

Sentimen Analisis pada Komentar Angket Dosen di STMIK STIKOM Indonesia

I Kadek Dwi Gandika Supartha¹, I Wawan Sudiarsa²

¹³ Prodi Sistem Komputer, STMIK STIKOM Indonesia Denpasar Bali

¹gandika.supartha@stiki-indnesia.ac.id

²sudiarsa@stiki-indonesia.ac.id

Abstrak— Dalam penelitian ini melakukan sentiment analisis terhadap komentar angket dosen di STMIK STIKOM Indonesia. Angket dosen digunakan untuk mengukur kinerja dosen disetiap akhir semester, namun akan sangat menyulitkan jika untuk melakukan evaluasi komentar secara manual. Untuk itulah diperlukan suatu aplikasi yang dapat menentukan sebuah komentar termasuk ke dalam sentimen positif atau negative secara otomatis. Metode yang digunakan yaitu Naïve Bayes Classifier (NBC), metode ini dipilih karena cukup sederhana, memiliki akurasi dan kecepatannya yang tinggi. Proses penentuan sentimen dimulai dari memasukkan data latih berupa komentar yang sentimennya sudah ditentukan sebelumnya. Selanjutnya mengubah setiap huruf besar pada komentar menjadi huruf kecil (case folding). Proses text mining pada data latih melalui tiga tahap yaitu tokenizing, filtering dan stemming. Proses stemming pada penelitian ini menggunakan Nazief dan Adriani Stemmer dan kemudian disimpan kedalam database. Selanjutnya yaitu memasukkan data uji berupa komentar baru yang belum diketahui sentimennya. Proses selanjutnya sama seperti pada data latih tetapi pada data uji setelah stemming ada proses analisis. Pada tahapan analisis komentar sebagai data latih yang ada didatabase diambil untuk melaksanakan perhitungan menggunakan metode Naive Bayes Classifier sehingga diketahui sentimen dari data uji yang telah diinputkan. Hasil pengujian menunjukkan bahwa semakin besar jumlah data latih maka presisi semakin besar. Presisi maksimum sebesar 80% dicapai pada data latih 250 dan data uji sebanyak 50.

Kata kunci— Sentimen, Angket, Naive Bayes Classification

Abstract— In this study, a sentiment analysis was conducted on the comments of a lecturer questionnaire at STMIK STIKOM Indonesia. Lecturer questionnaires are used to measure the performance of lecturers at the end of each semester, but it would be very difficult to evaluate comments manually. For this reason, we need an application that can determine whether a comment is included in a positive or negative sentiment automatically. The method used is the Naïve Bayes Classifier (NBC), this method was chosen because it is quite simple, has high accuracy and speed. The process of determining sentiment starts from entering training data in the form of comments whose sentiments have been predetermined. Next, change each uppercase letter in comments to lowercase (case folding). The text mining process on training data goes through three stages, namely tokenizing, filtering and stemming. The stemming process in this study used Nazief and Adriani Stemmer and then stored into the database. Next is to enter test data in the form of new comments whose sentiments are not yet known. The next process is the same as in the training data but in the test data after stemming there is an analysis process. In the analysis stage, comments as training data in the database are taken to perform calculations using the Naive Bayes Classifier method so that sentiment is known from the test data that has been inputted. The test results show that the greater the amount of training data, the greater the precision. The maximum precision of 80% was achieved on 250 training data and 50 test data. **Keywords**— Include 5 – 6 keywords or phrases, keywords—Sentiment, Questionnaire, Naive Bayes Classification

I. PENDAHULUAN

Penilaian kinerja seseorang sangat penting dilakukan dibidang usaha apapun yang dapat digunakan sebagai bahan evaluasi pekerja tersebut. Kinerja seseorang dinilai secara kualitatif atau berdasarkan kualitas kerjanya. Untuk mengubah data yang berbentuk kualitatif menjadi data kuantitatif dapat menggunakan angket. STMIK STIKOM Indonesia sudah menerapkan penggunaan angket untuk melakukan penilaian dan evaluasi kinerja dosen disetiap akhir semester. Angket yang digunakan memberikan pertanyaan mengenai kinerja seorang dosen dengan range penilaian mulai dari sangat setuju, setuju, tidak setuju dan sangat tidak setuju. Selain penilaian melalui pertanyaan, pada angket tersebut juga dilengkapi dengan komentar secara tertulis sehingga mahasiswa dapat memberikan komentar kepada setiap dosen. Akan tetapi, dengan banyaknya jumlah angket yang ada tentunya akan sangat menyulitkan untuk melakukan proses evaluasi secara manual, terutama evaluasi pada bagian komentar.

Sentiment analysis merupakan bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap dan emosi seseorang terhadap entitas seperti produk, jasa, organisasi, individu, isu, kejadian, topik beserta atributnya[1]. Akan tetapi proses analisis sentimen sendiri memiliki beberapa masalah teknis. Salah satu masalah teknis dalam

melakukan analisa sentimen adalah mengetahui orientasi pendapat, apakah pendapat tersebut merupakan pendapat positif atau negative[2]. Untuk mengatasi masalah tersebut dibutuhkan suatu metode yang dapat menilai sentimen pendapat apakah suatu pendapat, dalam hal ini adalah komentar pada angket, merupakan komentar positif atau komentar negatif.

Salah satu metode klasifikasi yang dapat digunakan untuk melakukan sentiment analisis adalah metode *Naïve Bayes* yang sering disebut dengan *Naïve Bayes Classifier (NBC)*. Kelebihan NBC adalah sederhana tetapi memiliki akurasi yang tinggi. Algoritma NBC yang sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi[3].

Beberapa penelitian sebelumnya yang berkaitan dengan *Naive Bayes* seperti Haditsah Annur tentang klasifikasi masyarakat miskin menggunakan metode *Naive Bayes*, diperoleh tingkat akurasi sebesar 73%, nilai *Precision* sebesar 92% dan *Recall* sebesar 86%[4]. Hasil penelitian Gustientiedina tentang memprediksi tingkat kepuasan mahasiswa terhadap pelayanan akademis di beberapa perguruan tinggi di Pekanbaru dengan menggunakan metode *Naive Bayes*. Hasil pengujian menunjukkan bahwa *Naive Bayes* cocok digunakan untuk mengukur tingkat kepuasan

mahasiswa terhadap pelayanan akademis dengan akurasi sebesar 96,71% dengan nilai *precision* sebesar 96,15% dan nilai *recallnya* sebesar 98,43%[5].

Berdasarkan uraian maka akan dilakukan penelitian tentang sentiment analisis pada komentar angket dosen di STMIK STIKOM Indonesia dengan menggunakan metode *Naïve Bayes Classification*, diharapkan hasilnya dapat membantu mempercepat melakukan analisis sentimen terhadap angket dosen sehingga dapat membantu dalam mengukur kinerja dari dosen di STMIK STIKOM Indonesia.

II. METODOLOGI PENELITIAN

Bagian ini menjelaskan secara rinci tentang penelitian yang dilakukan.

A. Text Mining

Text mining memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapat dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. *Text mining* adalah teknologi baru yang digunakan untuk data perusahaan yang selalu bertambah sehingga data teks yang tidak terstruktur tersebut dapat dianalisis[6]. Tahapan secara umum dalam *text mining* adalah *tokenizing*, *filtering*, *stemming* dan *analyzing*. *Tokenizing* yaitu tahap pemotongan string input berdasarkan kata penyusunnya. *Filtering* adalah tahap mengambil kata penting dari hasil token. *Stemming* adalah tahap mencari kata dasar hasil filtering. *Analyzing* adalah tahap penentuan *probabilitas* kata pada dokumen.

B. Sentiment Analysis

Analisis sentimen adalah proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negative[7]. Besarnya pengaruh dan manfaat dari analisis sentimen, menyebabkan penelitian ataupun aplikasi mengenai analisis sentiment berkembang pesat, bahkan di Amerika kurang lebih 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen. Pada dasarnya analisis sentiment merupakan klasifikasi, tetapi kenyataannya tidak semudah proses klasifikasi biasa karena terkait penggunaan bahasa.

C. Naïve Bayes Classifier (NBC)

Naïve Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan *probabilitas* sederhana yang berakar pada Teorema *Bayes* dan memiliki asumsi ketidak tergantungan (*independent*) yang tinggi dari masing – masing kondisi atau kejadian[8]. Bentuk umum teorema *bayes* adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesa data X merupakan suatu kelas spesifik

P(H|X) = *Probabilitas* hipotesis H berdasarkan kondisi X (*posterior probability*)

P(H) = *Probabilitas* hipotesis H (*prior probability*)

D. Nazief and Andriani Stemmer

Algoritma Nazief & Adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani, memiliki tahap-tahap sebagai berikut: Pertama cari kata yang akan diistem dalam kamus kata dasar. Jika ditemukan maka diasumsikan kata adalah *root word*. Maka algoritma berhenti. *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Hilangkan *derivation prefixes* DP {“di-”, “ke-”, “se-”, “me-”, “be-”, “pe”, “te-”} dengan iterasi maksimum adalah 3 kali[9].

E. Stop Word

Stop words adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna[10]. Contoh *stop words* untuk bahasa Inggris diantaranya “of”, “the”. Sedangkan untuk bahasa Indonesia diantaranya “yang”, “di”, “ke”. Selain bertujuan untuk *information retrieval*, *stop words* digunakan untuk klasifikasi atau untuk mengurangi jumlah kata yang harus diproses.

F. Performance Pengujian

Mengukur unjuk kerja sistem sentiment analisis dengan menganalisa hasil klasifikasi. Dalam penelitian ini menggunakan *confusion matrix* yang merupakan sebuah metode untuk menganalisa seberapa baik sebuah model klasifikasi. Akurasi yang dihasilkan dalam bentuk *persentase* dari kelompok latih yang telah diklasifikasi dengan benar. *Precision* dan *recall* digunakan untuk mengukur kinerja sistem. *Precision* adalah kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan. *Recal* merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Accuray* adalah tingkat kedekatan antara nilai yang didapat terhadap nilai sebenarnya. *Precision*, *recall* dan *accuracy* dapat dihitung dengan menggunakan *confusion matrix* dapat dilihat pada table berikut ini[11].

TABEL I
CONFUSION MATRIX

	<i>True</i>	<i>False</i>
<i>True (Positive)</i>	TP <i>(True Positive)</i> <i>Correct result</i>	FP <i>(False Positive)</i> <i>Unexpected result</i>
<i>False (Negative)</i>	FN <i>(False Negative)</i> <i>Missing result</i>	TN <i>(True Negative)</i> <i>Correct absence of result</i>

Precision dapat dihitung menggunakan persamaan berikut ini:

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Recall dapat dihitung dengan menggunakan persamaan berikut ini:

$$recal = \frac{TP}{TP + FN} \quad (3)$$

Untuk menghitung *accuracy* dapat menggunakan persamaan berikut ini:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Keterangan:

TP = Banyak data dengan nilai sebenarnya positif dengan nilai prediksi positif

FP = Banyak data dengan nilai sebenarnya negative dan nilai prediksi positif

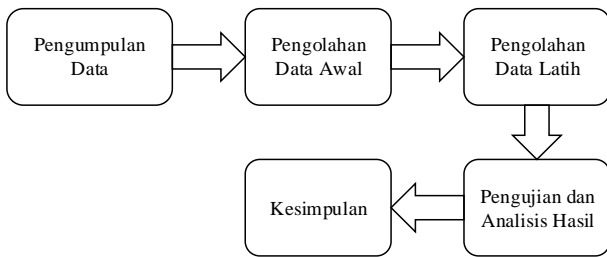
FN = Banyak data dengan nilai sebenarnya positif dan nilai prediksi negative

TN = Banyak data dengan nilai sebenarnya negatif dan nilai prediksi negative

Ukuran besaran *precision*, *recall* dan *accuracy* biasanya diberi dalam bentuk presentase anrata 1 sampai 100%. Sebuah sistem akan dianggap baik jika tingkat *precision*, *recal* dan *acuracynya* tinggi.

G. Alur Penelitian

Alur penelitian dapat dilihat pada Gambar dibawah ini:



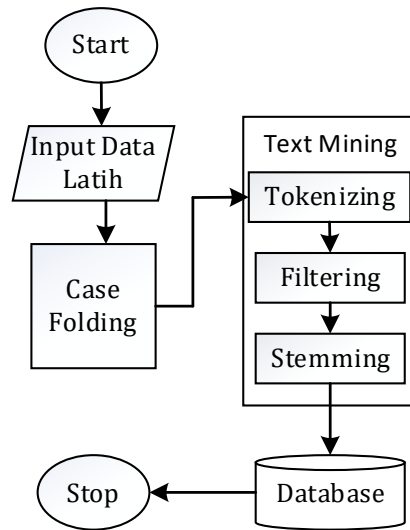
Gambar 1. Alur Penelitian

Pengumpulan data awal dilakukan untuk mengidentifikasi masalah yang ada, dimulai dengan melakukan wawancara dengan bagian akademik untuk mendapatkan data yang diperlukan. Pengolahan data awal dilakukan dengan cara mengolah data yang sudah didapatkan agar cocok dengan proses klasifikasi sentiment contohnya seperti menghilangkan kata sambung, memperbaiki kata yang kurang tepat dan lain sebagainya. Selanjutnya yaitu pengolahan data latih berupa komentar-komentar pada angket dosen. Setelah pengolahan data latih yang sudah dimasukkan ke dalam database, pengujian siap dilakukan dengan cara menguci komentar-komentar pada angket dosen yang belum diketahui sentimennya.

Pada analisis akan dilakukan pengujian dengan data uji untuk mengetahui seberapa akurat metode *Naïve Bayes* dalam melakukan klasifikasi sentiment positif dan negatif. Tahap terakhir yaitu melakukan penarikan kesimpulan dari analisis yang telah dilakukan.

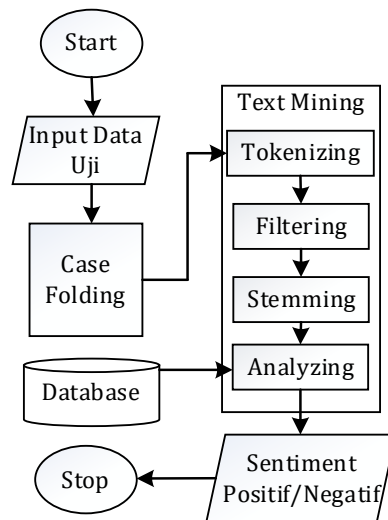
H. Flow Cart Sistem

Flow chart sistem sentimen angket dosen terdiri dari dua flowchart yaitu *flow chart* untuk data latih dan *flow chart* untuk data uji yang dapat dilihat pada gambar di bawah ini:



Gambar 2. Flowchart Data Latih

Untuk melakukan pelatihan diawali dengan memasukkan data latih berupa komentar angket dosen kemudian mengubah setiap huruf besar menjadi huruf kecil (*case folding*). Terakhir yaitu melakukan proses *text mining* pada setiap data latih, tetapi untuk data latih proses hanya dilakukan sampai tahap *stemming*.



Gambar 3. Flowchart Data Uji

Untuk pengujian dimulai dari input data uji baru kemudian prosesnya sama seperti *flowchart* data latih perbedaannya pada pengujian yaitu proses analisis dengan menggunakan data latih yang diambil dari *database*. Proses analisis menggunakan metode *Naïve Bayes Classifier*, hasil yang didapat adalah data uji tersebut termasuk sentiment positif atau negative.

I. Proses Text Mining

Proses *text mining* melalui empat tahap yaitu *tokenizing*, *filtering*, *stemming*, dan *analyzing*. Pada tahap *analyzing* komentar-komentar sebagai data latih yang telah disimpan di database sebelumnya akan diambil untuk melakukan perhitungan menggunakan metode NBC. Berikut ini adalah contoh perhitungan metode NBC dengan menggunakan

contoh data sebanyak 4 komentar sebagai data latih dan 1 komentar sebagai data uji. Komentar tersebut antara lain:

TABEL II
CONTOH DATA LATIH DAN DATA UJI

Komentar	Data Komentar	Sentimen
Komentar Uji KU	Diharapkan tetap mengajar, karena saya mudah paham dengan apa yang dosen ini ajarkan, dan cara mengajarnya tidak membosankan	??
Komentar 1 K2	Saya senang dengan cara mengajar bapak, saya jadi mudah paham	Positif
Komentar 2 K2	Cara mengajar sangat bagus, membuat saya cepat memahami mata kuliah ini	Positif
Komentar 3 K3	Saya rasa dosen ini sangat kurang humoris jadi mengikuti mata kuliah ini jadi kurang menarik dan membosankan	Negatif
Komentar 4 K4	Dosen ini tidak disiplin, sering datang terlambat dan saat mengajar kurang contoh sehingga sulit dipahami	Negatif

Hasil dari *tokenizing*, *filtering*, dan *stemming* dari setiap komentar pada Tabel I adalah sebagai berikut:

TABEL III
HASIL TOKENIZING, FILTERING DAN STEMMING

Komentar	Data Komentar	Sentimen
KU	harap, ajar, mudah, paham, dosen, ajar, mengajar, bosan	?
K1	senang, ajar, mudah, paham	Positif
K2	ajar, bagus, cepat, paham, mata kuliah	Positif
K3	dosen, humoris, ikut, mata kuliah, tarik, bosan	Negatif
K4	dosen, disiplin, lambat, ajar, contoh, sulit, paham	Negatif

Untuk mengestimasi probabilitas kemunculan sebuah kata sebagai sebuah sentimen positif atau negatif dengan melihat kumpulan data latih sentimen positif dan negatif dan menghitung seberapa sering kata tersebut muncul dalam setiap kelas. Hal ini yang membuat *training* ini sebagai pembelajaran terbimbing. Sehingga persamaan untuk menyelesaikan permasalahan ini adalah sebagai berikut:

$$P(F_1, \dots, F_n) = \frac{1}{Z} \prod_{i=1}^n p(F_i|C) \quad (5)$$

P(Positif|harap, tetap, ajar, mudah, paham, ajar, bosan)=

$$\frac{1}{2} x (1 x 1 x 2 x 1 x 2 x 2 x 1)$$

$$\frac{1}{2} x 8 = 4$$

P(Negatif|Harap, tetap, ajar, mudah, paham, ajar, bosan)=

$$\frac{1}{2} x (1 x 1 x 1 x 1 x 1 x 1 x 1)$$

$$\frac{1}{2} x 1 = 0,5$$

Selanjutnya adalah menganalisa sentimen dengan menggunakan metode Naive Bayes. Dari hasil perhitungan diatas, dapat dilihat hasil dari sentimen positif adalah 4 dan hasil dari perhitungan sentimen negatif adalah 0,5. Karena hasil setimen positif lebih besar dari sentimen negatif, maka dapat disimpulkan data uji tersebut masuk ke dalam sentimen positif

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan terdiri dari implementasi sistem, pelatihan dan pengujian.

A. Implementasi Sistem

Implementasi sistem dapat dilihat pada Gambar di bawah ini

Gambar 4. Form Utama

Pada *form* utama ini terdapat *tab* uji komentar dan data latih. Pada *tab* uji komentar digunakan untuk menguji komentar angket dosen dan pada *tab* data latih dapat digunakan untuk menginputkan data latih. Tampilan *tab* data latih dapat dilihat pada Gambar 5 di bawah ini

NO	Komentar	Sentimen	Hapus
1	Memberi support kepada mahasiswa	Positif	Hapus
2	menghargai mahasiswa yang terlambat / yang datang ke kampus	Positif	Hapus
3	Cara mengajar terlalu cepat terus suaranya kecil	Negatif	Hapus
4	Lebih humoris, karena bapak terkesan galek	Negatif	Hapus
5	Penyampaian kuliah sangat menarik	Positif	Hapus
6	Mohon saat mengajar jangan cepat	Negatif	Hapus
7	Proses mengajar terlalu keku	Negatif	Hapus
8	Bahasanya terlalu tinggi dalam penyampaian materi	Negatif	Hapus
9	Saya senang dengan cara mengajar Bapak saya jadi mudah mengerti	Positif	Hapus

Gambar 5. Form Data Latih

B. Pelatihan dan Pengujian

Sebelum melakukan pengujian data latih harus diinput ke dalam database, total keseluruhan data latih yang akan digunakan yaitu 250 data latih yang terdiri dari 125 data latih sentiment positif dan negative. *User* menginputkan komentar beserta *sentiment* positif atau *negative*, kemudian komentar akan mengalami proses *text mining* yang terdiri dari *tokenizing*, *filtering* dan *stemming*.

Total keseluruhan data uji yang akan digunakan adalah 50 data uji. Secara umum tahapan pengujian adalah dengan menginputkan data uji pada *form* utama yang selanjutnya mengalami proses *text mining* dan pengujian sentiment komentar dengan metode *Naïve Bayes Classifier*. *Probabilitas* yang telah dihitung pada tahap *analyzing* akan digunakan untuk mengetahui hasil pengujian komentar. Dalam penelitian ini akan dilakukan beberapa kombinasi data latih dan data uji. Kombinasi data latih yang akan digunakan antara lain 200, 210, 220, 230, 240, 250, sedangkan untuk kombinasi data uji digunakan 20 dan 50 data uji, dimana jumlah data latih yang memiliki sentiment positif sama dengan jumlah data latih yang memiliki sentiment negatif.

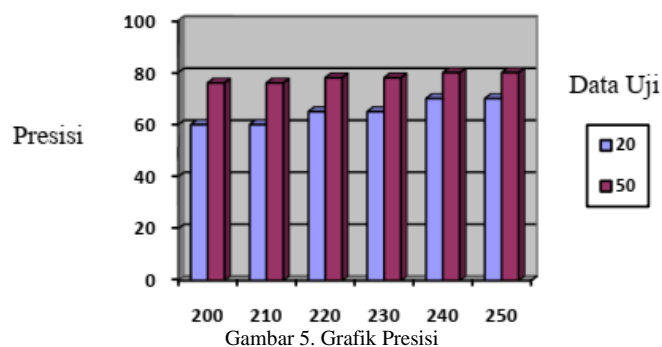
C. Analisis Hasil Pengujian

Dari hasil pengujian dengan menggunakan data uji diluar data latih yang telah dilakukan didapatkan hasil dan presisi yang dihitung dengan rumus pada persamaan (2) sebagai berikut.

TABEL IV
HASIL PENGUJIAN

Jumlah Data Uji	Jumlah Data Latih	Jumlah Data Benar	Presisi
20	200	12	60%
	210	12	60%
	220	13	65%
	230	13	65%
	240	14	70%
50	250	14	70%
	200	38	76%
	210	38	76%
	220	39	78%
	230	39	78%
	240	40	80%
	250	40	80%

Dari hasil pengujian pada Tabel IV dapat dilihat bahwa nilai presisi berbanding lurus dengan jumlah data latih, semakin besar jumlah data latih maka nilai presisi semakin besar.



Berdasarkan grafik pada Gambar 4.5 nilai presisi maksimum dicapai pada data latih berjumlah 240 dan 250. Sistem melakukan kesalahan dalam menentukan sentiment, prosentase kesalahan yang dilakukan oleh sistem sebesar 20% dan prosentasi benar sebesar 80%. Hal ini dikarenakan peluang kata yang terdapat pada komentar uji hanya muncul pada data latih dengan sentiment tertentu, sehingga sistem memberikan hasil yang kurang tepat, maka perlu diperhatikan jumlah kata pada data latih untuk setiap sentimennya.

IV. KESIMPULAN

Hasil analisis pengujian menunjukkan bahwa kinerja dari metode *Naïve Bayes Classifier* dalam melakukan sentiment analisis angket dosen di STMIK STIKOM Indonesia membarikan hasil yang cukup baik yaitu pada jumlah data latih sebanyak 250 dan data uji sebanyak 50 dengan presisi sebesar 80%. Untuk meningkatkan presisi dibutuhkan penggunaan data latih yang lebih banyak dan variasi kata yang lebih beragam.

Penelitian ini bisa dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil yang lebih baik.

REFERENSI

- [1] Liu, Bing, "Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers".2012
- [2] Liu, Bing. *Sentiment Analysis: A Multi-Faceted Problem*. University of Illinois at Chicago.2010
- [3] Saraswati, N. W, S, "Naïve Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis", SESINDO, 2013
- [4] Annur, H., "Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes," *ILKOM Jurnal Ilmiah*, Vol. 10, No. 2, Hal.. 160-165, ISSN: 2548-7779, 2018
- [5] Gustientiedina, Siddik M, Desnelita Y., Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis" *Jurnal Infomedia* Vol.2 No.4, ISSN: 2527-9858, 2019
- [6] Francis, L., and Flynn, M.2010. "Text Mining Handbook Casualty Actuarial Society E Forum". Spring. 2010
- [7] Nur, Yusuf. dan Santika, D. "Analisis Sentimen Pada Dokumen Berbahasa Indonesia Dengan Pendekatan Support Vector Machine". Universitas Bina Nusantara. Jakarta. November. 2011.
- [8] Prasetio, E., *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi, 2012.
- [9] Nazief, Bobby, Mirna Adriani, "Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia", *Fakulty of Computer Science University of Indonesia*
- [10] Wibisono Y., dan Khodra, M. L., *Clustering Berita Berbahasa Indonesia*, *Jurnal FPMIPA UPI*. 2005
- [11] J. Han, J., Kamber, M., Pei, *Data Mining Concept And Techniques*, III ed. Amerika: Library of Congress Cataloging, 2012.