

MOVIE REVIEW SENTIMEN ANALISIS DENGAN METODE NAÏVE BAYES BASE ON FEATURE SELECTION

Andilala¹

¹Program Studi Teknik Informatika, Fakultas Teknik
Universitas Muhammadiyah Bengkulu

¹andilalaumb@gmail.com

Abstrak: Perkembangan internet dewasa ini berkembang sangat pesat yang secara tidak langsung mendorong penciptaan konten *web* pribadi yang melibatkan sentimen seperti *blog*, *tweets*, *web* forum dan jenis-jenis media sosial. Manusia sering kali mengambil keputusan yang didasari oleh masukan dari teman, saudara, kolega dan lain-lain. Dengan didukung ketersediaan pertumbuhan dan popularitas sumber daya yang kaya akan pendapat seperti *review* situs *online* dan *blog* pribadi. Misalnya, pengungkapan perasaan pribadi yang memungkinkan pengguna untuk membahas masalah sehari-hari, bertukar pandangan politik, mengevaluasi layanan dan produk. Pada penelitian ini diterapkan metode *opinion mining* dengan menggunakan algoritma *Naïve Bayes* berbasis *Feature Selection*. Pengujian metode ini menggunakan *dataset Movie Review* yang di *download* dari situs <http://www.cs.cornell.edu/People/pabo/movie-review-data/>. Data tersebut awalnya dikumpulkan dari *IMDb* (*Internet Movie Database*). Tujuan dari penelitian ini adalah untuk meningkatkan akurasi dari algoritma *Naïve Bayes* dalam mengklasifikasikan dokumen. Akurasi yang dicapai dalam penelitian ini adalah sebesar 95,70% melampaui akurasi yang dijadikan *base line* dalam penelitian ini adalah sebesar 94,85%.

Kata Kunci : *Naïve Bayes, Selection Feature, Opinion mining, Information Gain, Sentiment Analysis.*

Abstract: *The development of internet today is growing very fast which indirectly encourage the creation of personalized web content that involves sentiments such as blogs, tweets, web forums and other types of social media. Humans often make decisions based on input from friends, relatives, colleagues and others, with the support of growth and popularity of the availability of resources such opinion online review sites and personal blogs, for example, disclosure of the personal feelings that allow users to discuss everyday problems, exchanging political views, evaluate services and products. In this research, opinion mining methods are applied using the Naive Bayes algorithm-based Feature Selection. These methods use the dataset Movie Review <http://www.cs.cornell.edu/People/pabo/movie-review-data/> downloaded from the site. The data was originally collected from the IMDb (Internet Movie Database). The purpose of this study was to improve the accuracy of the Naive Bayes algorithm in classifying documents. accuracy is achieved in this study amounted to*

95.70% accuracy which exceeds the base line made in this study amounted to 94.85%.

Keywords: *Naïve Bayes, Selection Feature, Opinion mining, Information Gain, Sentiment analysis.*

I. PENDAHULUAN

Perkembangan internet dewasa ini berkembang sangat pesat dengan ditandai meningkatnya pengguna internet dunia. Kemajuan dari *Web 2.0* mendorong penciptaan konten *web* pribadi yang melibatkan sentimen seperti *blog*, *tweets*, *web* forum dan jenis-jenis media sosial. Dengan adanya keragaman aplikasi yang tumbuh di internet sebagian orang atau suatu badan dapat mengambil keuntungan dengan memanfaatkan pendekatan *opinion mining*, misalnya kampanye pemasaran dapat

menerima dan mengevaluasi umpan balik dari basis pengguna yang besar, politisi dapat memperkirakan popularitas mereka, produsen dapat mengidentifikasi kelemahan dari produk mereka, pemerintah dapat meningkatkan kualitas layanan informasi yang ditawarkan kepada masyarakat sehingga pemerintah lebih bertanggungjawab kepada masyarakatnya dan dapat memajukan transparansi di sektor publik [1]. Salah satu tujuan dari *opinion mining* dan *sentiment analysis* adalah untuk membantu orang menemukan informasi yang berharga dari sejumlah data yang tidak terstruktur [2].

Sentiment Analysis atau *Opinion Mining* baru-baru ini menjadi topik menarik yang mencoba untuk menggabungkan statistik, kecerdasan buatan dan teknologi. *Data Mining* dalam kerangka terpadu [3] *opinion mining* adalah informasi tekstual yang berada di dalam *web* dan berisi tentang fakta dan opini. Opini merupakan pernyataan subjektif yang mencerminkan persepsi seseorang terhadap sesuatu peristiwa, misalnya tentang opini-opini yang berkembang seperti krisis di Libya dan Suriah, perdebatan tentang krisis ekonomi di Yunani, dan *downrating* atas kredibilitas Amerika Serikat adalah beberapa topik kontroversial yang dimuat dalam berita sehari-hari. Menganalisa *rating movie* untuk mengetahui tingkat pendapatan dari pemutaran suatu *movie* [4]. Dan *review movie* juga dapat dimanfaatkan atau digunakan oleh penikmat film untuk memutuskan film apa yang mau ditonton dan para *producer* film dapat memanfaatkan opini tersebut untuk mengetahui penilaian yang diberikan oleh para penikmat film tentang film-film yang paling banyak dikritik.

Ada banyak metode yang telah digunakan dalam bidang *opinion mining* dan *sentiment*

analysis diantaranya adalah *Naive Bayes (NB)* dan *Support Vector Machine (SVM)*, masing-masing metode tersebut mempunyai kelebihan dan kekurangan masing-masing. Misalnya *SVM* yang memiliki *accuracy* cukup tinggi dalam pengklasifikasian dokumen seperti penelitian [1] [5] [6] [7] [8], tetapi *SVM* memiliki kelemahan yaitu dari segi kinerja dan kecepatan komputasi yang lambat [9]. Sedangkan *NB* memiliki akurasi paling rendah dalam pengklasifikasian dokumen seperti *accuracy* yang di hasilkan dalam penelitian [1] [4] [6], tetapi kelebihan *NB* dari metode yang lain khususnya dari *SVM* adalah *NB* cukup mudah digunakan dan memiliki *accuracy* yang lebih baik jika menggunakan dataset yang besar [9] [10].

Dalam penelitiannya [6] mengatakan bahwa *abasi et al* dan *Au gamon* menyarankan untuk menambahkan langkah seleksi fitur untuk proses *Machine Learning* jika ingin meningkatkan kinerja dari klasifikasi. Seleksi fitur atau pemilihan atribut dapat dilakukan berdasarkan perhitungan bobot yang diperoleh. Dalam penelitian ini, seleksi fitur yang digunakan adalah *Information Gain (IG)* yang mengacu ke penelitian [11] [12] berdasarkan skor fitur yang merupakan kualitas istilah dalam *dataset* dokumen. Sebuah istilah dengan nilai tinggi berarti penting atau relevan dengan *dataset*. Maka tujuan dari penelitian ini adalah untuk memperbaiki keakuratan metode *Naive Bayes* dalam mengklasifikasikan dokumen dengan pemilihan fitur merujuk ke penelitian [7] [6] [11] [12] [13]. Dan penelitian ini nantinya dapat dimanfaatkan untuk mengetahui tingkat *sentiment* yang ada di forum *web* seperti : politisi dapat memanfaatkannya untuk mengetahui popularitas mereka, produsen suatu produk dapat memanfaatkannya untuk mengetahui opini-opini

yang ada tentang produk mereka.

II. TINJAUAN PUSTAKA

A. Penelitian Terkait

Bo Pang et al [4] melakukan penelitian untuk mengetahui *rating* sebuah *movie* dengan cara melihat dan mengklasifikasikan *rating* ke label positif atau negatif dari sebuah *movie*, *dataset* yang digunakan dari *Internet Movie Database (IMDb)*, memanfaatkan metode klasifikasi *machine learning Naïve Bayes Classifier (NBC)*, *Maksimum Entropi (ME)* dan *Support Vector Machine (SVM)* dan menggunakan pemilihan fitur *unigram*, *n-gram* dan *POS*. Hasil dari penelitian tersebut dinyatakan bahwa algoritma terbaik dan efektif untuk klasifikasi terbaik adalah SVM, sedangkan NBC adalah klasifikasi terburuk.

Penelitian *Jingnian et al* [14] mereka melakukan penelitian dengan menkomparasi beberapa metode *feature selection* seperti *information gain (IG)*, *Multi-class Odds Ratio (MOR)* dan *Class Discriminating Measure (CDM)*, *Extended Odds Ratio (EOR)* dan *Weighted Odds Ratio (WOR)* dan *MC-OR* menggunakan multikelas *dataset* yaitu *dataset routers* dan *dataset* berbahasa Cina. *Accuracy* tertinggi yang mereka dapatkan adalah dengan pemilihan *feature CMD* dan *MOR* dengan *accuracy* sebesar 85,60 %.

Selanjutnya penelitian *Somayajulu et al* [7] mereka melakukan penelitian tentang *review movie* dengan menggunakan *dataset* dari *IMDb* dan menggunakan empat metode klasifikasi yaitu, *Naïve Bayes (NB)*, *Support vector Machine (SVM)*, *Complemented Naïve Bayes (CNB)* dan *Discriminative Partitioning using Bayesian Networks (DPBN)* dengan pemilihan fitur *Information Gain (IG)*.

Selanjutnya *abasi et al* [5] mereka melakukan penelitian tentang klasifikasi *sentiment* berbagai bahasa dengan menggunakan pendekatan SVM dan pemilihan fitur *Entropy Weighted Genetic Algorithm (EWGA)*, *Information Gain (IG)* dan *Genetic Algorithm (GA)*, mereka mengembangkan *feature selection EWGA* dengan memanfaatkan informasi yang dihasilkan dari *IG. Dataset* untuk klasifikasi yang mereka gunakan adalah bahasa inggris dan bahasa Arab. *Accuracy* tertinggi yang mereka dapatkan adalah sebesar 95,55% untuk *EWGA*.

B. Landasan teori

a. Text Mining

Text mining adalah salah satu bidang khusus dari *data mining* [15], *text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Permasalahan yang dihadapi pada *text mining* sama dengan permasalahan yang terdapat pada *data mining*, yaitu jumlah data yang besar, dimensi data yang tinggi dan struktur yang terus berubah, dan data *noise*.

b. Opinion Mining

Dalam bukunya *Bo pang* [3] mengatakan bahwa *opinion mining* dan *sentiment analysis* adalah menunjukkan bidang studi yang sama (daerah sub analysis subjektif) yang bertujuan

untuk mengetahui sebuah opini termasuk positif atau negatif. Ada dua cara untuk melakukan atau *opinion mining* yaitu *machine learning* dan *Semantic orientation* [2]. *Machine learning* dilakukan dengan cara mengumpulkan dan menyeleksi opini yang ada di *web*, selanjutnya opini-opini tersebut di berikan label positif dan negatif, sedangkan *semantic orientation* merupakan kebalikan dari *machine learning* dan digunakan secara *real time*.

c. *Movie Review*

Review movie memiliki karakteristik yang unik. Ketika seseorang menulis *review film*, ia mungkin berkomentar beberapa elemen dari sebuah film (*screen-play, vision effect, music*) dan ada juga dengan orang-orang yang terlibat di dalam film (sutradara, penulis skenario, aktor).

d. *Information Gain*

Feature adalah seluruh kata yang muncul dalam *training set*. *Set* ini biasanya sangat besar yaitu satu dimensi untuk setiap kata unik sehingga memiliki *feature space* yang sangat besar. *Feature selection* adalah metode untuk mereduksi dimensi *feature space* dengan cara memilih kata-kata yang paling *informative*. Salah satu metode yang terbukti sangat efektif dalam melakukan *feature selection* adalah *Information Gain (IG)*, Metode *IG* berfungsi menghitung jumlah bit informasi yang dihasilkan dengan jalan mengamati kemunculan sebuah kata dalam dokumen. Sebuah kata akan memiliki nilai *IG* yang tinggi jika kata tersebut muncul di banyak dokumen dalam suatu kelas tertentu.

Tujuan penerapan *IG* adalah informasi atribut atau fitur diukur dengan kemurnian. Ini merupakan jumlah informasi bahwa fitur ini membawa dan membantu dalam mengklasifikasikan contoh

baru berdasarkan kata ini saja. Persamaan *Information Gain* [25] seperti di bawah ini :

$$\text{Infogain} = \frac{\sum_{j=1}^k p(C_j) \log P(C_j) + P(W)}{\sum_{j=1}^k p(C_j/W) \log P(C_j/W) + P(\bar{W})} - \frac{\sum_{j=1}^k p(C_j/(\bar{W})) \log P(C_j/(\bar{W}))}{P(\bar{W})}$$

Dimana $P(c_j)$ adalah rasio jumlah kategori dokumen C_j untuk jumlah pelatihan seluruh dokumen, $P(w)$ adalah rasio jumlah dokumen yang mencakup term W untuk jumlah semua dokumen pelatihan, $P(C_j/W)$ adalah rasio jumlah dokumen yang meliputi term w yang merupakan milik kategori C_j untuk jumlah dokumen yang meliputi term W disemua sampel pelatihan, $P(\bar{w})$

adalah jumlah rasio dari jumlah dokumen yang tidak termasuk term W untuk jumlah semua dokumen pelatihan, $P(C_j/\bar{w})$ adalah rasio dokumen yang tidak termasuk term w tetapi milik kategori C_j untuk jumlah dokumen yang tidak termasuk term dalam semua sampel pelatihan.

e. *TFIDF*

Metode pembobotan adalah dimana tiap kata atau simbol dihitung sebagai satu fitur. Adapun perhitungan bobot digunakan adalah *Term Frequency-Inverse Document Frequency (TFIDF)*.

Bobot *TFIDF* adalah statistik numerik yang mencerminkan betapa pentingnya sebuah kata dalam suatu dokumen dalam koleksi atau *corpus*. Hal ini sering digunakan sebagai faktor bobot dalam pencarian informasi dan pertambangan teks. Nilai *tf-idf* meningkat secara proporsional dengan berapa kali suatu kata muncul dalam dokumen, tetapi diimbangi dengan frekuensi dari kata dalam *corpus*, yang membantu untuk mengendalikan fakta bahwa beberapa kata

umumnya lebih umum dibanding yang lain. Salah satu fungsi peringkat paling sederhana dihitung dengan menjumlahkan $tf * IDF$ sebagai mana ditunjukkan dengan persamaan dibawah ini :

$$w_{ij} = tf_{ij} \times (\log(D/df_j) + 1) \dots (1)$$

dimana tf_{ij} adalah jumlah kemunculan term pada kategori W_{ij} , dan D adalah jumlah dokumen yang dijadikan training, df_j adalah probabilitas kemunculan term pada kategori W .

f. *Naïve Bayes*

Algoritma mesin pembelajaran memainkan peran penting dalam melatih sistem dengan kategori yang telah ditetapkan di antaranya *Naïve Bayes* yang memiliki beberapa fakta menarik, mudah diterapkan dan memiliki akurasi yang lebih baik di *dataset* besar [10]. Pada *NBC* setiap *record* direpresentasikan dalam pasangan atribut $\langle a_1, a_2, \dots, a_n \rangle$ dimana a_1 adalah kata pertama, a_2 adalah kata kedua dan seterusnya.

Sedangkan V adalah himpunan kategori dokumen. klasifikasi, pendekatan Bayes akan menghasilkan label kategori yang paling tinggi probabilitasnya (V_{MAP}) dengan masukan atribut $\langle a_1, a_2, \dots, a_n \rangle$ $V_{MAP} = \arg_{v_j \in V} \max P(v_j | a_1, a_2, \dots, a_n) \dots (2)$

Teorema Bayes menyatakan :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \dots (3)$$

Menggunakan teorema Bayes ini, persamaan (16) ini dapat ditulis :

$$V_{MAP} = \arg_{v_j \in V} \max \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} \dots (4)$$

$P(a_1, a_2, \dots, a_n)$ nilainya konstan untuk semua v_j sehingga persamaan ini dapat ditulis sebagai berikut :

$$V_{MAP} = \arg_{v_j \in V} \max P(a_1, a_2, \dots, a_n | v_j) P(v_j) \dots (5)$$

Tingkat kesulitan menghitung $P(a_1, a_2, \dots, a_n | v_j)$

menjadi tinggi karena jumlah term $P(a_1, a_2, \dots, a_n | v_j)$ bisa jadi akan sangat besar. Ini disebabkan jumlah term tersebut sama dengan jumlah kombinasi posisi kata dikali dengan jumlah kategori. *Naïve Bayes Classifier* menyederhanakan hal ini dengan mengasumsikan bahwa dalam setiap kategori, setiap kata independen satu sama lain. Dengan kata lain : $P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \dots (6)$

Substitusi persamaan ini akan menghasilkan :

$$V_{MAP} = \arg_{v_j \in V} \max P(v_j) \prod_i P(a_i | v_j) \dots (7)$$

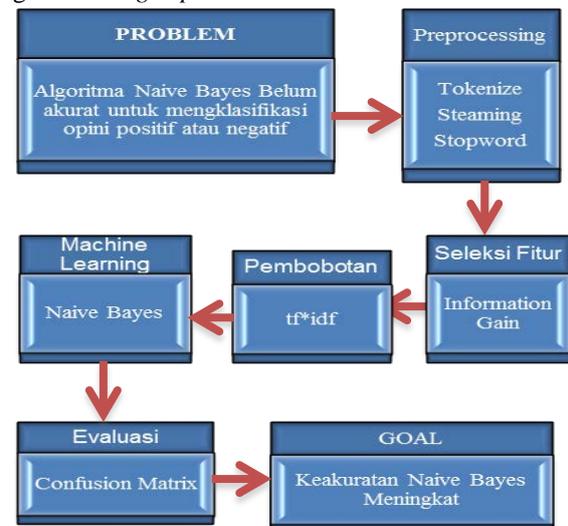
$P(v_j)$ dan probabilitas kata w_k untuk setiap kategori $P(w_k | v_j)$ dihitung pada saat pelatihan.

$$P(v_j) = \frac{|docs\ j|}{|Contoh|} \dots (8)$$

$$P(w_k | v_j) = \frac{nk+1}{n+|kosakata|} \dots (9)$$

Di mana $|docs\ j|$ adalah jumlah dokumen pada kategori j dan $|Contoh|$ adalah jumlah dokumen yang digunakan dalam pelatihan. Sedangkan n_k adalah jumlah kemunculan kata w_k pada kategori v_j dan $|kosakata|$ adalah jumlah kata yang unik (*distinct*) pada semua data pelatihan. Jumlah kata dalam tiap kelas dinyatakan sebagai n .

g. *Kerangka pemikiran*



Gambar 1. Kerangka Pemikiran

III. METODE PENELITIAN

Penelitian ini adalah diawali dengan mengambil dan mendownload data dari situs <http://www.cs.cornell.edu/People/pabo/movie-review-data/>, Data tersebut awalnya dikumpulkan dari *IMDb (Internet Movie Database)* [2] [4]. *Dataset* yang digunakan dalam penelitian ini adalah dataset *Movie Review* sama seperti dataset yang digunakan oleh penelitian [4] [2] [6] [16] [15] data tersebut telah bersih dan siap digunakan sebagai *corpus*, *dataset* tersebut berjumlah 2000 *file* teks yang telah di berikan label yaitu 1000 positif dan 1000 negatif.

Setelah data didapatkan, dilakukan pengujian model dengan diawali langkah *preprocessing* yang terdiri dari *tokenize* adalah proses yang memecah aliran teks menjadi kata, frasa, simbol, atau unsur-unsur bermakna lain yang disebut token. Daftar token menjadi masukan untuk diproses lebih lanjut seperti parsing atau pertambahan teks. Tokenisasi adalah hal yang kompleks untuk program komputer karena beberapa karakter dapat ditemukan sebagai token delimiters. Delimiter adalah karakter *spasi*, *tab* dan *baris baru* “*newline*”, sedangkan karakter () < > ! ? “ kadang kala dijadikan delimiter namun kadang kala bukan tergantung pada lingkungannya [16].

Setelah melalui proses *tokenize* dan kalimat sudah dalam bentuk frasa, langkah selanjutnya adalah melalui tahapan *Stopword* yaitu menghapus daftar kata-kata yang tidak dipakai dalam pemrosesan bahasa alami, dengan cara disaring sebelum pengolahan dokumen teks. Mereka umumnya dianggap sebagai 'kata-kata fungsional' yang tidak membawa makna seperti kata-kata

“*affected*, *affecting*, *affects*, *after*,

afterwards, *again*, *became*, *because*, *become*, *becomes*, *Becoming*, *been*, *before*, *became*, *because*, *become*, *becomes*, *becoming*, *been*, *before*, dll.

Setelah melalui tahapan penyaringan kata yang di lakukan oleh *stopword*, langkah selanjutnya adalah proses *Steaming* yang digunakan untuk mengurangi ukuran dari himpunan fitur awal dan untuk menghapus salah eja atau kata-kata dasar yang sama. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama.

Setelah melalui tahapan *preprocessing*, langkah selanjutnya adalah menyeleksi *feature* yang akan dijadikan *training* dalam mesin pembelajaran, seleksi fitur yang digunakan dalam penelitian ini adalah *Informationn Gain (IG)* untuk bisa menghitung nilai dari IG nilai *Entropy* harus di ketahui terlebih dahulu, *entropy* bisa dikatakan sebagai bit untuk menyatakan suatu kelas. Semakin kecil nilai *entropy* maka semakin baik untuk digunakan dalam mengekstraksi suatu kelas. Tujuan dari *gain* adalah mengukur seberapa baik suatu atribut memisahkan *training example* ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan *gain*. Setelah nilai *gain* dari *term* tiap-tiap dokumen training di ketahui, langkah selanjutnya adalah melakukan pembobotan untuk tiap-tiap term tersebut, pembobotan yang dilakukan dalam penelitian ini adalah pembobotan *tf*idf*. Prinsip kerja dari pembobotan yang digunakan merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. Semakin sedikit dokumen yang mengandung term yang dimaksud, maka nilai *idf* semakin besar. Jika setiap dokumen dalam koleksi

mengandung term yang bersangkutan, maka nilai dari idf dari term tersebut adalah nol. Hal ini menunjukkan bahwa sebuah term yang muncul pada setiap dokumen dalam koleksi tidak berguna untuk membedakan dokumen berdasarkan topik tertentu.

Penghitungan bobot dari term tertentu dalam sebuah dokumen dengan menggunakan $tf*idf$ menunjukkan bahwa deskripsi terbaik dari dokumen adalah term yang banyak muncul dalam dokumen tersebut dan sangat sedikit muncul pada dokumen yang lain. Demikian juga sebuah term yang muncul dalam jumlah yang sedang dalam proporsi yang cukup dalam dokumen dikoleksi yang diberikan juga akan menjadi descriptor yang baik. Bobot terendah akan diberikan pada term yang muncul sangat jarang pada beberapa dokumen (*low-frequency documents*) dan term yang muncul pada hampir atau seluruh dokumen (*high-frequency document*).

Setelah nilai bobot dari tiap term pada seluruh dokumen yang telah terseleksi di ketahui, langkah selanjutnya adalah penerapan metode *machine learning Naïve bayes* dengan tujuan untuk mengetahui klasifikasi atau *sentiment* dari data training apakah termasuk *sentiment* negatif atau *sentiment* positif.

IV. HASIL PENELITIAN

Berdasarkan dari penerapan model yang digunakan dalam penelitian maka didapatkan keakuratan klasifikasi dari algoritma *naïve bayes* dan *naïve bayes* berbasis *feature selection* sebagai berikut:

A. Naïve Bayes

Keakuratan klasifikasi dari *naïve bayes* untuk klasifikasi *document movie review* ditunjukkan dengan *Accuracy* yang dihasilkan

adalah sebesar = 95,15%, *Precision* sebesar = 96,40% dan *recall* dihasilkan sebesar = 93,80% sebagaimana ditunjukkan pada gambar dibawah ini:

	true negatif	true positif	class precision
pred. negatif	965	62	93.96%
pred. positif	35	938	96.40%
class recall	96.50%	93.80%	

	true negatif	true positif	class precision
pred. negatif	965	62	93.96%
pred. positif	35	938	96.40%
class recall	96.50%	93.80%	

	true negatif	true positif	class precision
pred. negatif	965	62	93.96%
pred. positif	35	938	96.40%
class recall	96.50%	93.80%	

Gambar 2. Nilai *Precision* dan *Recall* Klasifikasi *Document Movie Review*

B. Naïve Bayes berbasis *feature selection*

Keakuratan klasifikasi dari *naïve bayes* berbasis *feature selection* untuk klasifikasi *document movie review* meningkat sebagaimana ditunjukkan dengan *accuracy* sebesar = 95,70%, *Precision* sebesar =96,44% dan *recall* sebesar = 94,90%. Sebagaimana di tunjukkan dalam gambar di bawah ini:

	true negatif	true positif	class precision
pred. negatif	966	51	94.98%
pred. positif	35	943	96.44%
class recall	96.50%	94.90%	

	true negatif	true positif	class precision
pred. negatif	965	51	94.98%
pred. positif	35	943	96.44%
class recall	96.50%	94.90%	

	true negatif	true positif	class precision
pred. negatif	965	51	94.98%
pred. positif	35	943	96.44%
class recall	96.50%	94.90%	

Gambar 3. Nilai *Precision* dan *Recall* *feature selection* Untuk Klasifikasi *Document Movie Review*

Gambar 3 diatas menunjukkan analisa evaluasi hasil *Naïve Bayes* dan *Fitur Selection* dengan selisih nilai akurasi sebesar 0,55, akurasi yang dihasilkan dalam penelitian ini dapat melampaui akurasi dari penelitian yang di jadikan baseline yaitu penelitian somayajulu [7] akurasi tertinggi sebesar 94,85%. dan penelitian [14] dengan akurasi tertinggi sebesar 85,60%.

Dalam uji coba yang dilakukan dengan *accuracy*, *Precision* dan *Recal* yang di hasilkan antara *naïve bayes* dan *naïve bayes* berbasis *selection feature* berbeda, perbedaan tersebut terutama terdapat dalam jumlah total dokumen positif benar diklasifikasikan positif dan dokumen negatif benar diklasifikasikan negatif.

Seperti pada pemodelan *naïve bayes*, jumlah total dokumen uji adalah sebesar 2000 dokumen yang terdiri dari 1000 positif dan 1000 negatif. Tetapi di dalam model *naïve bayes*, dokumen negatif yang benar diklasifikasikan negatif adalah sebesar 965 dokumen, dan dokumen negative yang diklasifikasikan ke positif sebesar 35 dokumen. Dokumen positif yang benar diklasifikasikan ke positif sebesar 938 dokumen dan dokumen positif yang masuk kekategori negatif sebesar 62 dokumen.

Setelah pemodelan *naïve bayes* di beri *selection feature*, terjadi sedikit perubahan dalam klasifikasi seperti yang terjadi pada dokumen kategori positif sedangkan dokumen yang berkategori negatif tidak terjadi perubahan dalam proses klasifikasi masih tetap seperti pemodelan *naïve bayes*. Peningkatan tersebut terjadi pada Dokumen positif yang benar diklasifikasikan ke positif sebesar 949 dokumen dan dokumen positif yang masuk kekategori negatif sebesar 51 dokumen.

V. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa :

Accuracy yang dihasilkan dari penelitian ini dapat melampaui *accuracy* dari penelitian yang dijadikan base line yaitu penelitian [14] [7] menggunakan metode NB dan comparasi beberapa fitur seleksi, *accuracy* terbesar didapatkan oleh CMD sebesar 85,60% dan [7] sebesar 94,85%.

Accuracy tersebut dilampaui tanpa memberi seleksi fitur terhadap algoritma *Naïve Bayes*, *naïve bayes* hanya diberikan pembobotan $tf*idf$, *accuracy* yang dihasilkan tanpa memberikan seleksi fitur adalah sebesar 95,15%. *Accuracy* yang dihasilkan setelah memberikan seleksi fitur adalah sebesar 95,70%, meningkatkan sebesar 0,9%. Peningkatan *accuracy* dalam pengklasifikasian dokumen dengan penambahan seleksi fitur, peningkatan *accuracy* yang dihasilkan memang tidak terlalu tinggi dari *naïve bayes* tanpa seleksi fitur.

REFERENSI

- [1] George Stylios et al., "Public Opinion Mining for Governmental Decisions," *Electronic Journal of e-Government*, vol. 8, no. 2, pp. 202-213, 2010.
- [2] Pimwadee Chaovalit and Lina Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches," *IEEE*, pp. 1-9, 2005.
- [3] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [4] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Tehniques," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79-86, July 2002.
- [5] Ahmed Abbasi, Hsinchun Chen , and Arab Salem , "Sentiment analysis in multiple languages : Feature selection for opinion clasification in web forums," *ACM Transactions on Information Systems*, vol. 26, no. 3, pp. 1-34, 2008.
- [6] Tim O Keefe and Irena Koprinska, "Feature Selection and Weighting Methods in Sentiment Analysis," *Proceedings of the 14th Australasian Document Computing Symposium*, pp. 1-8, Desember 2009.

- [7] Somayajulu , Siva RamaKhrisnha Reddy, and Ajay Dani, "Classification of Movie Reviews Using Complemented Naive Bayesian Classifier," *International Journal of Intelligent Computing Research (IJICR)*, vol. 1, pp. 162-167, 2010.
- [8] Long Sheng Chen and Chia Wei Chang, "A New Term Weighting Method by Introducing Class Information for Sentiment Classification of Textual Data," *Proceeding of International MultiConference of engineers and Computer Scientists, IMECS*, vol. 1, pp. 1-4, March 2011.
- [9] Colas Fabrice and Brazdil Pavel, "Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks," *Artificial Intelligence in Theory and Practice*, vol. 217, no. 1, pp. 169-178, 2006.
- [10] Vidhya K A and G Aghila , "A Survey of Naïve Bayes Machine Learning approach in Text Document Classification," *International Journal of Computer Science and Information Security, IJCSIS*, vol. 7, no. 2, pp. 206-211, 2010.
- [11] Erik Boiy, Pieter Hens, Koen Deschacht, and Marie Francine Moens, "Automatic Sentiment Analysis in On-line Text," *Proceedings ELPUB2007 Conference on Electronic Publishing*, pp. 349-360, June 2007.
- [12] Wen Zhang, Taketoshi Yoshida, and Xijin Tang, "A Study with Multi-Word Features in Text Classification," *Proceedings of the 51st Annual Meeting of the ISSS*, pp. 1-8, 2007.
- [13] Yelena Mejova, *Topic Tracking & Sentiment Analysis in Twitter*, University of Iowa, 2011.
- [14] Jingnian Chen, Houkuan Huang, Shengfeng Tian, and Youli Qu, "Feature selection for text classification with Naïve Bayes," in *Expert Systems with Applications*. Beijing, China: Elsevier Ltd, 2009, pp. 5432-5435.
- [15] Bruno Ohana and Brendan Tierney, "Supervised Learning Methods for Sentiment Classification with RapidMiner," *RapidMiner Community Meeting And Conference, RCOMM*, pp. 1-8, June 2011.
- [16] Rawia Awadallah, Maya Ramanath, and Gerhard Weikum, "Harmony and Dissonance: Organizing the People's Voices on Political Controversies," *Proceedings of The Fifth ACM International Conference On Web Search And Data Mining, ACM*, pp. 523-532, February 2012.
- [17] Louise Francis and Matt Flynn, *Text Mining Handbook.*: Spring, 2010.