian Data Opini
Dalam tahap ini akan dilakukan pembagian data opini yang diperoleh berdasarkan jumlah data yang dimiliki. Disini penulis akan membagi data opini yang diperoleh menjadi beberapa bagian

yaitu sebesar 1.000, 1.500, 2.000, 2.500, dan 3.000. Dari keseluruhan 12.500 data opini yang dimiliki penulis hanya mengamil sebagian kecil dari data tersebut karena sesuai dengan prinsip dan keunggulan algoritma SVM (*Support Vector Machine)* semakin kecil sampel data yang dimiliki tetapi beragam algoritma tersebut mampu memberikan model klasifikasi yang lebih baik dan lebih akurat, dibanding dengan jumlah data yang banyak tetapi dengan model klasifikasi yang kurang akurat. Karena alasan tersebut maka penulis hanya mengambil sebagian kecil dari keseluruhan data sebagai data *testing* untuk menemukan model klasifikasi berdasarkan algoritma *Support Vector Machine*.

2. Pengujian Model Klasifikasi Opini

a. Pembuatan Gudang Penyimpanan (repository) Baru

Sebelum dilakukan proses *data mining* terlebih dahulu akan dibuat tempat penyimpanan data dan proses atau biasa disebut dengan *repository baru* dalam *rapiminer untuk menyimpan seluruh* proses dan data dalam penelitian yang akan dilakukan dengan nama Klasifikasi Opini.

b. Persiapan Dokumen:

Pada akan dilakukan tahap persiapan dokumen atau yang disebut *document preparation*. Dalam tahap ini akan dilakukan proses-proses pengolahan data teks yang menggunakan metode *text mining* dalam *rapidminer* sebelum data opini tersebut siap untuk diolah dan digunakan dalam pemodelan klasifikasi yang dilakukan. Proses ini dilakukan pada *repository*.

"Klasifikasi Opini" yang terdiri dari beberapa tahap yaitu antara lain :

i. Import Document

Pada tahap ini akan dilakukan proses *import* data dokumen opini yang telah dipersiapkan di tahap sebelumnya ke dalam *rapidminer* untuk diolah lebih lanjut.

ii. Document Processing

Setelah melakukan pemilihan direktori dari dokumen yang dituju dan member nama pada *class* yang diinginkan pada list dan *parameter* kemudian, dilanjutkan dengan mengklik tombol *apply* untuk memasukkan data ke dalam operator tersebut. Setelah proses tersebut dilakukan maka langkah selanjutnya adalah dengan masuk ke proses lanjutan di dalam operator *process files from documents* dengan mengklik tombol *subprocess* pada operator tersebut.

iii. Performance & Classification Modelling

Pada tahap ini data opini yang sudah disiapkan pada tahap sebelumnya akan diterapkan pada algoritma *support vector machine* dan akan diukur performa keakuratannya untuk menemukan model klasifikasi yang paling akurat untuk setiap data teks yang digunkan untuk testing.

Maka berdasarkan hasil dan kesimpulan diatas dapat dibentuk sebuah tabel yang menunjukkan perbandingan seluruh hasil model prediksi yang telah dilakukan.

Tabel 1 Hasil Pemodelan dan Pengukuran Akurasi

Parameter/Sampel	1000	1500	2000	2500	3000
Klasifikasi Positif	92,41%	88,27%	88,01%	89,52%	88,20%
Klasifikasi Negatif	91,51%	84,01%	84,46%	83,37%	84,34%
Akurasi Total	91,96%	86,14%	86,23%	86,45%	86,27%
Error Total	8,04%	13,86%	13,77%	13,55%	13,73%
Bias (offset)	-0,014	-0,004	-0,009	-0,003	-0,024

Dari tabel1 terlihat bahwa dengan algoritma *support vector machine* menghasilkan model klasifikasi dengan akurasi yang sangat tinggi untuk mengelompokkan data opini. Ini terlihat dari rata-rata akurasi total dari sampel 1000 sampai dengan 3000 menghasilkan model dengan rata-rata lebih dari 80% dengan akurasi tertinggi diperoleh dari sampel 1000 data opini. Ini membuktikan bahwa metode klasifikasi opini menggunakan algoritma *support vector machine* layak digunakan sebagai model klasifikasi secara keseluruhan.

3. Implementasi Interface (GUI)

Pada tahap ini penulis akan melakukan perancangan GUI berdasarkan hasil yang diperoleh dari *rapidminer* untuk menghasilkan tampilan (*interface)* bagi pembaca agar hasil penelitian dapat terlihat dengan lebih jelas dan mudah dipahami. Perancangan akan dilakukan dengan aplikasi *Microsoft Access 2013.*

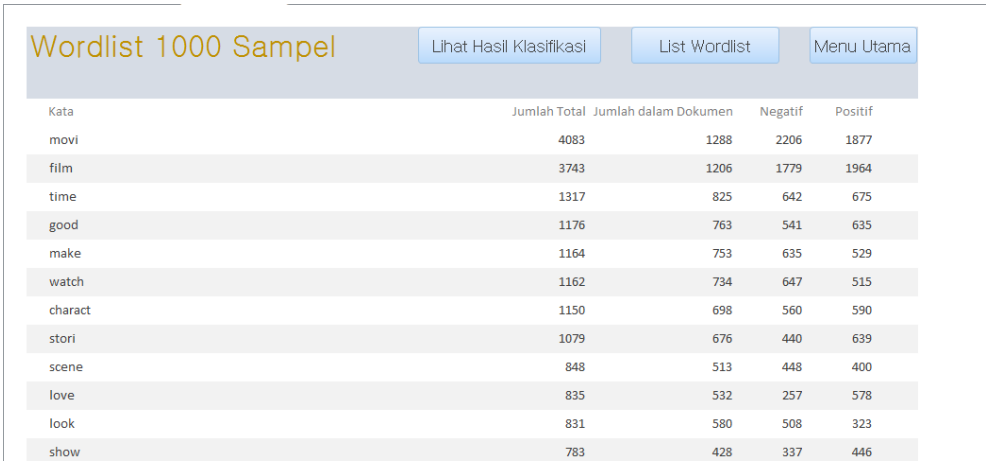
i. *Import Excel File*

Dalam tahap ini tipe atribut akan disesuaikan yaitu *short text* untuk atribut kata dan *integer* untuk sisanya.

Rancangan menu utama yang akan dibuat berdasarkan tabel tersebut kemudian akan terlihat seperti berikut.

Menu Utama

Sub-Menu Hasil Wordlist



Kata	Jumlah Total	Jumlah dalam Dokumen	Negatif	Positif
movi	4083	1288	2206	1877
film	3743	1206	1779	1964
time	1317	825	642	675
good	1176	763	541	635
make	1164	753	635	529
watch	1162	734	647	515
charact	1150	698	560	590
stori	1079	676	440	639
scene	848	513	448	400
love	835	532	257	578
look	831	580	508	323
show	783	428	337	446

Hasil Wordlist

5. KESIMPULAN

Dari hasil penelitian yang dilakukan untuk menganalisa dan mengklasifikasi opini penonton pada website IMDB.com dengan algoritma *support vector machine*. Penulis dapat menyimpulkan hal-hal sebagai berikut :

1. Dari hasil penelitian ini penulis dapat memberikan kemudahan untuk memproses data opini yang tidak terstruktur dengan melakukan pengolahan data berdasarkan metode *text mining* dan *data mining* dibanding dengan menggunakan metode pengolahan data biasa. Sehingga waktu yang diperlukan untuk memproses data tersebut untuk melakukan klasifikasi *sentiment analysis* tercapai dengan waktu yang relatif lebih singkat.
2. Ditemukannya bukti tingkat akurasi klasifikasi opini secara sistematis dengan menggunakan metode *data mining* dan *text mining* dengan

menggunakan algoritma *support vector machine*. Karena dengan diterapkannya proses klasifikasi secara sistematis maka penilaian berdasarkan *bias* (prasangka) individu dapat dikurangi karena penilaian dilakukan berdasarkan data opini yang telah didapat dan diteliti.

3. Ditemukan dan digunakannya teknik klasifikasi dengan algoritma *support vector machine* untuk memproses opini pentonton. Dimana hasilnya dapat terlihat dengan memprediksi opini negatif atau positif dari sebuah opini baru secara otomatis tanpa melakukan perbandingan secara manual.
4. Ditemukannya pola dan *wordlist* dari tiap kata dan opini penonton yang telah diteliti sehingga dapat melihat kata apa saja yang paling berpengaruh terhadap menentukan klasifikasi opini positif ataupun negatif.
5. Hasil akhir tentang persentase opini menunjukan: positif sebesar 92,41%, negatif sebesar 91,51%, dan eror (tidak terbaca) sebesar 1,88%.
6. Semakin besar jumlah data yang dipasang, semakin berkurang tingkat persentase hasilnya serta tingkat erornya semakin meningkat.

6. REKOMENDASI

Dalam penulisan penelitan karya akhir ini penulis memberi beberapa saran-saran lanjutan kepada para pembaca dan peneliti selanjutnya untuk bahan pertimbangan yaitu :

1. Karena dalam penelitian ini metode klasifikasi yang digunakan termasuk dalam klasifikasi *supervised learning* maka data yang digunakan merupakan data yang telah diklasifikasikan secara manual sebelumnya. Sehingga model klasifikasi yang diperoleh masih tergantung data yang terklasifikasi. Maka pada penelitan selanjutnya penulis berharap dapat memanfaatkan data yang tidak terklasifikasi sebelumnya (*unsupervised learning)* sehingga model klasifikasi yang didapat dapat benar-benar didapat secara sistematis.
2. Pengumpulan data opini dari imdb.com cukup sulit karena banyaknya jumlah data opini yang perlu dijadikan sampel, karena itu pada penelitian kali ini penulis mengambil sampel data imdb.com dari website lain yang menyediakan pengumpulan *dataset* opini secara gratis. Ini juga disebabkan karena alat atau aplikasi yang digunakan (*rapdiminer)* bersifat gratis sehingga kemampuan untuk melakukan *web crawling* masih terbatas. Alangkah baiknya bila penelitian selanjutnya dapat mengembangkan metode pengumpulan data dari website secara langsung tanpa mengandalkan website pengumpul data lainnya.
3. Penulis mengurangi jumlah sampel data yang digunakan untuk menemukan metode klasifikasi yang didapat dalam aplikasi *rapidminer* yang digunakan, ini disebabkan antara lain karena keterbatasan spesifikasi komputer yang digunakan oleh penulis dan penggunaan jumlah sampel data secara keseluruhan akan menyebabkan proses pemodelan data memakan waktu yang lama sehingga sampel data untuk pemodelan harus dikurangi untuk menghemat waktu pemrosesan dalam tahap pembuatan model klasifikasi. Untuk penelitian berikutnya disarankan agar peneliti selanjutnya dapat mengembangkan metode lain untuk mempercepat waktu pemrosesan dengan sampel data yang banyak dengan aplikasi *rapidminer* atau menggunakan aplikasi lain untuk melakukan proses *data mining* dan *text mining* dengan data sampel yang lebih banyak.
4. Pengunaan algoritma *support vector machine* masih sulit diterapkan penghitungannya jika dilakukan tanpa bantuan aplikasi, maka jika terjadi perubahan hasil atau model klasifikasi menggunakan aplikasi lain dengan metode yang sama dapat dijadikan sebagai referensi atau pembanding untuk menunjang penghitungan secara manual untuk kebutuhan publikasi.
5. Penelitian ini bersifat tidak baku, sehingga hasil dari penelitian dapat diperbaharui dan disempurnakan untuk menemukan hasil yang lebih maksimal dan lebih baik

7. DAFTAR REFERENSI

[1] Anggrawal, Charu C (2015), *Data Mining The Textbook,* New York : Springer International Publishing.

[2] Dean. Jared (2014), *Big Data, Data Mining, and Machine Learning : Value Creation for Business Leaders and Practitionerss,* New Jersey : John Wiley & Sons, Inc.

[3] Ertek, G et al (2012), *Text Mining with Rapidminer,* New Jersey : John Wiley & Sons, Inc.

[4] Esuli, Andrea dan Fabrizio Sebastiani, Jurnal : *Determining Term Subjectivity and Term Orientation for Opinion Mining,* Istituto di Scienza e Tecnologie dell'Informazione Consiglio Nazionale delle Ricerche Via Giuseppe Moruzzi, Italy.

[5] Han, Jiawei. et al (2012), *Data Mining Concepts and Techniques,* Edisi ke-3 Waltham : Elsevier Inc.

[6] J. Zaki, Mohammed dan Wagner Meira JR (2014), *Data Mining and Analysis Fundamentals Concepts and Algorithms,* New York : Cambridge University Press.

[7] Kudeyba, Stephan (2014), *Big Data, Mining, and Analytics : Components of Strategic Decision Making,* Boca Raton : Taylor and Francis Group. LLC.

[8] Laudon, Kenneth C. dan Jane P. Laudon, (2012), *Management Information Systems:Managing the Digital Firm*, Edisi ke-13, New Jersey : Pearson Prentice Hall.

[9] Liu, Bing (2012) *Sentiment Analysis and Opinion Mining,* London : Morgan & Claypool Publishers.

[10] M.Stair, Ralph dan George W. Reynolds (2012), *Fundamentals of Information Systems,* Edisi ke-6, Boston : Cengage Learning.

[11] Provott, Foster dan Tom Fawcett (2013), *Data Science for Business,* Sebastopol : O'Reilly Media, Inc.

[12] Sebastiani, Fabrizio. et al, Jurnal : *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,* Istituto di Scienza e Tecnologie dell'Informazione Consiglio Nazionale delle Ricerche Via Giuseppe Moruzzi, Italy.

[13] O.Brien, James A. (2010), *Management Information Systems,* Edisi ke-15, New York: Mc Graw Hill Irwin.

[14] Wallace, Patricia (2015), *Introduction to Information System,* Edisi ke-2, New Jersey : Preason Prentice Hall.

[15] Zhai, Cheng Xian dan Charu C Anggrawal (2012), *Mining Text Data,* New York : Springer International Publishing.