

Analisis Sentimen terhadap Opini Mahasiswa pada Survei Kepuasan terhadap Fasilitas dan Layanan Universitas Telkom dengan menggunakan Algoritma Naïve Bayes

Sentiment Analysis of Student Opinion on the Satisfaction Survey of Telkom University Facilities and Services using the Naïve Bayes Algorithm

Astri Asroviana Putri¹, Jondri², Rian Febrian Umbara³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹asroviana@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id,

³rianfebrianumbara@telkomuniversity.ac.id

Abstrak

Pada penelitian ini, dilakukan analisis sentimen dalam menentukan opini mahasiswa pada survei kepuasan terhadap fasilitas dan layanan Universitas Telkom. Opini tersebut sangat banyak dan tidak terstruktur. Maka dalam pengolahannya, diperlukan analisis sentiment dalam mengklasifikasikan opini mahasiswa. Ada tiga sentiment yang akan diklasifikasikan yaitu, sentiment positif, sentiment netral, dan sentiment negatif. Metode klasifikasi yang akan digunakan yaitu Naive Bayes Classifier.

Pada penelitian ini, data melewati tahap preprocessing menggunakan stemming dan stopword dan pembobotan kata menggunakan metode TF-IDF serta melakukan klasifikasi menggunakan algoritma Naive Bayes. Penelitian ini menghasilkan nilai rata-rata precision untuk kelas positif sebesar 52.51% dan kelas negatif sebesar 69.25%. Sedangkan nilai rata-rata recall untuk kelas positif sebesar 76.78% dan kelas negatif sebesar 61.05%

Kata kunci : Naïve bayes classifier, NLP, TFIDF

Abstract

In this journal, we will classify sentiments in determining student opinion in the satisfaction survey of Telkom University facilities and services. The opinion is very large and unstructured. It requires an analysis of sentiment in classifying student opinions. Where there are two sentiments that will be classified, positive sentiment and negative sentiment. The classification method that will be used is Naive Bayes Classifier.

In this study, the data passed the preprocessing stage using stemming and stopword and weighting the word using the TF-IDF method and classifying it using the Naive Bayes algorithm. This study resulted in an average precision value for the positive class of 52.51% and a negative class of 69.25%. While the average recall value for the positive class is 76.78% and the negative class is 61.05%

Keywords: Naïve bayes classifier, NLP, TFIDF

1. Pendahuluan

1.1 Latar Belakang

Dalam meningkatkan kualitas dari suatu Universitas, kuesioner kepuasan mahasiswa menjadi salah satu alat penting dalam mendukung keberhasilan suatu Universitas dalam meningkatkan pelayanannya dan fasilitasnya. Mahasiswa menjadi elemen dalam memberikan kritik dan saran bagi universitas tersebut. Kuesioner biasanya berupa pertanyaan yang dapat dijawab dengan mengisi di kolom yang tersedia. Kuesioner diberikan kepada sejumlah mahasiswa.

Beberapa kuesioner disebarkan kepada ribuan orang. Agar kuesioner dapat dimanfaatkan, perlu adanya metode untuk mengubah data kuesioner menjadi data yang bermanfaat. Metode yang dapat menyelesaikan permasalahan tersebut yaitu *Text Classification*. Klasifikasi (*Classification*) ialah tugas memilih label kelas yang tepat jika diberikan sebuah input[1]. Salah satu cabang penelitian pada *Text Classification* yaitu Sentimen Analisis yang merupakan metode untuk menganalisis suatu topik tertentu. Salah satu metode klasifikasi yang akan digunakan yaitu Naive Bayes *classifier* yang termasuk *supervised learning*. Pada supervised learning, digunakan data training untuk proses latihan. Pada saat prediksi, maka test data atau dokumen yang ingin diklasifikasikan akan dikenal menggunakan model *classifier* yang telah dihasilkan agar dapat ditentukan masuk ke kelas mana dokumen tersebut. Teknik *supervised learning* biasanya sangat buruk dalam menganalisis sentimen. Dalam melatih suatu data, metode Naive Bayes yang lebih berbobot dapat memperlihatkan hasil yang sangat baik ([8],[9]).

Berdasarkan penelitian yang telah dilakukan (Shuo Xu, 2018) didapatkan bahwa metode Multinomial Naive Bayes sangat efektif dalam klasifikasi teks. Dengan semakin bertambahnya jumlah data training, nilai akurasi semakin tinggi. Pada penelitian tersebut, ekstraksi fitur yang digunakan yaitu metode TFIDF. Maka, pada Tugas Akhir ini penulis akan mengklasifikasikan sentimen opini mahasiswa pada survei kepuasan terhadap fasilitas dan layanan Universitas Telkom menggunakan algoritma Naive Bayes dan berdistribusi Multinomial. Lalu, menggunakan metode TFIDF sebagai ekstraksi fitur.

Salah satu metode klasifikasi yang akan digunakan yaitu Naive Bayes *classifier* yang termasuk *supervised learning*. Pada supervised learning, digunakan data training untuk proses latihan. Pada saat prediksi, maka test data atau dokumen yang ingin diklasifikasikan akan dikenal menggunakan model *classifier* yang telah dihasilkan agar dapat ditentukan masuk ke kelas mana dokumen tersebut.

1.2 Topik dan Batasannya

Permasalahan yang akan diselesaikan pada penelitian ini yaitu bagaimana mengklasifikasikan dan menganalisis sentimen yang positif, sentimen yang negative, dan sentimen netral pada survei kepuasan terhadap fasilitas dan layanan Universitas Telkom dengan metode Naive Bayes dan bagaimana kinerja sistem klasifikasi yang dibangun. Penulis hanya berfokus mengklasifikasikan sentimen positif, negatif, dan netral. Data yang dipakai merupakan data isian saran dari semua Mahasiswa dari berbagai Fakultas pada Universitas Telkom yang berupa survei kepuasan terhadap fasilitas dan layanan Universitas Telkom.

1.3 Tujuan

Adapun tujuan dalam melakukan penelitian ini yaitu menghasilkan sistem yang dapat mengklasifikasikan sentimen positif dan sentimen negatif pada survei kepuasan terhadap fasilitas dan layanan Universitas Telkom dengan metode Naive Bayes.

2. Studi Terkait

2.1 Preprocessing

Analisis Sentimen, atau yang dikenal juga dengan *Opinion Mining* adalah ilmu yang menganalisis opini seseorang, sentimen, perilaku, penilaian dan emosi terhadap suatu entitas, yaitu seperti produk, pelayanan, organisasi, dan individu. Terdapat beberapa jenis lainnya yang memiliki tugas yang sedikit berbeda seperti *opinion extraction*, *sentiment mining*, *subjectivity analysis*, *emotion analysis*, *review mining*[2]. Metode analisis sentimen digunakan untuk mengklasifikasikan sentimen sentimen yang memiliki aspek tertentu. Bisa berupa hal yang mengandung kata kata positif maupun negatif.

Tahap *preprocessing* merupakan tahap *text processing* terhadap kalimat opini dalam kuesioner. Ada dua tahap yang akan dilakukan dalam tahap *preprocessing* ini yaitu *stopword* dan *stemming*. *Stopword* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Sedangkan *stemming* adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya.

2.2 Metode TF-IDF

Pada penelitian ini, penghitungan bobot dokumen diolah dalam bentuk vektor dengan tujuan untuk mengurangi bobot dari kata yang muncul di banyak interval. Kata – kata ini biasanya kata yang sedang populer muncul dan tidak menunjukkan topik apapun namun tidak termasuk dalam daftar *stopwords*. Penghitungan bobot dokumen yang diolah dalam bentuk vektor dengan *term* (istilah) yang berhasil dikenali perhitungannya berdasar metode TF-IDF yang merupakan integrasi metode TF-IDF yang merupakan integrasi *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).

TF adalah algoritma pembobotan heuristik yang menentukan bobot dokumen berdasarkan kemunculan *term*. Semakin sering sebuah istilah muncul, semakin tinggi bobot dokumen untuk istilah tersebut, dan sebaliknya. Lalu, IDF merupakan banyaknya istilah tertentu dalam keseluruhan dokumen yang dijelaskan dalam rumus yaitu :

$$w(i, j) = tfidf(d, t_j) = Nd_i, t_j \cdot \log \frac{|C|}{Nt_j} \quad (1)$$

Keterangan :

Nd_i, t_j = jumlah term t_j terjadi dalam dokumen d_i

Nt_j = jumlah dokumen di dalam C koleksi

2.3 Naive Bayes Classifier

Dalam algoritma Naive Bayes Classifier setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_i adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori opini. Pada saat klasifikasi, algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan, dimana aturan Naive Bayes untuk klasifikasi adalah sebagai berikut:

$$c = \arg \max_{V_j} \left(\frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \quad (1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kelas (V_j) sehingga persamaan dapat ditulis sebagai Persamaan (2).

$$c = \arg \max_{V_j} (P(x_1, x_2, \dots, x_n | V_j) P(V_j)) \quad (2)$$

Persamaan diatas dapat disederhanakan menjadi Persamaan (3).

$$c = \arg \max_{V_j} P(V_j) \prod_{i=1}^n (P(x_i | V_j)) \quad (3)$$

Keterangan :

V_j = Kelas Opini

$x_1, x_2, x_3, \dots, x_n$ = Kata kata yang terdapat pada data

$P(x_i | V_j)$ = Probabilitas x_i pada kelas V_j (Probabilitas kata kata pada tiap kelas opini)

$P(V_j)$ = Probabilitas dari V_j (Probabilitas dari tiap kelas opini)

Untuk $P(V_j)$ dan $P(x_i|V_j)$ dihitung saat pelatihan sesuai Persamaan (4) dan Persamaan (5).

$$P(V_j) = \frac{\sum_{i=1}^{|D|} P(V_j|d_i)}{|D|} \quad (4)$$

$$P(x_i|V_j) = \frac{\sum_{i=1}^{|D|} N_{t,i} \cdot P(V_j|d_i) + 1}{\sum_{i=1}^{|V|} \sum_{i=1}^{|D|} N_{t,i} \cdot P(V_j|d_i) + |V|} \quad (5)$$

Dimana $P(V_j|d_i)$ dapat diestimasi sebagai persamaan berikut, persamaan (6)

$$P(V_j|d_i) \propto P(V_j) \prod_{i=1}^n (P(x_i|V_j)) \quad (6)$$

Keterangan :

$|D|$ = Jumlah data

$|V|$ = Jumlah kata

$N_{t,i}$ = Banyak kata yang muncul pada suatu dokumen

d_i = Dokumen ke-n

$P(V_j|d_i)$ = Probabilitas V_j terhadap d_i (Probabilitas dari setiap kelas opini di setiap dokumen yang diujikan).

2.4 Confusion Matrix

Tahap evaluasi dilakukan dengan membuat *confusion matrix* dan menghitung *Recall* dan *Precision*.. Untuk proses evaluasi ini, dimana akan memberikan akurasi atau performa teks sentimen yang telah dilakukan. Dengan menggunakan *confusion matrix* yang dimana akan memudahkan dalam mendapatkan tingkat akurasi dari klasifikasi sentimen ini. Cara membuat *confusion matrix* yaitu dengan menghitung *Recall* dan *Precision* terlebih dahulu. Dengan cara seperti berikut :

$$Recall = \frac{TP}{TP+FN} = \frac{TP}{Predicted Results} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} = \frac{TP}{Actual Results} \quad (2)$$

$$Accuracy = \frac{True Positive + True Negative}{Total} \quad (3)$$

Keterangan :

TP : True Positive

FP : False Positive

FN : False Negative

2.5 F-Measure

F-measure dimaksudkan untuk menggabungkan *Recall* dan *Precision* menjadi satu ukuran yang efektif dalam pencarian akurasi yang tepat dan dapat dirumuskan seperti berikut :

$$F - Measure = \frac{2RP}{R+P} \quad (1)$$

Keterangan :

R = Recall ,

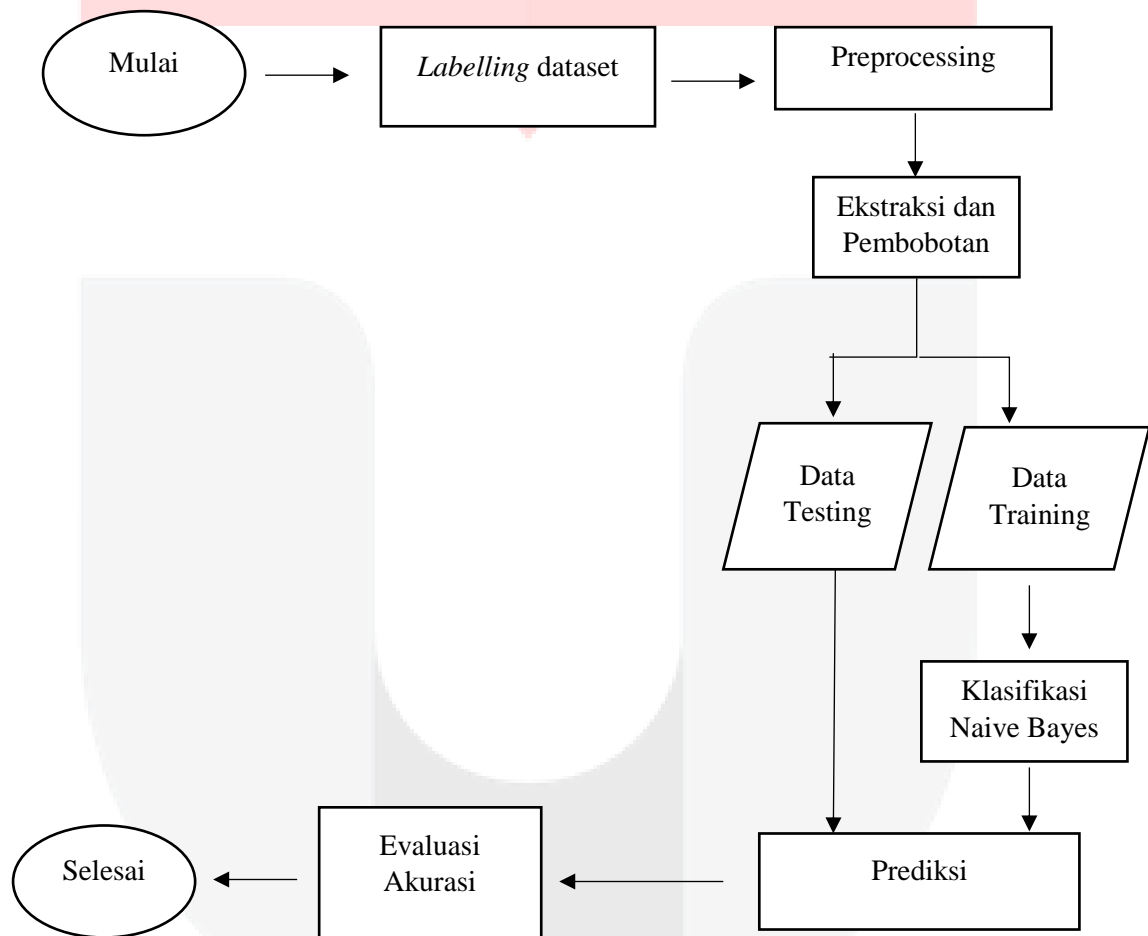
$P = Precision$

3. Sistem yang dibangun

3.1 Flowchart

Pada bagian ini, model *classifier* yang akan digunakan untuk mengklasifikasikan sentimen yaitu Naive Bayes *classifier*. Sebelum klasifikasi sentimen, data harus melalui tahap preprocessing terlebih dahulu. Dimana pada tahap preprocessing terdapat 2 tahap yaitu : *stemming* dan *stopword*. Setelah melalui tahap preprocessing, data melalui tahap seleksi fitur. Seleksi fitur yang akan digunakan yaitu Metode TF-IDF. Data yang telah melalui tahap seleksi fitur dinamakan data training.

Lalu, data akan dimasukkan ke dalam model *classifier* untuk diklasifikasi sentimennya yang berupa sentimen positif dan sentimen negatif. Setelah itu, dilakukan tahap evaluasi yang bertujuan untuk mengukur kinerja dari model *classifier* yang telah digunakan. *Precision* dan *recall* akan sering digunakan pada tahap evaluasi dalam pembuatan *confusion* matriks dan pengukuran F-measure.



Gambar 3.1. Alur Kerja Analisis Sentimen

3.2 Dataset

Dataset yang akan digunakan yaitu data survei kepuasan terhadap fasilitas dan layanan di Universitas Telkom. Jumlah data yang diproses yaitu sebanyak 20000 data. dimana, rentang waktu pengambilan data yaitu di tahun 2017 – 2018. Data ini tersimpan dalam bentuk *.csv*. kemudian data yang digunakan

diberikan label positif, negatif, ataupun netral secara manual. Label positif sebanyak 7232, Label negatif sebanyak 9364, dan label netral sebanyak 3402.

Label	Komentar	Inisial
Positif	Terus lanjutkan	1
Netral	tidak ada	0
Negatif	pelayanan kurang baik	-1

Tabel 3.1. Contoh data yang diberi label

3.3 Preprocessing

Proses *processing* dilakukan untuk mengolah data dengan tujuan untuk membersihkan data. Berikut tahap tahap yang dilakukan dalam proses *preprocessing* ini, yaitu sebagai berikut :

a. Stemming

Proses *stemming* ini digunakan untuk mengubah kata berimbuhan menjadi kata dasar. Data yang memiliki kata berimbuhan akan diubah menjadi kata kata dasar. Contoh, kata membetulkan diubah menjadi kata betul.

b. Stopword

Proses *Stopword* berguna untuk menghilangkan kata kata yang dianggap tidak memiliki makna. *Stopword* memiliki kosakata umum yang biasanya muncul dalam jumlah yang besar dan tidak memiliki makna. Contohnya preposisi seperti “di”, “ke”, dll.

3.4 Pembobotan Kata

Pembobotan kata dilakukan menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Tujuan dari perhitungan TF-IDF ini adalah untuk memberikan bobot pada data berdasarkan seberapa besar frekuensi kemunculan kata tersebut dalam dokumen. Pada proses ini akan dilakukan perhitungan untuk mencari nilai TF (*Term Frequency*) untuk mengetahui jumlah keberadaan sebuah *term* (istilah) didalam sebuah dokumen. Kemudian dilakukan perhitungan df (*document frequency*) untuk mengetahui banyaknya dokumen dimana sebuah term tersebut muncul. Selanjutnya perhitungan IDF (*Inverse Document Frequency*) yang berfungsi untuk mengurangi bobot sebuah term jika kemunculannya banyak tersebar pada setiap dokumen yang dimiliki.

3.5 Klasifikasi Sentimen menggunakan Naive Bayes Classifier

Pada proses ini dilakukan untuk mencari nilai probabilitas tertinggi dari data uji dalam kategori yang paling tepat. Pada penelitian ini, distribusi multinomial menjadi acuan dalam algoritma Naive Bayes. Karena, distribusi multinomial digunakan untuk penentuan probabilitas hasil yang dikategorikan ke dalam lebih dari dua kategori. Ada beberapa tahap yang dilakukan yaitu menghitung probabilitas prior, menghitung probabilitas kondisional, dan menghitung posterior untuk pemilihan kelas.

3.6 Evaluasi

Evaluasi dilakukan untuk menguji hasil dari klasifikasi dengan menggunakan *confusion matrix* dan *F-measure*. *Confusion matrix* dilakukan dengan menghitung nilai *Recall* dan *Precision*. Sedangkan,

F-measure menggabungkan nilai *Recall* dan *Precision* untuk mendapatkan nilai akurasi tepat.

4. Hasil dan Evaluasi

Setelah melakukan tahap klasifikasi menggunakan algoritma Naive Bayes selanjutnya dilakukan tahap evaluasi menggunakan *confusion matrix* dan *F-measure*. Pada tahap *confusion matrix* dilakukan dengan partisi data. Hasil pengujian akurasi didapatkan dengan membagi partisi data hasil klasifikasi data *training* dan data *testing*. Berikut hasil pengujian klasifikasi menggunakan *confusion matrix* :

		Truth data				
		Class 1	Class 2	Class 3	Classification overall	Producer Accuracy (Precision)
Classifier results	Class 1	3503	1531	1468	6502	53.876%
	Class 2	0	0	0	0	No data
	Class 3	1002	14	2293	3309	69.296%
	Truth overall	4505	1545	3761	9811	
User Accuracy (Recall)		77.758%	0%	60.968%		

Overall accuracy (OA): 59.077%

Kappa¹: 0.277

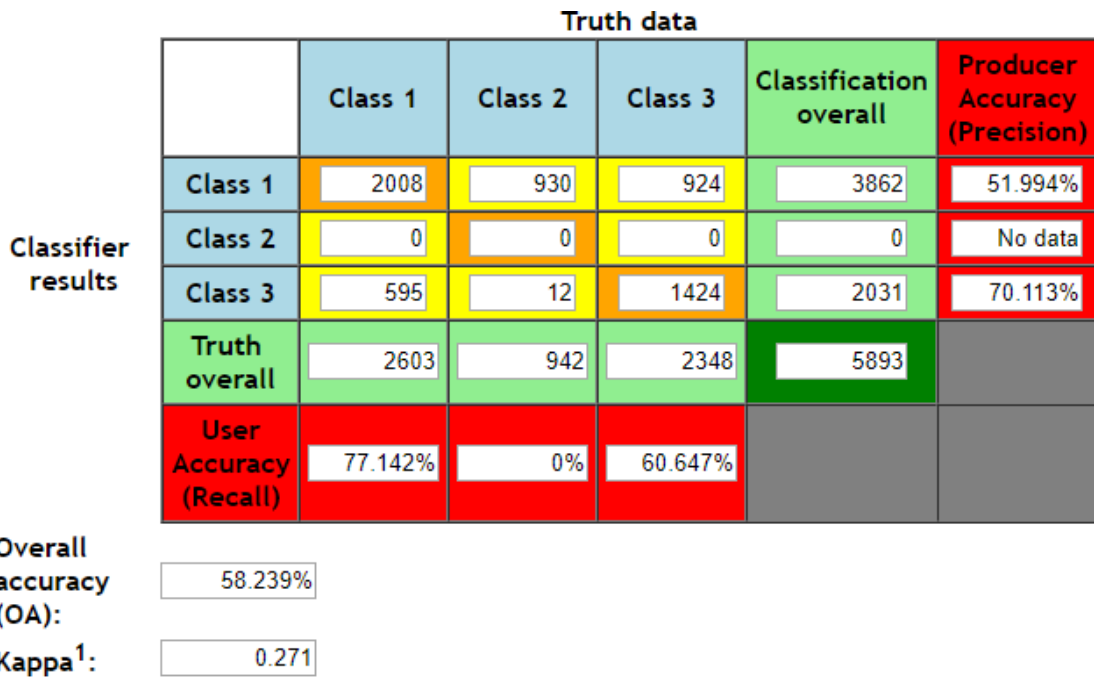
Gambar 4.1 Hasil Perhitungan Confusion Matrix dari perbandingan jumlah data training dan data testing 50:50

		Truth data				
		Class 1	Class 2	Class 3	Classification overall	Producer Accuracy (Precision)
Classifier results	Class 1	2783	1210	1155	5148	54.06%
	Class 2	0	0	0	0	No data
	Class 3	840	13	1862	2715	68.582%
	Truth overall	3623	1223	3017	7863	
User Accuracy (Recall)		76.815%	0%	61.717%		

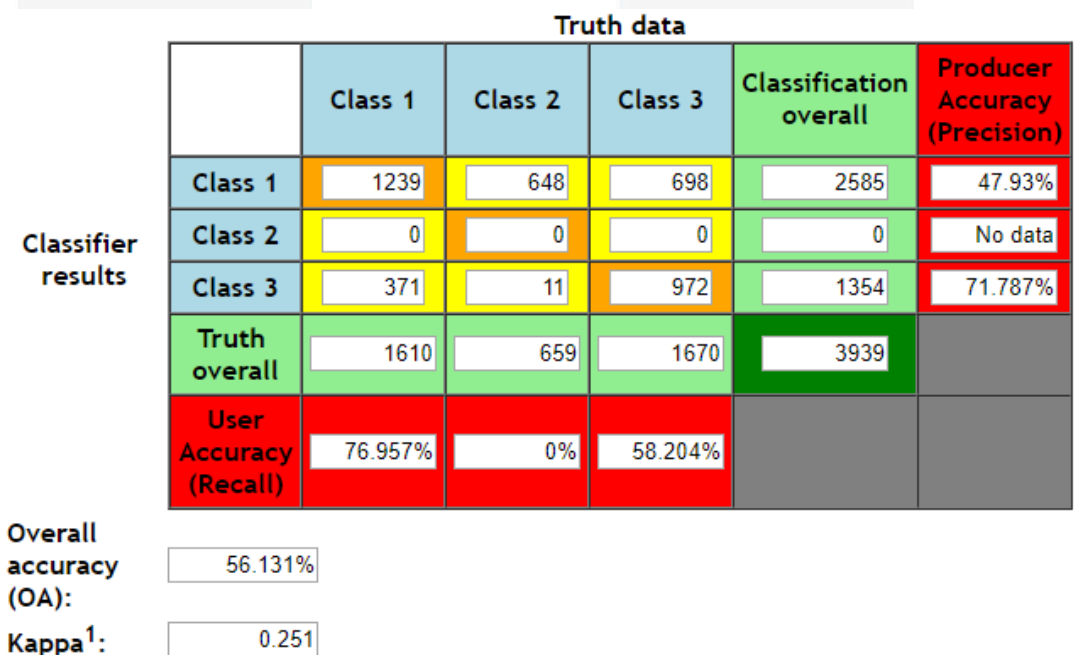
Overall accuracy (OA): 59.074%

Kappa¹: 0.277

Gambar 4.2 Hasil Perhitungan Confusion Matrix dari perbandingan jumlah data training dan data testing 60:40



Gambar 4.3 Hasil Perhitungan Confusion Matrix dari perbandingan jumlah data training dan data testing 70:30



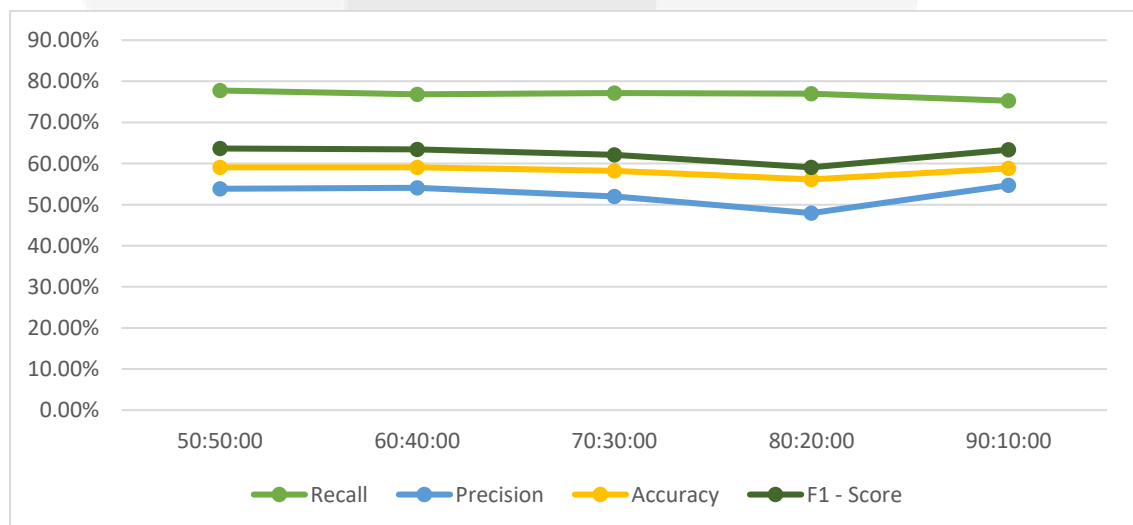
Gambar 4.3 Hasil Perhitungan Confusion Matrix dari perbandingan jumlah data training dan data testing 80:20

		Truth data				
		Class 1	Class 2	Class 3	Classification overall	Producer Accuracy (Precision)
Classifier results	Class 1	697	316	261	1274	54.71%
	Class 2	0	0	0	0	No data
	Class 3	229	2	458	689	66.473%
	Truth overall	926	318	719	1963	
User Accuracy (Recall)		75.27%	0%	63.7%		
Overall accuracy (OA):		58.839%				
Kappa ¹ :		0.272				

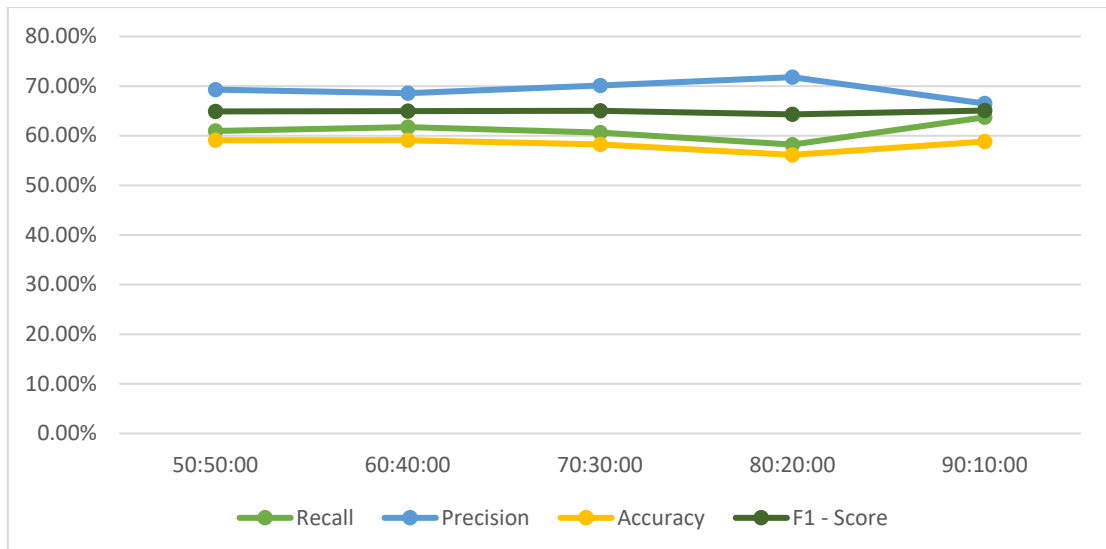
Gambar 4.3 Hasil Perhitungan Confusion Matrix dari perbandingan jumlah data training dan data testing 90:10

Dari perhitungan tersebut, *class 1* mewakili kelas positif, *class 2* mewakili kelas netral, dan *class 3* mewakili kelas negatif. Kelas netral tidak memiliki hasil dikarenakan pada kelas netral terdapat banyak data kosong (*no text*). Sistem tidak akan mengklasifikasikan data kosong tersebut ke kelas positif maupun kelas negatif. Maka, sistem akan mengabaikan data kosong tersebut. Sehingga, kelas netral tidak memiliki nilai *recall* dan *precision*.

Pada kelas positif, nilai *recall* lebih tinggi daripada nilai *precision*. Hal ini berarti lebih sedikit *false negative*. Sedangkan pada kelas negatif, didapatkan nilai *precision* lebih tinggi daripada nilai *recall*. Hal ini berarti lebih banyak *false negative*. Perbedaan seperti ini, disebabkan oleh tidak seimbangnya jumlah data positif dan data negatif. Jumlah data negatif lebih besar dibandingkan dengan jumlah data positif. Maka, nilai akurasi menjadi rendah dan hasil prediksi cenderung ke kelas negatif. Berikut perbandingan nilai *recall*, *precision*, *F1-Score*, dan akurasi pada kelas positif maupun negatif.



Gambar 4.5 Perbandingan nilai *recall*, *precision*, *F1-Score*, dan akurasi pada kelas positif

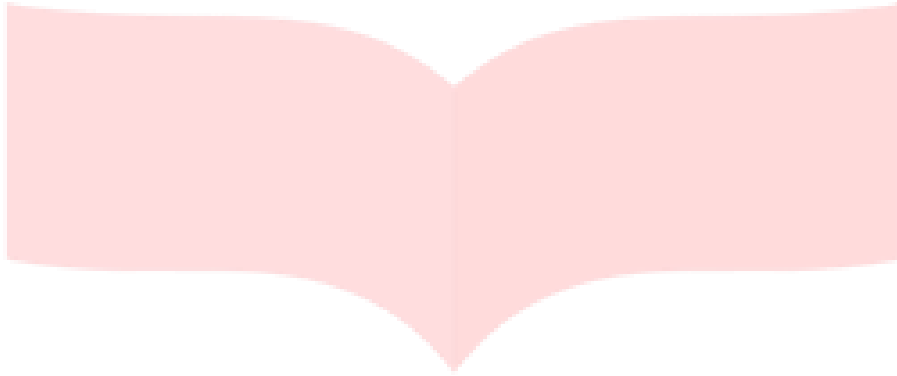


Gambar 4.6 Perbandingan nilai *recall*, *precision*, *F1-Score*, dan akurasi pada kelas negatif

Berdasarkan grafik tersebut, dapat dikatakan semakin banyak data training, maka semakin tinggi nilai akurasi. Sistem mengalami proses pembelajaran dari waktu ke waktu sehingga ketepatan dari sebuah sistem akan semakin tinggi.

5. Kesimpulan

Banyaknya data positif dan negatif yang tidak seimbang menyebabkan kecenderungan ke kelas tertentu. Pada penelitian ini, data negatif lebih banyak jumlahnya. Data negatif yang dilabel secara manual sebanyak 9364 dan data positif sebanyak 7232. Sehingga, hasil prediksi lebih cenderung ke kelas negatif. Hal ini dikarenakan, algoritma naive bayes bersifat probabilistik. Sehingga, menyebabkan akurasi yang cukup rendah. Sebaiknya, data yang sudah dilabel secara manual diperiksa kembali ke balai bahasa untuk ditindak lanjuti kebenarannya. Dengan adanya pakar, jumlah data positif maupun negatif menjadi lebih seimbang. Sehingga, menghasilkan akurasi yang tinggi.



DAFTAR PUSTAKA

- [1] He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472.
- [2] B. Liu, "Sentiment Analysis and Opinion Mining," 2012.
- [3] Pustejovsky, J., Stubbs A. (2012). *Natural Language Annotation for Machine Learning*. Beijing: O'Reilly
- [4] Jurafsky, Daniel dan Martin, H. James. 2007. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, New Jersey
- [5] Kamaruzaman, S.M., Chowdhury M.R. 2004. Text Categorization using Association Rule and Naive Bayes Classifier. *Asian Journal of Information Technology*. Vol. 3(9):657-665
- [6] Liu, B. 2010. *Handbook of Natural Language Processing*, chapter Sentiment Analysis and Analysis, 2nd Edition. Chapman & Hall/CRC Pres, New York
- [7] Tan, Songbo., Cheng, Xueqi., Wang, Yuefen., dan Xu, Hongbo. (2009). *Adapting Naïve Bayes to Domain Adaptation for Sentiment Analysis*. Chinese Academy of Sciences.
- [8] Escudero, G., Marquez, L., and Rigau, G., "Naive Bayes and Exemplar-based Approaches to Word Sense Disambiguation Revisited," in *Proceedings of the 14th European Conference*

[9] Lewis, D., "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," in Proceedings of 10th European Conference on Machine Learning, Chemnitz, Germany, 1998, pp. 4- 15.

[10] NLProcessor – *Text Analysis Toolkit*. 2000.<http://www.infogistics.com/textanalysis.html>

[11] Shaltout, N. A., El-Hefnawi, M., Rafea, A. & Moustafa, A., 2014. Information Gain as a Feature Selection Method for the Efficient Classification of Influenza Based on Viral Hosts. London, U.K, WCE

[12] Xu, Shuo. Bayesian Naive Bayes classifiers to text classification. China : Journal of Information Science, 2018.