

# ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING

Agatha Deolika<sup>1</sup>, Kusri<sup>2</sup>, Emha Taufiq Luthfi<sup>3</sup>

<sup>1,2,3</sup>Magister Teknik Informatika, Universitas Amikom, Yogyakarta

Jl. Ring Road Utara, Kabupaten Sleman, Daerah Istimewa Yogyakarta

[agatha.deolika@students.amikom.ac.id](mailto:agatha.deolika@students.amikom.ac.id)<sup>1</sup>, [kusri@amikom.ac.id](mailto:kusri@amikom.ac.id)<sup>2</sup>, [emhataufiqluthfi@amikom.ac.id](mailto:emhataufiqluthfi@amikom.ac.id)<sup>3</sup>

**Abstract** - In this era, we need to extract the text needed to visualize or need knowledge from a large collection of document texts. Text mining is the process of obtaining high quality information from text. High-quality information obtained because of attention to patterns and trends by reading statistical patterns. In the process of extracting the text, we need to pay for the words offered to give value/weight to the terms provided in a document. The weight given to the term depends on the method used. In weighting many words such as algorithms for example such as TF, IDF, RF, TF-IDF, TF.RF, TF.CHI, WIDF. This research will be analyzed and compared with the TF-IDF, TF.RF, and WIDF algorithms. For the test method, the naïve Bayes classification method will be used and the valuation analysis using the confusion matrix. With a dataset used as many as 130 documents in which 100 data transfer and 30 test data. Based on the analysis of the results of the classification that has been done, it can determine the weighting of TF.RF with naïf classification is better than weighting TF.IDF and WIDF with Accuracy values of 98.67%, Precision 93.81%, and Recall 96.67%.

**Keywords** - Text Mining, TF-IDF, TF-RF, WIDF, Classification, Naïve Bayes.

**Abstract** - Pada era sekarang ini pemanfaatan text mining sangatlah diperlukan untuk mevisualkan atau mengevaluasi pengetahuan dari kumpulan besar dari teks dokumen. Text mining adalah proses untuk memperoleh informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik. Pada proses teks mining terdapat pembobotan kata yang bertujuan untuk memberikan nilai/bobot pada term yang terdapat pada suatu dokumen. Bobot yang diberikan pada term tergantung kepada metode yang digunakan. Dalam pembobotan kata banyak sekali terdapat algoritma-algoritma contohnya seperti TF, Idf, RF, TF-IDF, TF.RF, TF.CHI, WIDF. Pada penelitian ini akan dianalisis dan dibandingkan algoritma TF-IDF, TF.RF, dan WIDF. Untuk metode pengujiannya akan digunakan metode klasifikasi naïve bayes dan analisis perbandingannya menggunakan confussion matrix. Dengan dataset yang digunakan sebanyak 130 dokumen yang mana 100 data training dan 30 data uji. Berdasarkan analisa pada hasil klasifikasi yang telah dilakukan, dapat disimpulkan bahwa pembobotan TF.RF dengan klasifikasi Naïve bayes lebih baik dari pembobotan TF.IDF dan WIDF dengan nilai Accuracy 98,67%, Precision 93,81%, dan Recall 96,67%.

**Kata kunci** - Text Mining, TF-IDF, TF-RF, WIDF, Klasifikasi, Naïve Bayes.

## I. PENDAHULUAN

Pada era sekarang ini pemanfaatan text mining sangatlah diperlukan untuk mevisualkan atau mengevaluasi pengetahuan dari kumpulan besar dari teks dokumen. Text mining adalah proses untuk memperoleh informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik. Pada proses teks mining terdapat pembobotan kata yang bertujuan untuk memberikan nilai/bobot pada term yang terdapat pada suatu dokumen. Bobot yang diberikan pada term tergantung kepada metode yang digunakan. Dalam pembobotan kata banyak sekali terdapat algoritma-algoritma contohnya seperti TF, Idf, RF, TF-IDF, TF.RF, WIDF.

Banyak pihak berupaya memodifikasi pembobotan kata, salah satu upaya yang dapat dilakukan adalah dengan memodifikasi atau menggabungkan beberapa metode pembobotan sehingga menjadi sebuah metode pembobotan yang baru. TF-IDF, TF.RF dan WIDF adalah beberapa contoh diantaranya. Term frequency inverse document frequency (TF-IDF) adalah metode pembobotan yang menggabungkan metode TF dan IDF. Metode ini diusulkan oleh Salton sebagai sebuah kombinasi metode yang dapat memberikan performansi yang lebih baik, khususnya dalam memperbaiki nilai recall dan precision. Weighted inverse document frequency (WIDF) adalah sebuah perluasan dari inverse document frequency (IDF)[1]. Term Frequency ·Relevance Frequency atau disebut juga TF-RF, adalah sebuah metode penggabungan antara metode TF dan RF

dengan tujuan untuk mendapatkan performansi yang lebih baik [2]. Pada penelitian yang dilakukan oleh Diki Susandi yang berjudul Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. Dari hasil pengujian precision dan recall didapat bahwa dengan menggunakan pembobotan TF.IDF ketepatan dan kelengkapan sistem dalam melakukan klasifikasi dokumen sebesar 70,7%. Sedangkan dengan menggunakan pembobotan WIDF ketepatan dan kelengkapan sistem dalam melakukan klasifikasi dokumen sebesar 70,6% [3]. Pada penelitian yang dilakukan oleh Thopo Martha Akbar yang berjudul analisis perbandingan metode pembobotan kata  $tf.idf$  dan  $tf.rf$  terhadap performansi kategorisasi teks, Berdasarkan hasil pengujian, terlihat bahwa nilai precision, recall, dan fmeasure untuk TF•RF lebih baik dibandingkan metode TF•IDF untuk sebagian besar pengujian yang dilakukan [4]

Berdasarkan dari study literature maka disini penulis ingin menanalisis 3 pembobotan kata yaitu TF-IDF, TF.RF, WIDF dengan menggunakan metode naïve Bayes yang akan dihitung secara masing-masing untuk mencari performa tertinggi dari kombinasi masing-masing pembobotan kata dan metode naïve bayes menggunakan metode *confussion matrix* dengan contoh kasus klasifikasi data aspirasi masyarakat pada aplikasi Layanan Aspirasi dan Pengaduan Online Rakyat (LAPOR!).

#### A. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan lainnya [5].

#### B. Term Weighting

Term weighting adalah sebuah metode pembobotan kata (term) untuk memberikan sebuah bobot atau nilai untuk kata (term) yang terkandung dalam sebuah dokumen. Bobot nilai ini menjadi ukuran besarnya jumlah dan tingkat kontribusi sebuah kata (term) untuk penentuan suatu kelas atau kategori dalam suatu dokumen. Terdapat beberapa metode pembobotan kata (term weighting) diantaranya adalah TF, TF-IDF, WIDF, dan TF-RF.

#### C. Term Frequency Inverse Document Frequency (TF-IDF)

Data yang telah melalui tahap preprocessing harus berbentuk numerik. Untuk mengubah data tersebut menjadi numerik yaitu menggunakan metode pembobotan TF-IDF. Metode Term Frequency Invers Document Frequency (TF-IDF) merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (term) terhadap dokumen dengan memberikan bobot setiap kata. Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut [6]. Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF perkata dengan bobot masing-masing kata adalah 1. Sedangkan nilai IDF diformulasikan pada Persamaan (1).

$$IDF(word) = \log \frac{td}{df} \quad (1)$$

$IDF(word)$  adalah nilai IDF dari setiap kata yang akan di cari,  $td$  adalah jumlah keseluruhan dokumen yang ada,  $df$  jumlah kemuculan kata pada semua dokumen.

#### D. Term Frequency. Relevance Frequency (TF.RF)

Metode TF.RF adalah metode dari penggabungan dua metode, yaitu metode TF dan metode RF. Penggabungan dua metode ini diharapkan dapat menghasilkan performansi yang lebih baik. Dengan persamaan :

$$TF * RF(d, t) = TF(d, t) * RF(c, t) \quad (2)$$

#### E. Weighted Inverse Document Frequency (WIDF)

Weighted Inverse Document Frequency atau disebut juga WIDF, adalah sebuah metode pembobotan yang muncul sebagai upaya perbaikan terhadap metode IDF. Kekurangan pada metode IDF adalah tidak adanya perlakuan yang berbeda untuk dokumen yang banyak memunculkan term yang sama dengan dokumen lainnya yang lebih sedikit memunculkan term yang sama. Metode IDF hanya akan menilai term berdasar kemunculan (ada atau tidaknya saja) term-nya saja pada suatu dokumen, tanpa melihat frekuensi (jumlah) term yang muncul pada dokumen tersebut. Hal ini berbeda dengan metode WIDF yang akan memperhitungkan frekuensi dari suatu term yang muncul pada suatu dokumen untuk menormalisasikan frekuensi term tersebut terhadap keseluruhan dokumen. Pada WIDF, bobot dari suatu term dihitung dengan menggunakan persamaan:

$$WIDF(d, t) = \frac{TF(d, f)}{\sum_{i \in D} TF(i, t)} \quad (3)$$

$$W(d, t) = WIDF(d, t)$$

#### F. Algoritma NBC

Naive Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas

dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian [7].

Olson dan Delen dalam bukunya, menjelaskan bahwa Naive Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar. Algoritma ini berdasar pada mengasumsikan bahwa atribut obyek adalah sesuatu yang independen. Probabilitas yang terlibat dalam membuat perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan [8].

Naive Bayes Classifier memiliki akurasi lebih baik dibanding dengan model classifier lainnya. Penelitian yang dilakukan oleh Xhemali, Hinde dan Stone dalam jurnalnya "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages" mengatakan bahwa "Naive Bayes Classifier memiliki tingkat akurasi yang lebih baik dibanding model classifier lainnya". Teorema bayes merupakan dasar aturan dari naive bayes classifier berikut teorema bayes akan disajikan pada persamaan (4).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

Dimana X merupakan data tuple hasil pengujian dari suatu set data yang telah ditentukan masuk ke ke dalam kelas tertentu. H merupakan suatu hipotesis yang akan menentukan X masuk ke dalam kelas C. P(H|X) merupakan peluang atau probabilitas X yang merupakan data tuple atau bukti yang diperoleh pada saat observasi masuk ke dalam kelas C, dengan kata lain mencari probabilitas X dimiliki oleh kelas C. P(H|X) merupakan probabilitas posterior, H dikondisikan pada X. Sebaliknya P(H) merupakan probabilitas prior, atau probabilitas sebelumnya. Kemudian P(X|H) merupakan probabilitas posterior dimana X dikondisikan pada H. Sedangkan P(X) merupakan probabilitas sebelumnya dari X [9].

Dengan aturan Bayes maka penelitian ini akan mengimplementasikan aturan bayes dengan studi kasus tertentu oleh karena itu aturan bayes dapat dinyatakan :

$$P(W_i|C) = \frac{\text{count}(w_i.C)+1}{\text{count}(C)+|V|} \quad (5)$$

- W<sub>i</sub> = adalah kata ke i
- W<sub>i</sub>.C = jumlah kata w<sub>i</sub> dalam C(count)
- |V| = total kata pada data testing
- Count ( C ) = jumlah kata pada berdasarkan kategori

#### G. Pengukuran performance

Pada tahap ini melakukan pengukuran performansi sistem classifier yang dibangun untuk mengetahui tingkat akurasi dengan menggunakan parameter performansi. Parameter performansi yang digunakan yaitu diantaranya nilai akurasi, precision,

dan recall. Berikut adalah tabel confusion matrix pada table 1.

Tabel 1. CONFUSION MATRIX

Kategori		True Value	
		True	False
Hasil klasifikasi	True	TP	FP
	False	FN	TN

Tabel 1 menunjukkan bahwa hasil klasifikasi benar sesuai dengan keputusan sistem (TP) dan Klasifikasi salah menurut sistem (FP). Sedangkan dokumen yang tidak termasuk dalam hasil klasifikasi kategori, adakalanya memang bukan anggota dari kategori itu (TN) dan adakalanya ternyata seharusnya menjadi anggota kategori tersebut (FN) [10]. Keempat parameter di atas digunakan untuk menghitung 3 metode evaluasi yakni :

*Recall*, yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah seluruh dokumen relevan. *Recall* memiliki rumusan sebagai berikut :

$$Recall = \frac{TP}{TP+FP} \quad (6)$$

*Precision* yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah dokumen yang terkenali. *Precision* memiliki persamaan sebagai berikut :

$$Precision = \frac{TP}{TP+FN} \quad (7)$$

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aslinya. Akurasi digunakan untuk mengevaluasi banyaknya label prediksi yang sesuai dengan label aktual. Semakin besar nilai akurasinya, maka performansi klasifikasi semakin baik. Berikut persamaannya:

$$Akurasi = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (8)$$

## II. METODE PENELITIAN

Penelitian yang akan diusulkan berjenis penelitian eksperimen. Penelitian dilakukan menggunakan dataset yang didapatkan dari aplikasi LAPOR!. Peneliti ini menggunakan data teks dengan bahasa Indonesia baik dan benar. Metode yang digunakan dalam penelitian ini adalah analisis perbandingan pembobotan kata teks mining yaitu TF-IDF, TF-RF, WIDF dengan menggunakan masing-masing metode klasifikasi Naive Bayes Classifier.

Dalam pengerjaan penelitian ini diperlukan langkah-langkah kegiatan penelitian untuk mendapatkan hasil yang maksimal, dapat dilihat alur penelitian pada Gambar 1. Langkah-langkah tersebut adalah sebagai berikut :

1. Pengumpulan Dokumen (Dataset) Pada tahap ini penelitian dilakukan dengan teknik mengumpulkan data-data yang berupa teks yang ada pada aplikasi Layanan Aspirasi dan Pengaduan Online Rakyat (LAPOR!).
2. Preprocessing data Text Mining untuk merubah data teks yang berupa dokumen menjadi perkata. Preprocessing dokumen dilakukan dengan memecah

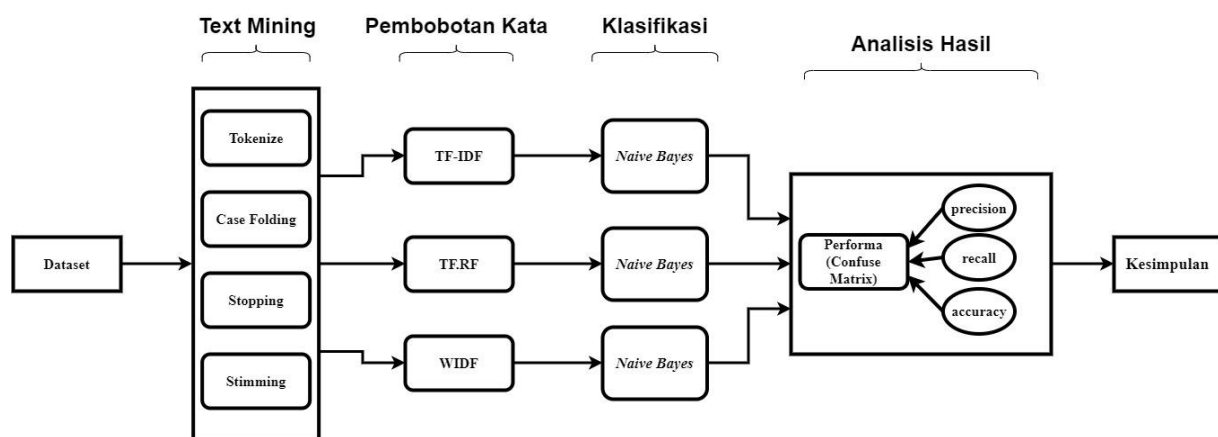
string teks dari dokumen yang panjang menjadi kalimat-kalimat menggunakan fungsi split(), dengan tanda titik ".", tanda tanya "?" dan tanda seru "!" sebagai delimiter untuk memotong string dokumen. Setelah proses memecah string selesai dilanjutkan dengan pemilahan kalimat menjadi perkata (tokenazing) dengan menggunakan fungsi explode() yang mana memisah kata berdasarkan spasi. Kemudian melakukan Case Folding yang mana merubah huruf capital menjadi huruf kecil dan setelah itu filter kata yang tidak penting (stoplist) yang pada akhir mengembalikan kedalam bentuk dasar (stemming).

3. Proses pembobotan kata, pada proses ini akan dilakukan perhitungan nilai pembobotan kata berdasarkan frekuensi kata terhadap dokumen, masing-masing pembobotan kata akan dihitung dan

nantinya akan menjadi masing-masing nilai input pada metode klasifikasi.

4. Proses Algoritma Naive Bayes Classifier ini akan melakukan proses traning dan testing dari dataset yang sudah ditentukan. Untuk proses testing nilai input yang digunakan adalah dari masing-masing pembobotan, maka dari itu pada penelitian ini akan dilakukan 3 kali proses atau kombinasi yaitu proses tf-idf dan naïve bayes, tf.rf dan naïve bayes, widf dan naïve bayes.

5. Analisa Hasil (Akurasi) dilakuakn dengan cara membandingkan antara hasil klasifikasi naïve bayes classifier dan hasil klasifikasi manual pada setiap kombinasi dengan menggunakan *confussion matrix*(*Recall, Precission, F-Measure*).



Gambar 1. Diagram Desain Sistem

### III. HASIL DAN PEMBAHASAN

Pada penelitian ini digunakan berbagai macam data, diantaranya data Traning, data Testing. Data Traning digunakan untuk membentuk sebuah model klasifikasi. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi dokumen data baru yang belum pernah ada. Data ini akan digunakan sebagai proses mining berupa 100 data dokumen yang sudah dikelompokan berdasarkan kategori dinas terkait. Data ini memiliki atribut Kalimat dan Hasil dan data ini hanya mecakup 5 Dinas yaitu PUPR, Pendidikan, Kesehatan, Pertanian, Dukcapil. Beberapa data training dapat dilihat pada Tabel 2.

Tabel 2. Data Traning

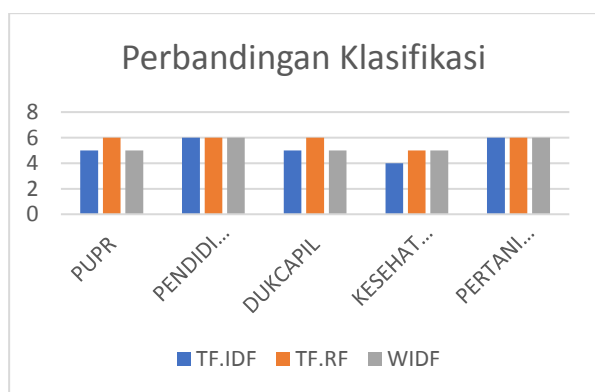
No	Kalimat	Hasil
1	Pelayanan pada puskesmas tolong ditingkatkan karena sering tenaga keseshatan tidak ada ditempat.	Kesehatan
2	Tolong agar sistem zona dihilangkan karena banyak anak sekolah kebingungan	Pendidikan
3	Apakah benar ada pembukaan penerimaan tenaga kontrak pada dinas PU	PUPR
4	Kepada Bupati tolong diusut perusakan sawah di desa trahean	Pertanian
5	Berapa bulan syarat saya bisa mendapatkan surat domisili?	Dukcapil
....		
100	Alat berat di banyak mengganggu jalan dan meakibatkan kemacetan	PUPR

Hasil klasifikasi sistem yang direpresentasikan melalui Tabel 3 diperoleh informasi bahwa dari 30 dokumen uji yang terdiri dari masing-masing 6 data perkelas. Setelah dilakukan pengujian didapatkan Nilain TF.IDF 26, TF.RF 29, WIDF 27 dokumen yang

akurasinya sesuai dengan data yang sdh di kelompokkan secara manual. Hasil sistem tersebut diperoleh melalui beberapa tahap yakni tahap text preprocessing dan klasifikasi teks dengan menggunakan Algoritma naïve bayes. Hasil dari pengelompokan algoritma naïve bayes yang sudah dikategorikan tersebut dibandingkan dengan data yang dikelompokkan secara manual, dapat dilihat pada Gambar 2

Tabel 3. Hasil Data Pengujian

No	Data	Manual	Naïve Bayes		
			TF.IDF	TF.RF	WIDF
1	Data 1	PUPR	Tidak Akurat	Akurat	Akurat
2	Data 2	PERTANIA N	Akurat	Akurat	Akurat
3	Data 3	PUPR	Akurat	Akurat	Akurat
4	Data 4	DUKCAPIL	Tidak Akurat	Akurat	Tidak Akurat
5	Data 5	KESEHATA N	Tidak akurat	Tidak Akurat	Tidak Akurat
...					
30	Data 30	PENDIDIK AN	Akurat	Akurat	Akurat



Gambar 2. Grafik Perbandingan Klasifikasi

Hasil klasifikasi sistem yang direpresentasikan melalui tabel 2 diperoleh informasi bahwa dari 30 dokumen uji didapatkan akurasi yang berbeda-beda. Hasil sistem tersebut diperoleh melalui beberapa tahap yakni tahap text preprocessing, pembobotan kata, dan klasifikasi teks dengan menggunakan Algoritma naïve bayes. Hasil dari pengelompokan algoritma naïve bayes yang sudah dikategorikan tersebut dibandingkan dengan data yang dikelompokkan secara manual, dapat dilihat pada gambar 2.

	TF.IDF	TF.RF	WIDF
Precision	92,67%	93,81%	91,42%
Recall	83,33%	96,67%	83,33%
Accuracy	94,67%	98,67%	94,67%

Berdasarkan hasil Accuracy menggunakan *confussion matrix* yang ditunjukkan pada table 4 dan gambar 2, diperoleh nilai tertinggi pada pembobotan kata TF.RF dengan nilai Accuracy 98,67%, Precision 93,81%, dan Recall 96,67%.

IV. KESIMPULAN

Berdasarkan analisa pada hasil klasifikasi yang telah dilakukan, dapat disimpulkan bahwa pembobotan TF.RF dengan klasifikasi Naïve bayes lebih baik dari pembobotan TF.IDF dan WIDF dengan nilai Accuracy 98,67%, Precision 93,81%, dan Recall 96,67%. Klasifikasi naïve bayes dapat digunakan untuk mengelompokkan atau klasifikasi text mining dengan baik.

V. SARAN

Berdasarkan penelitian diatas mungkin pada penelitian selanjutnya dapat menggunakan atau menerapkan metode filter kata pada proses text Preprosesing dan juga dapat melakukan pengujian menggunakan metode-metode klasifikasi lainnya.

DAFTAR PUSTAKA

[1] Tokunaga T. & Iwayama, M, Text Categorization Based On Weighted Inverse Document Frequency. Tokyo, Japan: Tokyo Institute of Technology, 1994.

[2] Lan, Man, A New Term Weighting Method for Text Categorization. National University of Singapore. Singapore, 2006.

[3] Diki Susandi, Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. Jurnal ProTekInfo Vol. 3 No. 1 ISSN: 2406-7741, 2016.

[4] Thopo Martha Akbar, Analisis Perbandingan Metode Pembobotan Kata Tf.Idf Dan Tf.Rf Terhadap Performansi Kategorisasi Teks. Tugas Akhir Teknik Informatika Universitas Telkom, 2012.

[5] Mandias, F. A., Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi, Konferensi Nasional Sistem & Informatika 2015, STIKOM Bali, 9-10 Oktober 2015, 2015.

[6] Fitri, Meisya. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia. Universitas Tanjungpura : Semarang, 2013.

[7] M. Syukri Mustafa, Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, Citec Journal, Vol. 4, No. 2 2460-4259, 2017.

- [8] Olsen, D. L., Delen, D, Advanced Data Mining Techniques, Penerbit pringer, USA, 2008.
- [9] Han, J., and M.Kamber, Data Mining Concept and Techniques Second Edition, Elsevier, San Frasco, 2006.
- [10] Fitri Handayani, Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. Vol. 7 No. 1, Jurnal Teknik Elektro, Semarang, 2015.