

# Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine

Evita Fitri<sup>1</sup>, Yuri Yuliani<sup>2</sup>, Susy Rosyida<sup>3</sup>, Windu Gata<sup>4</sup>

<sup>1</sup>Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya No.18, Kota Jakarta Pusat, (021)8005722, www.nusamandiri.ac.id, e-mail:  
[14002313@nusamandiri.co.id](mailto:14002313@nusamandiri.co.id)

<sup>2</sup>Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya No.18 Kota Jakarta Pusat, (021)8005722, www.nusamandiri.ac.id, e-mail:  
[14002311@nusamandiri.ac.id](mailto:14002311@nusamandiri.ac.id)

<sup>3</sup>Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya No.18 Kota Jakarta Pusat, (021)8005722, www.nusamandiri.ac.id, e-mail:  
[susy.sud@nusamandiri.ac.id](mailto:susy.sud@nusamandiri.ac.id)

<sup>4</sup>Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya No.18 Kota Jakarta Pusat, (021)8005722, www.nusamandiri.ac.id, e-mail:  
[windu@nusamandiri.ac.id](mailto:windu@nusamandiri.ac.id)

## ARTICLE INFO

Article history:

Received 28 May 2020  
Received in revised form 10 June 2020  
Accepted 29 July 2020  
Available online 31 July 2020

## ABSTRACT

The review of the users of one application is of great help to development in improving the quality of the application and may be the means for assessments that users feel satisfied or not. The study conducted a sentiment analysis of the Ruangguru application by testing the three classification models such as *Naive Bayes*, *Random Forest* and *Support Vectors Machine*. The study has yielded results that from Random Forest classification model 97,16% by using Cross Validation and an AUC score of 0.996. Then accuracy with the model of *Support Vector Machine* classification support results in accuracy rate of 96.01% to an AUC value of 0.543 and accuracy in the testing of Naive Bayes classification model was 94,16% of AUC score 0,999. This study shows that an increase in accuracy from previous studies of 7.16% with Random Forest's final cut as a Random Forest classification model with the best performance.

Keywords: Analisis Sentimen, Algoritma, Naive Bayes, Random Forest, Support Vector Machine

## 1. Pendahuluan

Pendidikan merupakan salah satu sektor yang dipengaruhi oleh perkembangan teknologi. Saat ini pesatnya kemajuan teknologi menjadi salah satu unsur penting dalam dunia pendidikan[1]. *Mobile learning* atau sering disebut juga sebagai *m-learning* didefinisikan sebagai *e-learning* melalui media perangkat komputasi *mobile* atau merupakan penyampaian bahan pembelajaran elektronik pada alat komputasi *mobile* yang bertujuan mempermudah penggunaannya untuk diakses dimana saja dan kapan saja[2]. Adapun aplikasi merupakan sebuah program yang dikembangkan untuk memenuhi kebutuhan pengguna[3].

Saat ini sudah banyak ketersediaan aplikasi belajar online atau situs website belajar online yang dapat diakses oleh setiap orang sebagai salah satu solusi untuk belajar bagi para

pelajar, kehadiran aplikasi bimbingan belajar online memberikan kemudahan serta terbilang cukup efektif dan efisien karna dapat dilakukan kapanpun dan dimana saja tanpa mengenal ruang dan waktu. Salah satu aplikasi tersebut ialah aplikasi belajar Ruangguru, aplikasi ini sebagai salah satu fasilitas bimbingan belajar yang cukup banyak digunakan khususnya dikalangan pelajar Indonesia dengan jumlah pengunduhan sebanyak 1-15 Juta lebih pengunduh. Aplikasi Ruangguru adalah salah satu aplikasi buatan dari PT. Ruang Raya Indonesia, perusahaan ini berdiri pada tahun 2014 dan bergerak dibidang pendidikan non formal. Ruangguru merupakan perusahaan teknologi terbesar di Indonesia yang berfokus pada layanan berbasis pendidikan dan telah memiliki lebih dari 15 juta pengguna serta mengelola 300.000 guru yang menawarkan jasa di lebih dari 100 bidang pelajaran. Ruangguru mengembangkan berbagai layanan belajar berbasis teknologi, termasuk layanan kelas virtual, platform ujian online, video belajar berlangganan, *marketplace* les privat, serta konten-konten pendidikan lainnya yang bisa diakses melalui web dan aplikasi Ruangguru[4].

Namun, saat ini bukan hanya satu atau dua aplikasi belajar online yang tersedia dan mudah digunakan, banyaknya aplikasi belajar online yang tersedia membuat pengguna lebih selektif dalam menentukan aplikasi apa yang cocok atau yang lebih baik untuk digunakan. Salah satu yang dapat mempengaruhi pengguna untuk memilih aplikasi tersebut dengan melihat ulasan-ulasan yang ada pada laman pengunduhan aplikasi tersebut. Ulasan-ulasan ini dapat menjadi salah satu bahan pertimbangan dalam penilaian sebuah aplikasi, apabila ulasan-ulasan ini dikumpulkan lalu diolah maka hasil tersebut akan dapat dijadikan sebuah kesimpulan aplikasi belajar online manakah yang memiliki sentimen terbaik.

Pada paper yang diteliti [5] bahwa analisa sentimen merupakan riset komputasional dari opini, sentimen dan emosi yang di ekspresikan secara tekstual. Adapun dalam analisis sentimen pada umumnya melakukan suatu proses pengelompokkan dari teks yang ada didalam baik dokumen, kalimat atau suatu pendapat sehingga menghasilkan suatu nilai yang mengartikan apakah dokumen, kalimat dan pendapat tersebut bernilai positif atau negatif.

Penelitian [6] dengan judul analisis sentimen *review* aplikasi ruangguru menggunakan algoritma *Support Vector Machine*. Penelitian ini menerapkan metode klasifikasi algoritma *Support Vector Machine*, dan dilakukan pengujian menggunakan kombinasi dari pembagian data latih dan data uji serta menggunakan sistem *K-Fold Cross Validation*. Adapun pembagian data training sebesar 60% dan data testing 40%. Hasil dari penelitian ini mendapatkan nilai akurasi tertinggi berada di kisaran 90%.

Penelitian selanjutnya [7] yaitu mengenai analisis sentimen dan *topic modelling* pada sebuah aplikasi belajar yaitu aplikasi belajar Ruangguru dengan menggunakan data dari komentar yang didapat pada website youtube. Pada penelitian ini pengumpulan data dilakukan dengan cara *scrapping* yang menggunakan aplikasi *ScrapeStrom* untuk memudahkan pengambilan data yang valid dan akurat dan data tersebut dianalisa dengan menggunakan aplikasi Orange. Hasil dari data yang telah diolah ialah mengkatagorikan seberapa jumlah respon atau komentar positif dan negatif atas pelayananan aplikasi tersebut, dan hasil dari *modelling topic* tersebut bahwa jumlah komentar positif lebih mendominasi bila dibandingkan dengan komentar negatif.

Berdasarkan uraian diatas, maka peneliti mengusulkan penelitian analisis sentimen terhadap *review* aplikasi belajar online yang pada kasus ini mengambil *sample* aplikasi Ruangguru dengan menerapkan metode klasifikasi *Naive Bayes*, *Random Forest* dan *Support Vector Machine*,

## 2. Metode Penelitian

### 2.1. Metode *Naive Bayes Classifier*

*Naive Bayes* Merupakan salah satu algoritma data mining yang penggunaannya mudah serta pemrosesannya memiliki waktu yang cepat, mudah diimplementasikan dengan struktur yang cukup sederhana dan memiliki tingkat efektifitas yg tinggi [8]. Klasifikasi *Naive Bayes* adalah salah satu pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*, klasifikasi ini didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* yang terbukti memiliki akurasi dan kecepatan yang tinggi saat diimplementasikan kedalam database dengan dataset yang besar[9].

Adapun persamaan dari dasar metoda pada teorema bayes tersebut sebagai berikut:

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

Keterangan:

- X : Data dengan *class* yang belum diketahui  
 H : Hipotesis data X yang merupakan suatu *class* spesifik  
 P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)  
 P(H) : Probabilitas hipotesis H (prior probabilitas)  
 P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H  
 P(X) : Probabilitas X

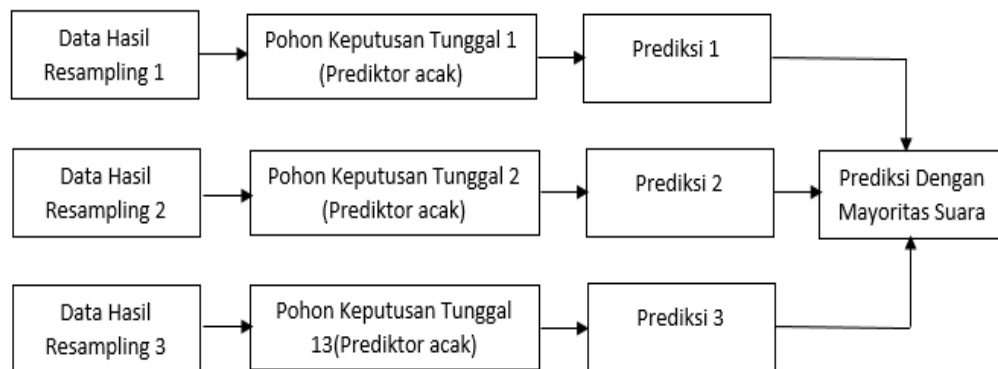
## 2.2. Metode *Random Forest*

Salah satu metode yang digunakan untuk pengklasifikasian dan regresi adalah *Random Forest*. Metode *Random Forest* merupakan sebuah *esamble* (kumpulan) metode pembelajaran menggunakan pohon keputusan sebagai *base classifier* yang dibangun dan dikombinasikan, adapun beberapa aspek penting dalam metode *Random Forest* diantaranya melakukan *bootstrap sampling* untuk membangun pohon prediksi, masing-masing pohon keputusan memprediksi dengan prediktor acak dan *Random Forest* sendiri melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan cara *majority vote* untuk klasifikasi dan juga rata-rata untuk regresi [10].

Adapun *Random Forest* dapat dibangun menggunakan *bagging* dengan pemilihan atribut acak. Metode *CART (Classification and Regression Tree)* sendiri dapat digunakan untuk menumbuhkan pohon keputusan, pohon keputusan tersebut tumbuh hingga ukuran maksimum dan tidak akan dipangkas sehingga dihasilkan kumpulan pohon yang kemudian disebut *forest*. Adapun kelebihan dari metode *Forest* sebagai berikut[11]:

1. Hasil akurasi bagus
2. Relatif kuat terhadap *outliers* dan noise
3. Lebih cepat dibandingkan dengan *bagging* dan *boosting*
4. Sifatnya yang sederhana dan mudah dipararelkan

Selanjutnya gambaran algoritma sederhana *Random Forest* dapat dilihat sebagai berikut[11]:



Gambar 1. Algoritma sederhana *Random Forest*

## 2.3. Metode *Support Vector Machine*

Pengklasifikasian menggunakan *Support Vector Machine* merupakan tehnik *machine learning* yang cukup populer untuk pengklasifikasian teks serta memiliki performa yang baik pada banyak domain serta mampu mengidentifikasi *hyperplane* secara terpisah diantara dua *class* yang berbeda sehingga hasilnya termaksimalkan dan dapat memaksimalkan pula jarak antara data yang paling dekat dengan *hyperplane* [12]. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara suatu kelas dengan kelas lain, *Support Vector Machine* melakukan pencarian nilai *hyperplane* dengan menggunakan *support vector* dan nilai *margin* [13].

Selanjutnya metode ini dapat digunakan sebagai salah satu metode yang handal dalam menyelesaikan masalah klasifikasi data, permasalahan tersebut dipecahkan dengan menyelesaikan persamaan *Lagrangian* yang merupakan bentuk dual dari *Support Vector Machine* melalui *quadratic programming*[14]. *Support Vector Machine* sendiri memiliki prinsip dasar linier *classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun pada metode ini telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi, pada ruang berdimensi tinggi akan dicari *hyperplane* yang dapat memaksimalkan jarak (margin) antara kelas data, adapun persamaannya sebagai berikut[15]:

Menurut Santosa, *Hyperplane* klasifikasi linier dapat dinotasikan:

$$f(x) = \mathbf{w}^T \mathbf{x} + b$$

sehingga menurut Vapnik dan Cortes diperoleh persamaan:

$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1 \text{ untuk } y_i = +1$$

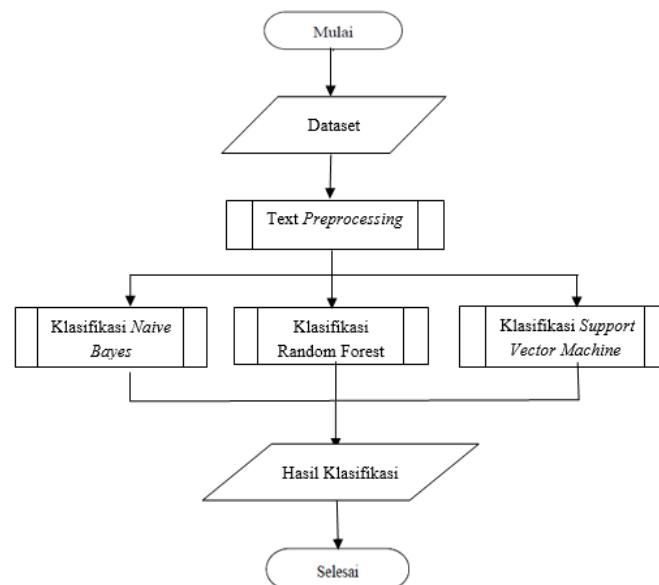
$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \leq -1 \text{ untuk } y_i = -1$$

Dengan  $x_i$  = himpunan data training ,  $i = 1, 2, \dots, n$  dan  $y_i$  = label kelas dari  $x_i$

Untuk mendapatkan *hyperplane* terbaik adalah dengan mencari *hyperplane* yang terletak di tengah-tengah antara dua bidang pembatas kelas dan untuk mendapatkan *hyperplane* terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda[15].

#### 2.4. Metode Usulan

Penelitian ini adalah sebuah eksperimen analisis sentimen terhadap *review* pelayanan aplikasi belajar online Ruangguru dengan model yang digunakan ialah klasifikasi *Naive Bayes*, *Random Forest* dan *Support Vector Machine*. Adapun tahapan pada penelitian analisis sentimen ini yaitu dengan mengumpulkan dataset dengan langkah awal memproses dataset dengan *text preprocessing* kemudian data tersebut di proses dengan model klasifikasi yang telah ditetapkan untuk mengetahui hasil akhirnya apakah sentimen tersebut masuk kedalam katagori sentimen positif atau sentimen negatif serta untuk mengetahui tingkat akurasi dan AUC (*Area Under Curve*) dari masing masing model klasifikasinya. Diagram alir secara umum dapat digambarkan sebagai berikut:



Gambar 2. Diagram Alir Sistem

#### 2.5. Pengumpulan Dataset

Pada penelitian ini, dataset didapat dari *review* aplikasi Ruangguru di Google Play Store. Pada proses ini data diambil menggunakan teknik *scrapping* menggunakan *web scrapping* dari website google play store. Data diambil sebanyak 1629 dengan bentuk *text review* dalam rentang waktu 01 Maret sampai dengan 30 Maret 2020. Data tersebut berjumlah 1629 sebelum dilakukannya *text preprocessing*.

## 2.6. Metode Pengolahan Dataset

Dataset yang telah diperoleh dari proses *scrapping* masih memiliki struktur data yang tidak terstruktur, data yang sembarang dan tidak beraturan. Oleh karena ini dibutuhkan tahap *preprocessing* data sebelum dataset diuji dengan sebuah model. Tahap ini dilakukan untuk membersihkan data dari noise dan mengubah data menjadi data yang terstruktur, adapun tahap dalam *preprocessing* data pada penelitian ini yaitu:

- Tokenize*, dimana dalam proses ini dilakukannya pemecahan kata pada kalimat *review*. Tahap ini juga menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks.
- Stemming*, dimana pada proses ini menemukan kata dasar dengan menghilangkan semua imbuhan yang menyatu pada kata.
- Case Folding*, dimana proses tersebut memanfaatkan fitur *transform cases* yang bertujuan untuk menyeragamkan seluruh teks kedalam huru kecil semua (*lowercase*).
- Stopword Removal*, dimana pada proses ini dilakukannya penghilangan kata yang termasuk kedalam kategori *stopword*. *Stopword* merupakan kata yang sering muncul namun dianggap tidak memiliki arti.

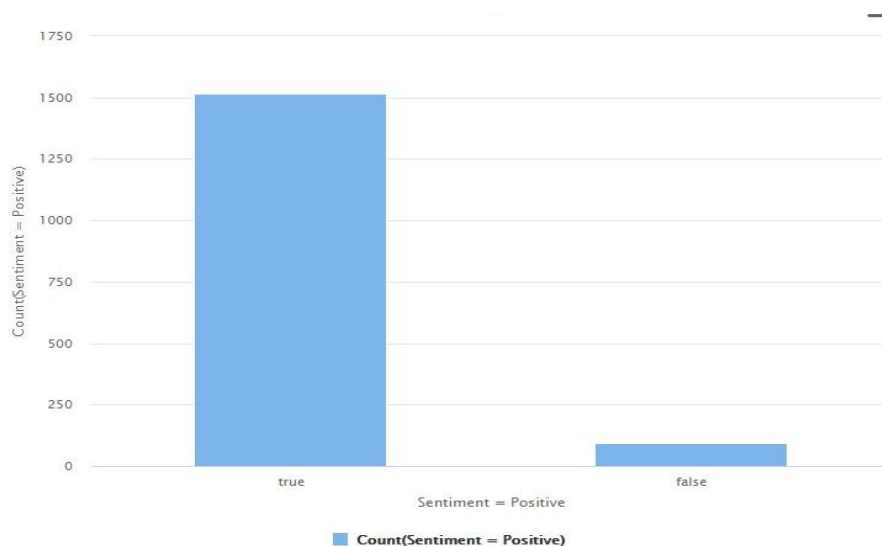


Gambar 3. Proses Pengolahan Dataset

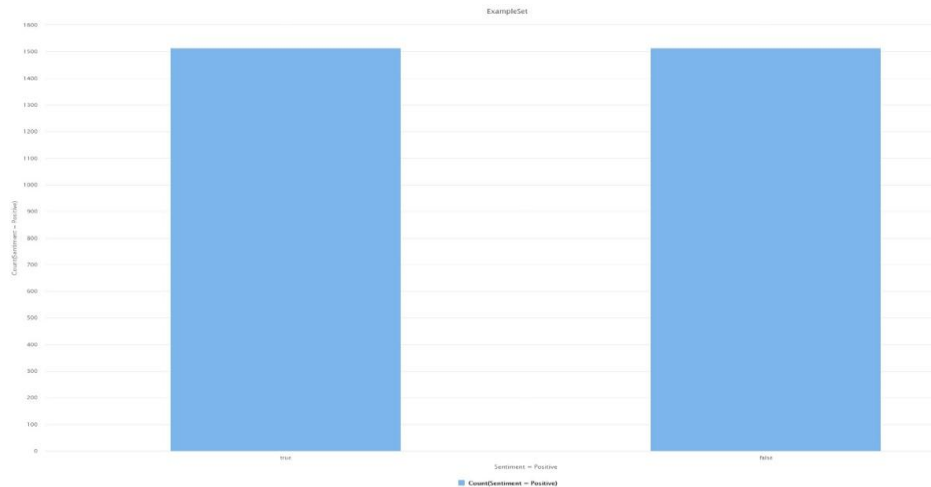
## 3. Hasil dan Analisa

### 3.1. Dataset

Diperoleh dataset dari *review* aplikasi Ruangguru di Google Play Store pada tahap *scrapping* berjumlah 1629 data dengan rentang waktu selama kurang lebih 30 hari pada tanggal 1 Maret 2020 hingga 30 Maret 2020 dan dilakukan *text preprocessing* pada dataset tersebut. Kemudian dilakukannya pelabelan atau pembuatan *class* pada dataset tersebut. Pada penelitian ini ditemukannya ketidak seimbangan pada *class* maka diterapkannya metode SMOTE untuk menangani ketidak seimbangan *class* tersebut sehingga dapat menghasilkan keakuratan dari proses klasifikasi yang dilakukan, pada penelitian ini setelah dilakukannya *text preprocessing* menghasilkan jumlah dataset sebanyak 1515 *instance*, adapun perbandingan data sebelum penerapan tehnik SMOTE dan sesudah dilakukannya penerapan tehnik SMOTE dapat dilihat pada gambar berikut:



Gambar 4. Dataset sebelum proses SMOTE



Gambar 5. Dataset setelah proses SMOTE

### 3.2. Pengujian

#### 3.2.1. Pengujian Menggunakan Klasifikasi *Naive Bayes* dengan *Cross Validation*

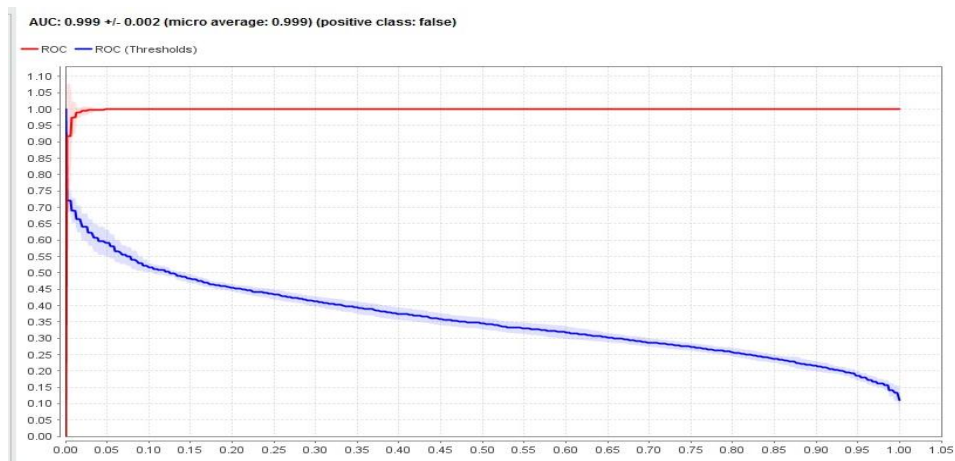
Dalam penelitian ini, seluruh tahapan uji coba algoritma klasifikasi pada dataset yang telah diolah menggunakan *tools* RapidMiner, pada tahap uji coba pertama menggunakan klasifikasi *Naive Bayes* dengan metode *Cross Validation* serta nilai K-Fold sebagai parameter, pada tahap ini nilai K-Fold yang digunakan ialah K 10 dengan hasil akurasi sebagai berikut:

accuracy: 94.16% +/- 1.37% (micro average: 94.16%)

	true true	true false	class precision
pred. true	1338	0	100.00%
pred. false	177	1515	89.54%
class recall	88.32%	100.00%	

Gambar 6. Hasil pengujian berdasarkan algoritma *Naive Bayes* dengan K-Fold

Pada gambar diatas dapat diketahui bahwa pada tahap pengujian menggunakan algoritma *Naive Bayes* dengan nilai K-Fold *Cross Validation* 10 menghasilkan akurasi dengan jumlah 94,16% serta nilai *precision* nya dari masing-masing *prediction true* dan *prediction false* bernilai 100,00% dan 89,54% adapun *recall* dari masing-masing *prediction* bernilai 100,00% dan 88,32%. Adapun distribusi AUC (*Area Under Curve*) dari pengujian ini dapat dilihat sebagai berikut:



Gambar 7. Grafik AUC klasifikasi *Naive Bayes*

Pada hasil uji coba diatas menunjukkan bahwa grafik ROC dengan nilai AUC (*Area Under Curve*) dengan menggunakan model klasifikasi *Naive Bayes* serta metode *Cross Validation* sebesar 0,999.

**3.2.2. Pengujian Menggunakan Klasifikasi *Random Forest* dengan *Cross Validation***

Selanjutnya uji coba dengan menggunakan model algoritma *Random Forest*. Pada tahap ini dataset diuji menggunakan metode *Cross Validation* dengan nilai K-Fold yang digunakan ialah K 10 dengan hasil akurasi sebagai berikut:

accuracy: 97.16% +/- 1.37% (micro average: 97.16%)

	true true	true false	class precision
pred. true	1491	62	96.01%
pred. false	24	1453	98.38%
class recall	98.42%	95.91%	

Gambar 8. Hasil pengujian berdasarkan algoritma *Random Forest* dengan K-Fold

Pada gambar tersebut dilihat bahwa hasil dari uji coba menggunakan model algoritma *Random Forest* menghasilkan nilai akurasi yang cukup tinggi bila dibandingkan dengan uji coba menggunakan algoritma *Naive Bayes*. Pada pengujian menggunakan model algoritma *Random Forest* ini jumlah akurasi yang dihasilkan bernilai 97,16%. Dengan nilai *precision* nya dari masing-masing *prediction true* dan *prediction false* bernilai 96,01% dan 98,38% adapun *recall* dari masing-masing *prediction* bernilai 98,42% dan 95,91%. Adapun distribusi AUC (*Area Under Curve*) dari pengujian ini dapat dilihat sebagai berikut:



Gambar 9. Grafik AUC klasifikasi *Random Forest*

Dari hasil uji coba tersebut menunjukkan bahwa grafik ROC dengan nilai AUC (*Area Under Curve*) dengan menggunakan model klasifikasi *Random Forest* serta metode *Cross Validation* sebesar 0,996.

**3.2.3. Pengujian Menggunakan Klasifikasi *Support Vector Machine* dengan *Cross Validation***

Pada pengujian selanjutnya yaitu dengan menggunakan model algoritma *Support Vector Machine* serta menggunakan nilai K-Fold yang sama dengan pengujian sebelumnya yakni 10, pada pengujian menggunakan *Cross Validation* ini menghasilkan nilai akurasi yang berbeda dengan kedua algoritma sebelumnya namun tidak cukup kecil bila dibandingkan dengan algoritma *Naive Bayes*, adapun nilai akurasi nya sebagai berikut:

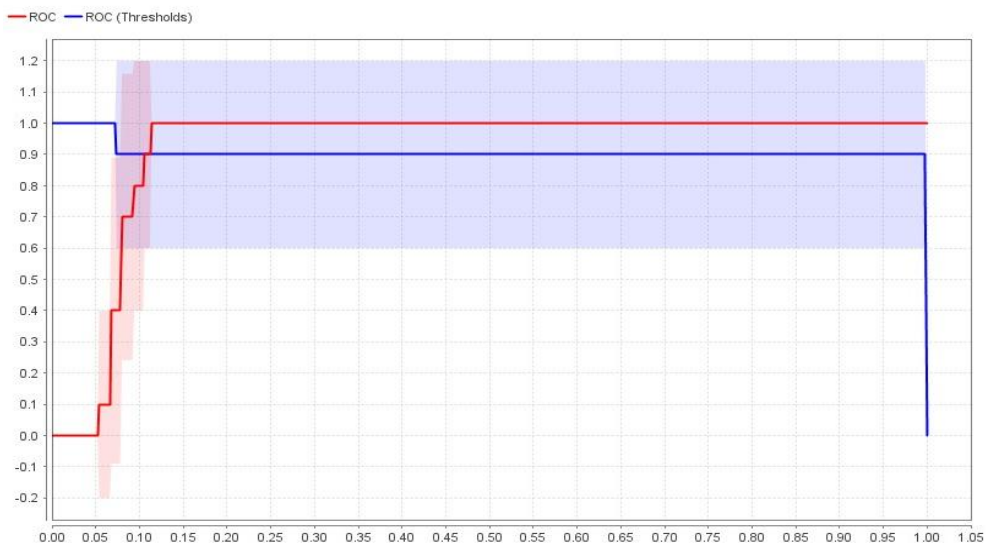
accuracy: 96.01% +/- 0.94% (micro average: 96.01%)

	true true	true false	class precision
pred. true	1394	0	100.00%
pred. false	121	1515	92.60%
class recall	92.01%	100.00%	

Gambar 10. Hasil pengujian berdasarkan algoritma *Support Vector Machine* dengan K-Fold

Dapat diketahui bahwa dari gambar tersebut pada pengujian ini menghasilkan nilai akurasi yang cukup tinggi bila dibandingkan dengan algoritma klasifikasi *Naive Bayes* yakni bernilai 96,01% dari hasil sebelumnya yaitu *Naive Bayes* 94,16% serta nilai *precision* nya dari masing-masing *prediction true* dan *prediction false* bernilai 100,00% dan 92,60% adapun *recall* dari masing-masing *prediction* bernilai 92,01% dan 100,00%. Untuk distribusi AUC (*Area Under Curve*) dari pengujian ini dapat dilihat sebagai berikut:

AUC: 0.543 +/- 0.137 (micro average: 0.543) (positive class: false)



Gambar 11. Gambar 7. AUC Klasifikasi *Support Vector Machine*

Adapun pada gambar diatas menunjukkan bahwa grafik ROC dengan nilai AUC (*Area Under Curve*) dengan menggunakan model klasifikasi *Support Vector Machine* serta metode *Cross Validation* sebesar 0,543.

Bila dilihat dari hasil pengujian ketiga algoritma tersebut diketahui bahwa model klasifikasi dengan algoritma *Random Forest* memiliki tingkat akurasi yang cukup tinggi dibandingkan dengan *Naive Bayes* ataupun *Support Vector Machine*, hal ini dapat dilihat dari tabel perbandingan berikut :

Tabel 1. Hasil pengujian dari ketiga algoritma

No	Klasifikasi	Akurasi	AUC
1	<i>Naive Bayes</i>	94,16 %	0,999
2	<i>Random Forest</i>	97,16 %	0,996



3	Support Vector Machine	96,01 %	0,543
---	------------------------	---------	-------

Hasil tersebut menunjukkan jumlah akurasi dari masing-masing algoritma dengan nilai *precision* dan *recall* yang berbeda-beda serta ROC Curve dengan nilai AUC (*Area Under Curve*) yang berbeda. Adapun dalam uji coba ketiga model algoritma yang telah dilakukan diatas menggunakan metode *Cross Validation* dengan masing masing nilai K 10 dan menghasilkan nilai akurasi lebih baik dari nilai K-Fold yang lainnya.

#### 4. Kesimpulan

Berdasarkan hasil analisa dan hasil dari beberapa pengujian model klasifikasi diatas, dapat diperoleh beberapa kesimpulan diantaranya diketahui bahwa *review* atau komentar di Google Play Store yang diberikan oleh pengguna terhadap aplikasi Ruangguru bila dilihat dari tabel hasil *preprocessing* dataset menunjukkan jumlah komentar positif lebih banyak diberikan dibandingkan dengan komentar atau *review* negatif dari para pengguna aplikasi Ruangguru, hal ini dapat disimpulkan bahwa dengan tingkat penilaian respon positif lebih tinggi pada aplikasi Ruangguru para pengguna merasa puas dengan fitur-fitur yang diberikan oleh aplikasi Ruangguru.

Pada penelitian ini pun penulis dapat menerapkan beberapa model algoritma klasifikasi untuk menentukan nilai akurasi tertinggi serta menghasilkan nilai AUC (*Area Under Curve*) dari masing-masing model klasifikasi yang digunakan dengan dataset yang telah di *scrapping* serta melalui tahap *text preprocessing* sehingga diketahui hasil akurasi tertinggi berada dikisaran >94% dan masing - masing pengujian menggunakan metode *Cross Validation* dengan nilai K Fold 10. Guna menghitung performa terbaik dari masing-masing algoritma.

Adapun hasil selanjutnya bahwa pada pengujian diatas didapat nilai akurasi tertinggi yaitu ada pada pengujian dengan menggunakan model algoritma *Random Forest* dengan jumlah akurasi sebesar 97,16% serta nilai AUC 0,996 kemudian disusul algoritma *Support Vector Machine* yang menghasilkan akurasi sebesar 96,01% dengan nilai AUC sebesar 0,543 dan nilai akurasi terendah ada pada pengujian menggunakan algoritma *Naive Bayes* dengan nilai akurasi sebesar 94,16% serta nilai AUC sebesar 0,999. Dengan demikian dapat disimpulkan bahwa hasil performa dari pengujian diatas menunjukkan *Random Forest* memiliki nilai akurasi tertinggi dari kedua algoritma lainnya yang telah diuji serta menunjukkan adanya kenaikan akurasi dari penelitian sebelumnya dengan peningkatan akurasi sebesar 7,16%.

#### References

- [1] H. Wahyono, "Pemanfaatan Teknologi Informasi dalam Penilaian Hasil Belajar pada Generasi Milenial di Era Revolusi Industri 4.0," *Proceeding Biol. Educ.*, vol. 3, no. 1, pp. 192–201, 2019.
- [2] T. Listyorini and A. Widodo, "Perancangan Mobile Learning Mata Kuliah Sistem Operasi Berbasis Android," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 3, no. 1, p. 25, 2013, doi: 10.24176/simet.v3i1.85.
- [3] Y. & I. B. . Maryono, *Teknologi Informasi & Komunikasi*. Jakarta: Yudhistira Ghalia Indonesia, 2008.
- [4] Ruangguru, "Profile RuangGuru," 2020. [https://ruangguru.com/general/about?utm\\_source=bimbelrg&utm\\_medium=referral&utm\\_campaign=footer](https://ruangguru.com/general/about?utm_source=bimbelrg&utm_medium=referral&utm_campaign=footer).
- [5] F. Gunawan, M. A. Fauzi, and P. P. Adikara, "Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile)," *Syst. Inf. Syst. Informatics J.*, vol. 3, no. 2, pp. 1–6, 2017, doi: 10.29080/systemic.v3i2.234.
- [6] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, p. 258, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [7] M. R. Firdaus, F. M. Rizki, and F. M. Gaus, "Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru," vol. 4, pp. 66–76, 2020.
- [8] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," vol. 1 No.12, pp. 1733–1741, 2017.
- [9] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol.

*Analisis Sentmen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine (Evita Fitri)*

- 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [10] A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [11] L. Ratnawati and D. R. Sulistyningrum, “Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel,” vol. 8, no. 2, 2019.
- [12] D. Gunawan, R. Dwiza, D. Ardiansyah, F. Akba, and S. Alfariz, “Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023,” *J. Tek. Komput. AMIK BSI*, vol. VI No.1, 2020, doi: 10.31294/jtk.v4i2.
- [13] A. Novantirani, M. K. Sabariah, and V. Effendy, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.
- [14] R. R. Fiska, “Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera),” *SATIN - Sains dan Teknol. Inf.*, vol. 3, no. 1, p. 15, 2017, doi: 10.33372/stn.v3i1.200.
- [15] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, “Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang,” *J. GAUSSIAN*, vol. Volume 3, pp. 811–820, 2014, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>.