

## **Klasifikasi dan Analisa Sentimen Kuesioner Fasilitas dan Layanan untuk Universitas Qomaruddin Gresik**

**Muhammad Ajir Muzakki<sup>1</sup>, Gunawan<sup>2</sup>**

<sup>1,2</sup> Program Pasca Sarjana, Magister Teknologi Informasi

Institut Sain dan Teknologi Terpadu Surabaya

Jalan Ngagel Jaya Tengah 73 - 77 Surabaya

<sup>1</sup>ajirmz@gmail.com, <sup>2</sup>gunawan@stts.edu

### **Abstrak**

Universitas Qomaruddin merupakan salah satu lembaga pendidikan tinggi swasta yang selalu memprioritaskan atau mengutamakan kepuasan mahasiswanya sebagai pelanggan atau customer dengan memberikan pelayanan yang terbaik dan berkualitas. Dalam meningkatkan fasilitas dan layanan tiap semester dilakukan kuesioner untuk memperoleh umpan balik terkait pelayanan akademik, proses pembelajaran maupun sarana dan prasarana kampus. Bentuk kuesioner berupa pertanyaan dengan jawaban optional dan uraian. Pada data uraian memerlukan penanganan lebih lanjut, mengingat data uraian ini sangat memungkinkan mengungkap hal-hal yang belum terungkap melalui butir pertanyaan dengan jawaban optional yang jumlahnya terbatas. Tujuan Penelitian ini adalah melakukan klasifikasi terhadap data kuesioner mahasiswa dengan algoritma rocchio dan analisa sentimen dengan algoritma Naïve Bayes. Algoritma rocchio digunakan untuk klasifikasi pada kategori fasilitas dan layanan yaitu Sarana prasarana (sarpras), Biro Administrasi Akademik (BAA), Biro Administrasi Keuangan (BAK), Biro Sistem Informasi (BSI), Lain-lain. Sedangkan Algoritma Naïve bayes digunakan untuk analisa sentimen, yaitu positif, netral dan negative. Berdasarkan hasil uji coba yang telah dilakukan dengan 10 kali uji coba diperoleh, nilai akurasi untuk klasifikasi kategori fasilitas dan layanan memiliki nilai rata-rata 83.27 sedangkan untuk klasifikasi sentimen memiliki nilai rata-rata 80.67 sehingga dapat dikatakan performa sistem cukup stabil.

**Kata kunci**— Klasifikasi, Analisa Sentimen, Algoritma Rocchio, Algoritma Naïve Bayes

### **Abstract**

Qomaruddin University is a private collage that always prioritizes student satisfaction as customers by providing the best and qualified services. In improving facilities and services , each semester questionnaires are conducted to obtain feedback on academic services, learning processes and campus facilities and infrastructure. The form of the questionnaire is in the form of questions with optional answers and descriptions. In the descriptive data requires further handling, considering that this descriptive data makes it very possible to reveal things that have not been revealed through question items with a limited number of optional answers. The purpose of this research is to classify the student questionnaire data using the Rocchio algorithm and sentimen analysis using the Naïve Bayes algorithm. The Rocchio algorithm is used to classify facilities and services, namely Infrastructure (Sarpras), Academic Administration Bureau (BAA), Financial Administration Bureau (BAK), Information Systems Bureau (BSI), etc. The Naïve Bayes algorithm is used for sentimen analysis, namely positive, neutral and negative. Based on the results of trials that have been carried out with 10 trials obtained, the accuracy value for the classification of facilities and services category has an average value of 83.27 while for sentimen analysis has an average value of 80.67 so it can be said that the system performance is quite stable.

**Keywords**— Classification, Sentimen Analysis, Rocchio Algorithm, Naïve Bayes Algorithm

## 1. PENDAHULUAN

Dewasa ini kesadaran masyarakat akan pemenuhan kebutuhan terhadap pendidikan formal terutama pendidikan tinggi semakin meningkat, hal ini menjadikan perguruan tinggi sebagai sektor yang sangat strategis yang diharapkan untuk menghasilkan sumber daya yang berkualitas. Penjaminan mutu dan kualitas pendidikan di perguruan tinggi adalah sebuah proses penetapan dan pemenuhan standart mutu pengelolaan pendidikan tinggi secara konsisten dan berkelanjutan. Keadaan persaingan yang kompetitif antar perguruan tinggi menuntut lembaga pendidikan harus serius memperhatikan kualitas pendidikan dan kelembagaan sehingga lembaga pendidikan tersebut mampu unggul dalam persaingan tersebut (Taman et al., 2013). Hal ini harus menjadi fokus dari semua pihak yang terlibat di dalam pengelolaan lembaga pendidikan terutama perguruan tinggi, tak terkecuali Universitas Qomaruddin.

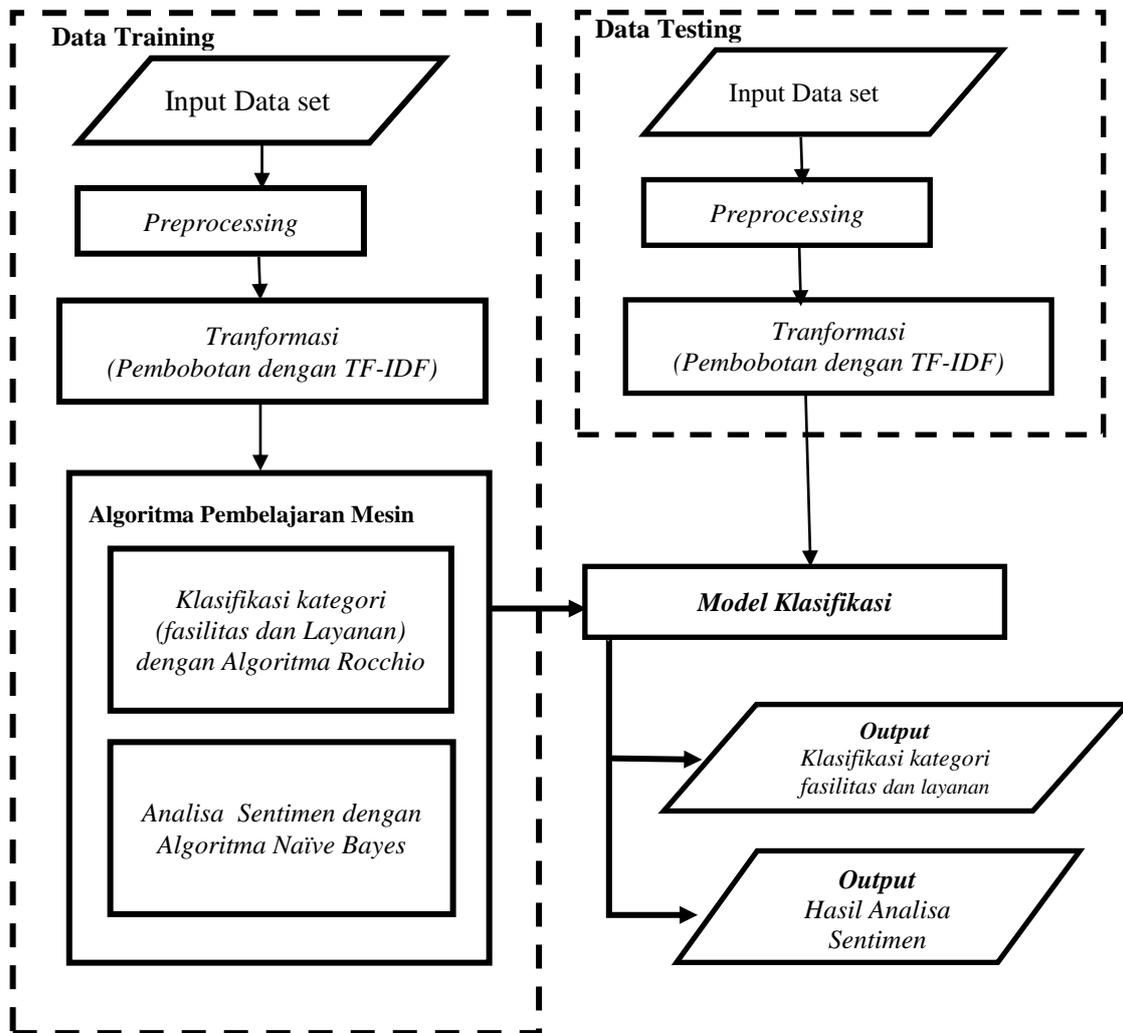
Universitas Qomaruddin (UQ) sebagai salah satu lembaga pendidikan tinggi swasta yang menerapkan sebuah konsep dimana selalu memprioritaskan atau mengutamakan kepuasan mahasiswanya sebagai pelanggan atau customer dengan memberikan pelayanan yang terbaik dan berkualitas serta terus meningkatkan kinerja organisasi UQ yang pada akhirnya diharapkan dapat meningkatkan loyalitas dari mahasiswanya sebagai pelanggan terhadap lembaga. Beberapa bidang atau bagian yang dikembangkan dan ditingkatkan secara berkesinambungan meliputi kurikulum program studi, proses pembelajaran, sumber daya manusia baik dosen, pegawai, teknisi, mahasiswa, sarana dan prasarana, lingkungan akademik, kerjasama dalam dan luar negeri.

Dalam tiap semester, UQ melakukan kuesioner terhadap mahasiswa untuk memperoleh umpan balik terkait pelayanan akademik, proses pembelajaran maupun sarana dan prasarana kampus. Kuesioner berisi pertanyaan dengan jawaban optional dan isian saran. Adapun data isian saran pada kuesioner dapat berupa saran, pendapat atau keluhan mahasiswa terkait fasilitas dan proses yang terdapat dalam UQ. Namun belum ada klasifikasi unit kerja yang jadi tujuan dari data kuesioner tersebut. Hal ini tentu perlu penanganan lebih lanjut, mengingat data saran ini sangat memungkinkan mengungkap hal-hal yang belum terungkap melalui butir pertanyaan yang jumlahnya terbatas, dalam rangka perbaikan manajemen Universitas Qomaruddin. Untuk mengungkap data isian saran pada kuesioner mahasiswa, akan dilakukan klasifikasi dan analisa sentimen. *Text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Mooney, 2006). Sedangkan analisa Sentimen menurut Feldman (Feldman, 2013) adalah suatu pekerjaan yang dimaksudkan untuk menentukan kecenderungan dari suatu opini yang terkandung dalam tulisan. Berdasarkan uraian tersebut di atas, dalam penelitian ini akan dilakukan klasifikasi kategori pada fasilitas dan layanan dari data kuesioner mahasiswa dengan menggunakan algoritma Rocchio dan analisis sentimen dari data kuesioner mahasiswa dengan menggunakan algoritma Naïve Bayes.

## 2. METODE PENELITIAN

### 2.1 Deskripsi Sistem

Sistem yang dibangun dalam penelitian ini adalah suatu sistem yang dapat digunakan untuk melakukan klasifikasi berdasarkan kategori fasilitas dan layanan dengan algoritma Rocchio, dan analisa sentimen dengan algoritma Naïve Bayes. Adapun gambaran umum dari sistem yang akan dibuat dapat dilihat pada gambar 1.



**Gambar 1.** Sistem Klasifikasi dan Analisa Sentimen

### 2.2 Data Set

Sistem dimulai dengan *input data set* yang diambil dari data kuesioner terhadap mahasiswa di Universitas Qomaruddin Gresik (UQG) melalui sistem informasi akademik UQG. Data diambil secara langsung dari *database* sistem informasi akademik UQG tanpa melalui proses *crawling*. *Data set* dalam penelitian ini merupakan kumpulan jawaban data kuesioner pada kolom isian yang dapat berupa saran atau pendapat sedangkan jawaban yang berbentuk opsional tidak digunakan dalam penelitian ini. Jumlah *data set* yang digunakan sebanyak 810 data. Adapun panjang jawaban kuesioner tidak lebih dari 500 karakter. Jumlah kuesioner pada masing-masing layanan 162 data, sedangkan masing-masing kategori sentimen sebanyak 270 data.

### 2.3 *Preprocessing*

*Preprocessing* bertujuan untuk mengolah data teks yang tidak terstruktur menjadi data yang terstruktur, memiliki kesamaan bentuk dan mengurangi jumlah kata, pendek kata *Preprocessing* didefinisikan merubah teks menjadi term index (Andriyani, 2019). Tahapan *Preprocessing* data penelitian meliputi :

2.3.1 *Cleansing*, pembersihan data dilakukan untuk menghilangkan karakter-karakter dan simbol yang dianggap tidak penting seperti tanda ‘.’, ‘,’,’:’. Dalam melakukan pembersihan data terlebih dahulu teks diubah menjadi huruf kecil semua, selanjutnya karakter dan simbol dibuang dari penyusun kata (*term*).

2.3.2 *Tokenizing*, proses pemenggalan setiap kata kalimat. Dalam tahap ini pemisahan kata dilakukan berdasarkan karakter spasi. Kata yang disusun oleh karakter angka akan dihilangkan.

2.3.3 *Filtering*, bertujuan untuk menghilangkan kata-kata yang kurang penting (*stopword*). Hal ini untuk menghindari kata yang sering muncul *stopword* bukan kata yang mewakili dokumen. Daftar kata termasuk dalam *stopword* disimpan dalam *stoplist*. Selanjutnya daftar *stoplist* tersebut digunakan sebagai acuan untuk mengeliminasi *stopword* dalam teks saran.

2.3.4 *Stemming*, proses untuk menemukan kata dasar dari setiap kata. Kata dasar diperoleh dengan membuang imbuhan yang menempel pada kata. Pada tahapan ini setiap kata dibandingkan dengan daftar kata dasar (*stem*), jika sama maka akan disimpan ke dalam *database*. Jika berbeda maka akan dicari kata dasarnya.

### 2.4 *Tranformasi*

Proses tranformasi data pada dasarnya bertujuan mengubah data dalam bentuk teks menjadi angka. Proses tranformasi data dilakukan dengan menghitung bobot setiap kata yang muncul hasil *preprocessing*. Penelitian ini menggunakan TF-IDF untuk menghitung bobot kata. Dalam TF-IDF bobot sebuah kata diperoleh dengan menghitung frekuensi kemunculann kata dalam sebuah dokumen, dengan mempertimbangkan frekuensi kemunculan kata di semua dokumen. Semakin sering kata muncul di banyak dokumen maka kata tersebut tidak unik dan tidak bisa digunakan sebagai kata yang mewakili sebuah dokumen.

### 2.5 *Klasifikasi Kategori Fasilitas dan Layanan dengan Algoritma Rocchio*

Algoritma *Rocchio* yaitu sebuah prosedur reformulasi *query* yang bertujuan membantu user dalam *information retrieval*. *Rocchio* melihat *feedback* sebagai suatu permasalahan dalam mencari *query* optimal, yaitu *query* yang mengoptimalkan selisih antara dokumen relevan dengan dokumen yang tidak relevan (Rizki, 2020). Dalam melakukan klasifikasi kategori fasilitas dan layanan pada penelitian ini akan menggunakan algoritma *Rocchio*. Setiap dokumen pada dataset akan diklasifikasikan ke kategori layanan Sarana prasarana (sarpras), layanan Biro Administrasi Akademik (BAA), layanan Biro Administrasi Keuangan (BAK), layanan Biro Sistem Informasi (BSI), dan lain-lain.

### 2.6 *Analisa Sentimen dengan Algoritma Naïve Bayes*

Dalam melakukan analisa sentimen pada penelitian ini akan menggunakan Algoritma Naïve Bayes. Naïve Bayes merupakan algoritma statistik, yang melaksanakan pengerjaan data terhadap numerik data dengan memakai probabilitas bayesian (Saputra et al., 2015). Setiap sentimen mahasiswa akan diklasifikasikan ke

kategori positif, netral atau negatif. Untuk mendapatkan model Naïve Bayes yang dapat melakukan klasifikasi, proses *training* akan dilakukan dengan menggunakan data *training*. Data *training* yang digunakan merupakan kumpulan dokumen sentimen mahasiswa yang dibagi ke dalam 3 kategori, yaitu positif, netral, dan negatif.

Proses *training* dengan algoritma Naïve Bayes dilakukan dengan tujuan untuk mendapatkan model yang dapat melakukan klasifikasi sentimen mahasiswa. Proses ini akan menghasilkan bobot untuk setiap *term* yang akan digunakan untuk proses pengujian. Model dari metode Naïve Bayes yang didapatkan dari proses *training* akan digunakan untuk klasifikasi kategori layanan pada dokumen sentimen mahasiswa. Model tersebut akan mengklasifikasi setiap dokumen sentimen mahasiswa ke dalam 3 kategori sentimen, yaitu positif, netral, dan negatif.

### 2.7 Evaluasi Kinerja Algoritma

*Confusion* matrik merupakan salah satu metode yang digunakan untuk mengukur kinerja dari suatu algoritma. *Confusion* Matrik dapat diartikan sebagai sebuah instrumen yang mempunyai fungsi untuk melaksanakan analisa apakah *classifier* tersebut benar dalam mengenali *tuple* dari kelas yang berbeda (Fibrianda dan Bhawiyuga, 2018). *Confusion matrix* dapat digunakan untuk menganalisa seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. TP (*True Positive*) dan TN (*True Negative*) memberikan informasi ketika klasifikasi benar, sedangkan FP (*False Positive*) dan FN (*False Negative*) memberikan informasi ketika klasifikasi salah.

**Tabel 1. Model Confusion Matrix**

		Prediksi	
		YES	NO
Aktual	YES	TP	FN
	NO	FP	TN

*Keterangan :*

- TP : Jumlah *true positive*
- TN : Jumlah *true negative*
- FP : Jumlah *false positive*
- FN : Jumlah *false negative*

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Preprocessing

*Preprocessing* bertujuan untuk mendapatkan dokumen yang lebih berkualitas. Tahapan *processing* dalam penelitian ini yaitu *cleansing*, *tokenizing*, *filtering*, dan *stemming*. Proses *Preprocessing* berhasil membuat dokumen yang awalnya masih mengandung banyak *noise* menjadi lebih bersih seperti yang terlihat di tabel 2.

**Tabel 2.** Hasil *Preprocessing*

No	Dokumen	
	Awal	Hasil <i>Preprocessing</i>
1	petugas BAA ada hari apa saja????	tugas BAA hari apa
2	Pelayanan BAA diharapkan bisa lebih baik lagi, bu.	layan BAA harap lebih baik bu
3	Percepat pelayanan BAA	cepat layan BAA
4	pelayanan BAA lebih efisien lagi	layan BAA lebih efisien
5	AC kurang dingin	AC kurang dingin

### 3.2 Klasifikasi dan Analisa Sentimen

Pada tahap ini dilakukan proses klasifikasi pada kategori fasilitas dan layanan dengan algoritma Rocchio, dan analisa sentimen dengan algoritma Naïve bayes. Sebelum dilakukan klasifikasi dan analisa sentimen, dataset yang sudah melewati proses *Preprocessing* dilakukan pembobotan TF-IDF. Kemudian *data set* dibagi menjadi dua, 70% untuk data latih dan 30% untuk data uji. Pada data latih dilakukan pelabelan secara manual. Hasil dari klasifikasi dapat dilihat pada tabel 3.

#### Segmen Program Proses Training dengan Rocchio

```
01: text_clf = Pipeline([('vect', CountVectorizer()),
02:                    ('tfidf', TfidfTransformer()),
03:                    ('clf', NearestCentroid()),
04:                    ])
05: text_clf.fit(train_X, train_Z)
```

#### Segmen Program Proses Training dengan metode Naïve Bayes

```
01: Naive = naive_bayes.MultinomialNB()
02: Naive.fit(Train_X_Tfidf,train_Y)
```

**Tabel 3.** Hasil Klasifikasi

No	Dokumen	Kategori		Sentimen	
		Fakta	Prediksi	Fakta	Prediksi
1	Mempercepat pelayanan BAA	BAA	BAA	-1	-1
2	Pelayanan BAA sangat tidak memuaskan semoga cepat sadar dan berubah	BAA	BAA	-1	-1
3	Saran saya, untuk yang meminta surat surat keterangan akademik supaya proses nya lebih di perpendek lagi.	BAA	lain	-1	0
4	BAK selalu ruwet validasi SPP	BAK	BAK	-1	1
5	Pelayanan ketika validasi setelah pembayaran SPP atau UKT terkadang sangat buruk mahasiswa/i harus antri dan terkadang banyak yg tdk tercatat oleh BAK di setiap prmbayaran nya.	BAK	BAK	-1	-1

### 3.3 Evaluasi Kinerja Algoritma

Untuk mengukur kinerja dari *classifier* digunakan *confusion* matrik. Berikut ini adalah hasil dari kinerja algoritma Rocchio dapat dilihat pada gambar 2.

	baa	bak	bsi	lain	sarpras
baa	49.0	0.0	0.0	1.0	1.0
bak	1.0	49.0	0.0	1.0	0.0
bsi	3.0	1.0	42.0	2.0	2.0
lain	1.0	0.0	4.0	35.0	11.0
sarpras	1.0	0.0	6.0	4.0	40.0

**Gambar 2.** *Confusion Matrix* Hasil Uji Coba Klasifikasi Fasilitas dan Layanan

Berdasarkan gambar 2 dapat diketahui bahwa kelas BAA dari data uji yang dilakukan percobaan memperoleh ketepatan prediksi benar 49 dari 51 apabila dihitung bentuk persentase sebesar 91,15%. Untuk kelas BAK prediksi tepat 49 dari 51 sehingga besar akurasinya 91,15%. Adapun kelas BSI diketahui prediksi tepat 42 dari 50 sehingga akurasinya sebesar 84%. Sedangkan kelas lain prediksi tepat 35 dari 51 sehingga persentase akurasi sebesar 68,63%. Kemudian kelas sarpras diketahui prediksi tepat 40 dari 51 sehingga besar akurasinya 78,43%. Seluruh prediksi tepat sebanyak 215 dibandingkan dengan *data testing* sebanyak 254 diperoleh rata-rata akurasi sebesar 84,65%. Hal ini menunjukkan bahwa sistem berhasil melakukan klasifikasi untuk kategori fasilitas dan layanan dengan baik.

Keberhasilan klasifikasi kategori fasilitas dengan algoritma Rocchio ini dapat dinilai baik karena memiliki nilai persentase kegagalan sebesar 15,35%. Nilai kegagalan ini dapat disebabkan karena kemungkinan adanya beberapa proses dalam *preprocessing* yang mungkin tidak harus dilakukan. Adapun kinerja dari algoritma Naïve Bayes dapat dilihat pada gambar 3.

	-1	0	1
-1	65.0	12.0	8.0
0	6.0	70.0	8.0
1	5.0	4.0	76.0

**Gambar 3.** *Confusion Matrix* Hasil Uji Coba analisa sentimen

Berdasarkan gambar 3 dapat diketahui bahwa kelas negatif setelah dilakukan percobaan memperoleh ketepatan prediksi benar 65 dari 85 apabila dihitung bentuk persentase sebesar 76,47%. Untuk kelas netral diperoleh prediksi tepat 70 dari 84 sehingga besar akurasinya 83,33%. Sedangkan kelas positif diketahui prediksi tepat 76 dari 85 sehingga akurasinya sebesar 89,41%. Seluruh prediksi tepat sebanyak 211 dibandingkan dengan *data testing* sebanyak 254 diperoleh rata-rata akurasi sebesar 83,07%. Hal ini menunjukkan bahwa sistem berhasil melakukan klasifikasi sentimen dengan baik.

Keberhasilan analisa sentimen dengan algoritma Naïve Bayes ini dapat dinilai baik karena memiliki nilai persentase kegagalan sebesar 16,93%. Nilai kegagalan ini dapat disebabkan karena kemungkinan adanya beberapa proses dalam *Preprocessing* yang mungkin tidak harus dilakukan. Selanjutnya proses pengujian sistem ini dilakukan sebanyak 10 kali untuk mendapatkan rata-rata nilai akurasi yang dapat dihasilkan oleh sistem. Hasil uji coba sistem dapat dilihat pada tabel 4.

**Tabel 4.** Akurasi 10 Kali Percobaan

No.	Klasifikasi Kategori Fasilitas dan Layanan dengan Rocchio	Analisa Sentimen dengan NB
1	84,65	83,07
2	85,83	82,28
3	82,28	80,71
4	82,28	81,1
5	83,07	77,56
6	81,1	81,89
7	85,04	82,68
8	83,07	79,13
9	81,89	81,1
10	83,46	77,17
<b>Rata-rata</b>	<b>83,267</b>	<b>80,669</b>

Dari tabel 4 dapat diketahui bahwa untuk klasifikasi fasilitas dan layanan dengan metode Rocchio diperoleh nilai akurasi paling kecil adalah 81,1% sedangkan nilai paling besar adalah 85,83%. Sedangkan untuk klasifikasi sentimen dengan metode Naïve Bayes diperoleh nilai akurasi paling kecil adalah 77,17% sedangkan nilai paling besar adalah 83,07%. Dengan demikian, performa sistem untuk klasifikasi sesuai dengan yang diharapkan. Dimana Nilai akurasi untuk Klasifikasi Kategori fasilitas dan layanan memiliki nilai rata-rata 83,27, sedangkan untuk klasifikasi sentimen memiliki nilai rata-rata 80,67.

#### 4. KESIMPULAN

Berikut kesimpulan dari hasil dan pembahasan yang telah dipaparkan sebelumnya :

1. Hasil pengujian menunjukkan bahwa sistem sukses menghasilkan model yang berisi *centroid* dari setiap kategori layanan dengan cukup baik.
2. Keberhasilan membuat model tersebut menyebabkan sistem dapat melakukan klasifikasi kategori layanan menggunakan metode Rocchio dengan cukup baik. Klasifikasi yang dilakukan oleh sistem berhasil mencapai rata-rata akurasi sebesar 83,27%.
3. Klasifikasi kategori layanan yang dihasilkan oleh sistem terbilang cukup stabil karena dengan 10 kali percobaan, sistem dapat menghasilkan klasifikasi dengan nilai akurasi paling kecil sebesar 81,1% dan akurasi paling besar sebesar 85,83%.
4. Pada 10 kali pengujian klasifikasi sentimen mahasiswa, sistem menunjukkan performa yang cukup bagus dengan rata-rata akurasi sebesar 80,67%.
5. Performa sistem untuk klasifikasi sentimen mahasiswa juga terbilang cukup stabil karena berhasil melakukan klasifikasi dengan akurasi paling kecil sebesar 77,17% dan akurasi paling besar sebesar 83,07%.
6. Selain kedua klasifikasi tersebut, sistem juga dapat melakukan analisa dan klasifikasi berdasarkan kategori layanan dan jenis sentimen mahasiswa. Hasil ini dapat menampilkan jumlah sentimen positif, netral, atau negative yang diperoleh

setiap kategori layanan. Hasil ini dapat dimanfaatkan oleh setiap unit kerja untuk mengetahui kepuasan mahasiswa terhadap layanan mereka.

## 5. SARAN

Berikut beberapa saran untuk pengembangan penelitian berikutnya :

1. *Preprocessing* di sistem ini hanya dapat menangani sentimen yang dari bahasa Indonesia. Oleh karena itu diperlukan perbaikan di bagian *Preprocessing* supaya sistem dapat melakukan *Preprocessing* untuk bahasa lain seperti bahasa Inggris yang saat ini sangatlah populer.
2. Ditambahkannya metode yang dapat memberikan kesimpulan berupa gambaran kondisi sentimen yang diperoleh setiap unit kerja. Sehingga dengan adanya fitur tersebut, setiap unit kerja dapat menggunakannya sebagai dasar pengambilan kebijakan untuk kedepannya.

## UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih banyak kepada pihak-pihak yang terlibat dalam penelitian ini terutama kepada keluarga Universitas Qomaruddin Gresik. Semoga menjadi amal baik dan bermanfaat kepada halayak ramai.

## DAFTAR PUSTAKA

- Andriyani, N.A., 2019. Optimasi Algoritma *K-Nearest Neighbor* dalam Mendeteksi Komentar *Spam* Berbahasa Indonesia pada Instagram Menggunakan *Convert Negation* dan *TF-IDF (Term Frequency-Inverse Document Frequency)* pada Tahap *Pre processing*.
- Feldman, R., 2013. Techniques and applications for sentiment analysis. *Communications of the ACM* 56, 82–89.
- Fibrianda, M.F., Bhawiyuga, A., 2018. Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN 2548, 964X.
- Mooney, J., 2006. *Machine Learning Text Categorization*. Austin: University of Texas.
- Rizki, R., 2020. Sistem Informasi Perpustakaan SMK Negeri 1 Indralaya Selatan Menggunakan Metode *Rocchio Relevance Feedback*. *Jurnal Teknologi Informasi Mura* 12, 76–89.
- Saputra, N., Adji, T.B., Permanasari, A.E., 2015. Analisis sentimen data presiden Jokowi dengan preprocessing normalisasi dan stemming menggunakan metode naive bayes dan SVM. *Jurnal Dinamika Informatika* 5.
- Taman, A., Sukirno, S., Sari, A.R., Setiawan, N., Pustikaningsih, A., 2013. Analisis Kualitas Pelayanan Terhadap Kepuasan Mahasiswa Pada Fakultas Ekonomi Universitas Negeri Yogyakarta. *Nominal: Barometer Riset Akuntansi dan Manajemen* 2, 99–111.