



# Analisa Sentimen Ulasan Kepuasan Aplikasi Ruangguru Dengan Metode Random Forest dan KNN

Salsa Ananda Rachmawati<sup>1</sup>, Herny Februariyanti<sup>\*2</sup>

<sup>1,2</sup>Universitas Stikubank; Jalan Tri Lomba Juang Semarang

e-mail: <sup>1</sup>salsaanandarachmawati@mhs.unisbank.ac.id , <sup>\*2</sup>hernyfeb@edu.unisbank.ac.id

## Abstrak

Salah satu sarana belajar online yang banyak dicari adalah Ruangguru. Banyaknya program pembelajaran online yang tersedia, Ruangguru harus terus meningkatkan fungsinya agar dapat memuaskan pelanggan dan memberikan pengalaman yang bermanfaat. Menggunakan algoritma Random Forest dan KNN dengan klasifikasi sentimen positif, negatif, atau netral, penelitian ini menggunakan 1.800 data review yang dibagi menjadi 2 kelompok: 1.394 data latih dari review dan 406 data uji. Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dan KNN dalam melakukan klasifikasi ulasan aplikasi Ruangguru ke dalam kelas bersentimen positif, negatif atau netral dan menganalisis nilai akurasi dari algoritma Random Forest dan KNN. Hasil penelitian menunjukkan akurasi algoritma Random Forest sebesar 96% dibandingkan dengan algoritma KNN yang sebesar 71%. Artinya, performa algoritma Random Forest lebih baik dibandingkan algoritma KNN dalam mengklasifikasikan ulasan aplikasi Ruangguru ke dalam kelas sentimen positif, negatif, atau netral, dengan akurasi sebesar 96%.

**Kata kunci;** KNN, Random Forest, Ruangguru

## Abstract

*Ruangguru is an online learning application that is in great demand. The large number of online learning applications available requires Ruangguru to continue to improve and improve its performance so that it can achieve user satisfaction and the application can be optimally useful for users. This research uses data from 1,800 reviews which is divided into 2, namely 1,394 training data from review data and 406 test data from review data using Random Forest and KNN methods with positive, negative or neutral sentiment classification. This research aims to apply the Random Forest and KNN algorithms in classifying Ruangguru application reviews into classes with positive, negative or neutral sentiment and analyzing the accuracy values of the Random Forest and KNN algorithms. The results of the research show that the accuracy of the random forest algorithm is 97% while the KNN algorithm is 71% so it can be concluded that the Random Forest algorithm is better than the KNN algorithm in classifying Ruangguru application reviews into positive, negative or neutral sentiment classes with an accuracy of 96 %.*

**Keywords;** KNN, Random Forest, Ruangguru

## 1. PENDAHULUAN

Salah satu sarana belajar *online* yang banyak dicari adalah Ruangguru. Banyaknya program pembelajaran *online* yang tersedia, Ruangguru harus terus meningkatkan fungsinya agar dapat memuaskan pelanggan dan memberikan

pengalaman yang bermanfaat [1]. Aplikasi Ruangguru sangat populer di Indonesia dan dapat diunduh secara gratis di *Google Play Store* dan *Appstore*. Pengguna dapat memberikan tanggapan positif maupun negatif terhadap aplikasi Ruangguru. Ruangguru menerima  $\geq 654$  ribu ulasan sehingga perlu dilakukan analisis sentimen untuk mengukur



bagaimana respon masyarakat terhadap Ruangguru [2].

Banyaknya ulasan dari pengguna aplikasi Ruangguru dapat dimanfaatkan untuk mencari sebuah informasi. Salah satu metode yang mungkin digunakan untuk analisis *review* adalah analisis sentimen. Teknik memahami, mengekstraksi, dan memproses data tekstual secara mekanis untuk memperoleh informasi sentimen dari kalimat opini yang dikenal sebagai analisis sentimen, atau penambangan opini [3].

Analisis sentimen dengan algoritma *Random Forest* dan *K-Nearest Neighbor* (KNN) digunakan dalam penelitian ini untuk menentukan apakah ulasan terhadap aplikasi Ruangguru positif, negatif, atau netral. Berdasarkan algoritma pohon keputusan, metode *Random Forest* merupakan metodologi pembelajaran *ensembel* yang menggunakan banyak pohon keputusan sebagai pengklasifikasi. Keputusan kelas terbanyak diambil dari pohon keputusan yang sedang digunakan untuk menentukan kelas-kelas yang muncul dari proses klasifikasi [4]. Dengan membandingkan bobot beberapa fitur yang ada, metode KNN menemukan kasus dengan memperkirakan kemiripan antara contoh baru (data pelatihan) dan kasus lama (data pengujian) [5].

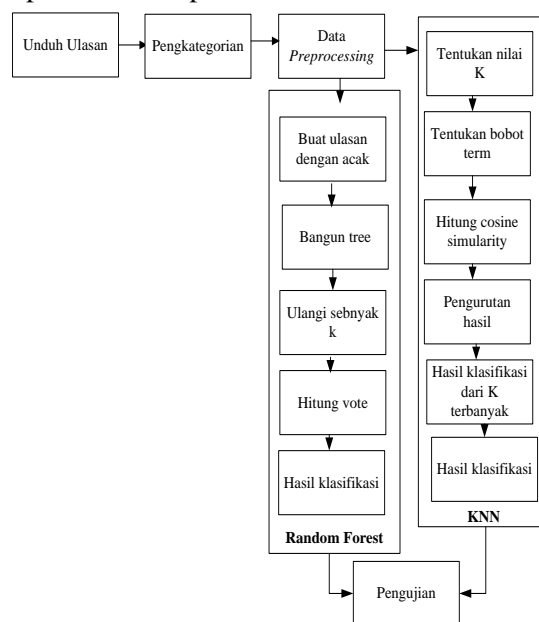
Penelitian oleh Isman dkk (2021), melakukan perbandingan metode *K-Nearest Neighbour* (KNN) dan metode *Local Binary Pattern Histogram* (LPBH) dengan hasil metode KNN memiliki akurasi 97,5 % lebih baik daripada metode LPBH dengan akurasi 94% dalam melakukan klasifikasi daun herbal [6]. Hasil penelitian oleh Leonardo dkk (2020) menghasilkan metode *Random Forest* memiliki akurasi 90 % lebih baik daripada metode *Naïve Bayes* dengan akurasi 85 % dalam melakukan prediksi keberhasilan klien telemarketing [7]. Pendekatan *Random Forest* digunakan dalam penelitian Alita & Rahman (2020) untuk menilai sentimen dalam proses deteksi sarkasme. Hasil penelitian menunjukkan peningkatan nilai *F1score* sebesar 11,27%, hasil akurasi sebesar 16,61%, nilai presisi sebesar 5,45%, dan nilai *recall* sebesar 9,64% [8]. Sentimen vaksin Sinovac dianalisis menggunakan metode *Random Forest* dalam penelitian yang dilakukan Alden et al. (2022), dengan temuan akurasi mencapai 79%, nilai presisi 85%, dan *recall* 90% [9].

Penelitian oleh Kasau dkk (2022) membandingkan *Random Forest* dan *Decision Tree* dengan hasil metode *Random Forest* dengan akurasi 97% lebih baik dari metode *Decision Tree* dengan akurasi 95 % [10]. Pendekatan KNN digunakan dalam penelitian Gata & Purnomo (2017), untuk menilai penggunaan aplikasi dan pengiriman SMS yang tidak sesuai dengan pekerjaan, dengan hasil akurasi mencapai 96,15% [11]. Penelitian oleh Aksan dkk (2022), menggunakan metode KNN untuk menganalisis dan melakukan klasifikasi pembangunan MRT Jakarta dengan tingkat akurasi 88,09% [12].

Penelitian selanjutnya akan menerapkan dan mengukur akurasi algoritma *Random Forest* dan KNN dalam melakukan klasifikasi ulasan aplikasi Ruangguru ke dalam kelas bersentimen positif, negatif atau netral.

## 2. METODE PENELITIAN

Proses klasifikasi analisa sentimen ulasan kepuasan pengguna aplikasi Ruangguru dengan metode *Random Forest* dan KNN diperlihatkan seperti Gambar 1.



Gambar 1 Proses Klasifikasi

Gambaran penyelesaian proses klasifikasi ulasan kepuasan pengguna aplikasi Ruangguru menggunakan algoritma *Random Forest* dan KNN sebagai berikut:

a. Unduh Ulasan

Sumber datanya adalah *Google Play Store* Ruangguru dari periode 2022 sampai

dengan 2023. Sebanyak 1.800 *review* yang diperoleh dari hasil *crawling review* aplikasi Ruangguru dalam format CSV dan dibagi menjadi 1.394 data untuk melatih model.

#### b. Pengkategorian

Proses selanjutnya melakukan kategori ulasan berdasarkan *rating* atau *score* bintang, jika *rating* ulasan lebih besar dari 3 maka ulasan akan dikategorikan sebagai ulasan positif, jika *rating* ulasan sama dengan 3 maka ulasan akan dikategorikan sebagai ulasan netral dan jika *rating* ulasan kurang dari 3 maka ulasan akan dikategorikan sebagai ulasan negatif [13] [14].

#### c. Data Preprocessing

Ulasan yang dikumpulkan akan menjalani sejumlah langkah pemrosesan teks, termasuk *case folding*, tokenisasi, pemfilteran, dan *stemming* untuk memperoleh struktur data yang mudah untuk diolah.

#### d. Proses Klasifikasi

Setelah tahap pra-pemrosesan, ulasan diberi label dengan bobot kata yang digunakan dalam analisis sentimen menggunakan metode TF-IDF. Nilai IDF didapatkan dari Persamaan (1) [15].

$$IDF = \log \frac{D}{dn} \quad (1)$$

Bobot tiap ulasan dihitung dengan Persamaan (2) [15].

$$W_{ij} = tf_{ij} \log \frac{D}{dn} \quad (2)$$

Berikut proses yang dilakukan algoritma *Random Forest* untuk mengkategorikan *review* aplikasi Ruangguru [16]:

- 1) Buat sampel data  $D_j$  dengan mengambil acak dari dataset  $D$ .
- 2) Buatlah pohon ke- $i$  ( $i=1,2,\dots,k$ ) dari sampel data  $D_j$ .
- 3) Lakukan langkah 1 dan 2 sebanyak  $k$  kali.
- 4) Hitung *information gain* untuk menentukan setiap node pada pohon dengan Persamaan (3).

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

Dimana nilai  $Info(D)$  dapat dicari dengan Persamaan (4) dan  $Info_A(D)$  dapat dicari dengan Persamaan (5).

$$Info(D) = - \sum_{i=1}^m P_i \log_2 (P_i) \quad (4)$$

$$Info_A(D) = \sum_{i=1}^m \frac{D_j}{D} \times Info(D) \quad (5)$$

- 5) Untuk menghitung *information gain* dari atribut yang memiliki nilai kontinu (numerik), maka harus ditentukan nilai pembelah (*split-point*) terbaik untuk mengelompokan nilai dari atribut tersebut.
- 6) Ambil data *test* dan gunakan *rule* dari setiap pohon untuk memprediksi keluaran klasifikasi dari data tersebut, simpan hasil yang didapat.
- 7) Hitung suara (*vote*) untuk setiap target yang diprediksi dari setiap pohon.
- 8) Pertimbangkan target prediksi yang terpilih dengan memilih target kelas yang paling banyak diprediksi sebagai hasil prediksi akhir dari metode *Random Forest*.
- 9) Hasil prediksi setiap pohon bisa saja ada yang berbeda dan ada yang sama, maka prediksi akhir akan dipilih berdasarkan prediksi kelas yang terbanyak diprediksi. Misalkan dari 100 pohon, 80 pohon memprediksi target adalah kelas positif dan sisanya adalah kelas negatif, maka prediksi akhir yang dipilih adalah kelas positif.

Langkah-langkah algoritma KNN dalam melakukan klasifikasi ulasan aplikasi Ruangguru adalah sebagai berikut [17]:

- 1) Menentukan parameter  $k$ , dimana penelitian ini menggunakan  $K=3$ .
- 2) Gunakan istilah pembobotan TF-IDF untuk menghitung bobot setiap kata.
- 3) Gunakan kesamaan kosinus untuk menentukan seberapa mirip ulasan satu sama lain.
- 4) Urutkan hasil perhitungan *cosine* kemiripan dari yang terbesar hingga yang terkecil.
- 5) Tentukan kemiripan maksimal pada *review* yang dikategorikan dengan mengambil  $K$  sebanyak-banyaknya.
- 6) Jika ulasan lebih banyak mengandung ulasan positif maka ulasan diklasifikasikan menjadi klasifikasi positif, sedangkan jika ulasan lebih banyak mengandung ulasan negatif maka ulasan diklasifikasikan menjadi klasifikasi negatif.

#### e. Pengujian

Tujuan dari langkah ini adalah untuk menguji hasil klasifikasi. Pengujian dilakukan untuk memastikan model yang diperoleh akurat dan selaras dengan tujuan yang ingin dicapai menggunakan *Confusion Matrix* yang seperti pada Persamaan (6) [18].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \quad (6)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Text Processing

##### a. Case Folding

Semua huruf akan diubah menjadi huruf kecil. Selain itu, prosedur pelipatan huruf besar-kecil akan menghilangkan tanda baca apa pun, termasuk [?/\$/./!2\_:")(-+,]. Proses *case folding* untuk kata “Aplikasi” diubah menjadi kata “aplikasi” (semua kata diubah menjadi huruf kecil) dapat dilihat seperti Gambar 2.

|   | sentiment | content   |
|---|-----------|---|
| 0 | NEGATIF   | Pada bagian home saya agak kesulitan mendapatk... |
| 1 | POSITIF   | Aplikasi yang sangat membantu disaat saya mera... |
| 2 | NETRAL    | Saya mempunyai beberapa kendala seperti materi... |
| 3 | NEGATIF   | Kelebihan: 1. Ada konsep kilat yang ga bikin n... |
| 4 | POSITIF   | Sayang banget... Padahal ruangguru udah bagus ... |

|   | sentiment | content   |
|---|-----------|---|
| 0 | NEGATIF   | pada bagian home saya agak kesulitan mendapatk... |
| 1 | POSITIF   | aplikasi yang sangat membantu disaat saya mera... |
| 2 | NETRAL    | saya mempunyai beberapa kendala seperti materi... |
| 3 | NEGATIF   | kelebihan: 1. ada konsep kilat yang ga bikin n... |
| 4 | POSITIF   | sayang banget... padahal ruangguru udah bagus ... |

Gambar 2 Case Folding

##### b. Tokenizing

Langkah ini memisahkan ulasan aplikasi Ruangguru menjadi beberapa bagian yang disebut token seperti Gambar 3.

|   | sentiment | content   |
|---|-----------|---|
| 0 | NEGATIF   | pada bagian home saya agak kesulitan mendapatk... |
| 1 | POSITIF   | aplikasi yang sangat membantu disaat saya mera... |
| 2 | NETRAL    | saya mempunyai beberapa kendala seperti materi... |
| 3 | NEGATIF   | kelebihan: 1. ada konsep kilat yang ga bikin n... |
| 4 | POSITIF   | sayang banget... padahal ruangguru udah bagus ... |



|   |         |   |
|---|---------|---|
| 0 | NEGATIF | [pada, bagian, home, saya, agak, kesulitan, me... |
| 1 | POSITIF | [aplikasi, yang, sangat, membantu, disaat, say... |
| 2 | NETRAL  | [saya, mempunyai, beberapa, kendala, seperti, ... |
| 3 | NEGATIF | [kelebihan, 1, ada, konsep, kilat, yang, ga, b... |
| 4 | POSITIF | [sayang, banget, padahal, ruangguru, udah, bag... |

Gambar 3 Tokenizing

Ini juga menghilangkan karakter khusus, angka, dan tanda baca, seperti mention (@), hashtag (#), titik (.), koma (,), dan tanda baca lainnya, dari tinjauan karena tidak ada hubungannya dengan nilai klasifikasi dan harus dibuang. Gambar 3 mengilustrasikan proses tokenisasi dalam tindakan, ketika pernyataan “sebenarnya sih udah bagus” dibagi menjadi kata-kata sesuai spasi.

##### c. Stopword Removal

Tahap penghapusan kata *stopword* diantaranya “karena”, “dalam”, “oleh”, “pada”, dan seterusnya. Proses *stopword removal* dapat dilihat seperti Gambar 4, dimana kata sambung seperti “sebenarnya” akan dihapus atau dihilangkan dan seterusnya.

|   |         |   |
|---|---------|---|
| 0 | NEGATIF | [pada, bagian, home, saya, agak, kesulitan, me... |
| 1 | POSITIF | [aplikasi, yang, sangat, membantu, disaat, say... |
| 2 | NETRAL  | [saya, mempunyai, beberapa, kendala, seperti, ... |
| 3 | NEGATIF | [kelebihan, 1, ada, konsep, kilat, yang, ga, b... |
| 4 | POSITIF | [sayang, banget, padahal, ruangguru, udah, bag... |

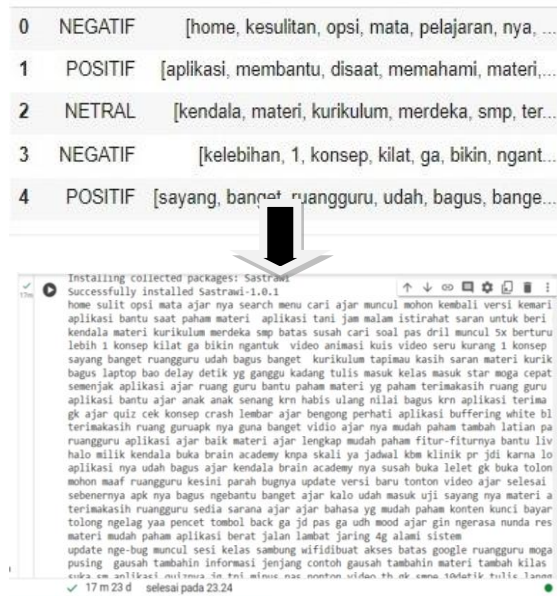


|   |         |   |
|---|---------|---|
| 0 | NEGATIF | [home, kesulitan, opsi, mata, pelajaran, nya, ... |
| 1 | POSITIF | [aplikasi, membantu, disaat, memahami, materi,... |
| 2 | NETRAL  | [kendala, materi, kurikulum, merdeka, smp, ter... |
| 3 | NEGATIF | [kelebihan, 1, konsep, kilat, ga, bikin, ngant... |
| 4 | POSITIF | [sayang, banget, ruangguru, udah, bagus, bange... |

Gambar 4 Stopword Removal

##### d. Stemming

Proses *stemming* dengan *Library Sastrawi* dapat dilihat seperti Gambar 5, dimana kata “materinya” akan diubah menjadi “materi” dan kata “aplikasinya” akan diubah menjadi kata “aplikasi” dan seterusnya.



Gambar 5 Stemming

3.2. Klasifikasi KNN

Langkah-langkah algoritma KNN dalam melakukan klasifikasi ulasan “terima kasih ruangguru, yang sangat sangat membantu saya dalam belajar” yang kemudian dilakukan *text processing* menjadi “terima kasih ruangguru bantu belajar” yang selanjutnya disebut sebagai Q adalah sebagai berikut:

- Menentukan parameter *k*, dimana penelitian ini menggunakan *K=3*.
- Data latih dalam pembentukan *term* menggunakan 5 data latih, dimana ulasan yang mengandung kata kunci diberi nilai 1. Nilai TF merupakan penjumlahan dari kata yang ditemukan dan IDF merupakan hasil dari Persamaan 1 dimana *n* merupakan jumlah ulasan yang dihitung. Dalam hal ini *n* bernilai 6 seperti yang ditampilkan Tabel 1 dan Tabel 2.

Tabel 1 Bobot Kata

| Term      | tf  |     |     |     |
|-----------|-----|-----|-----|-----|
|           | Q   | UL1 | UL2 | UL3 |
| terima    | 1.0 |     |     |     |
| kasih     | 1.0 |     |     |     |
| ruangguru | 1.0 |     |     |     |
| bantu     | 1.0 |     | 1.0 |     |
| belajar   | 1.0 |     |     |     |
| home      |     | 1.0 |     |     |
| sulit     |     | 1.0 |     |     |
| opsi      |     | 1.0 |     |     |
| mata      |     | 1.0 |     |     |
| pelajaran |     | 1.0 |     |     |
| cari      |     |     |     |     |
| muncul    |     |     |     |     |

|          |  |  |     |     |
|----------|--|--|-----|-----|
| bagus    |  |  |     |     |
| aplikasi |  |  | 1.0 |     |
| materi   |  |  | 1.0 | 1.0 |

Tabel 2 Bobot Kata Lanjutan

| Term      | tf  |     |     | idf       |
|-----------|-----|-----|-----|-----------|
|           | UL4 | UL5 | Df  | Log(n/df) |
| terima    |     |     | 1.0 | 0,78      |
| kasih     |     |     | 1.0 | 0,78      |
| ruangguru |     |     | 1.0 | 0,78      |
| bantu     |     |     | 2.0 | 0,48      |
| belajar   |     |     | 1.0 | 0,78      |
| home      |     |     | 1.0 | 0,78      |
| sulit     |     |     | 1.0 | 0,78      |
| opsi      |     | 1.0 | 2.0 | 0,48      |
| mata      |     | 1.0 | 2.0 | 0,48      |
| pelajaran |     | 1.0 | 2.0 | 0,48      |
| cari      | 1.0 |     | 1.0 | 0,78      |
| muncul    | 1.0 |     | 1.0 | 0,78      |
| bagus     | 1.0 |     | 1.0 | 0,78      |
| aplikasi  |     |     | 1.0 | 0,78      |
| materi    |     |     | 2.0 | 0,48      |

Proses selanjutnya adalah menghitung bobot *W<sub>ij</sub>* dengan Persamaan 2 yang hasilnya ditampilkan seperti Tabel 3.

Tabel 3 Hitung Bobot

| <i>W<sub>ij</sub>=tf.idf</i> |      |      |      |      |      |
|------------------------------|------|------|------|------|------|
| Q                            | UL1  | UL2  | UL3  | UL4  | UL5  |
| 0,78                         |      |      |      |      |      |
| 0,78                         |      |      |      |      |      |
| 0,78                         |      |      |      |      |      |
| 0,48                         |      | 0,48 |      |      |      |
| 0,78                         |      |      |      |      |      |
|                              | 0,78 |      |      |      |      |
|                              | 0,78 |      |      |      |      |
|                              | 0,48 |      |      |      | 0,48 |
|                              | 0,48 |      |      |      | 0,48 |
|                              | 0,48 |      |      |      | 0,48 |
|                              |      |      |      | 0,78 |      |
|                              |      |      |      | 0,78 |      |
|                              |      |      |      | 0,78 |      |
|                              |      | 0,78 |      |      |      |
|                              |      | 0,48 | 0,48 |      |      |

- Gunakan penghitungan kesamaan kosinus untuk menentukan seberapa mirip ulasan satu sama lain dengan Persamaan (7).

$$similarity(x,y) = \frac{\sum_{i=1}^t x_i y_i}{\sqrt{\sum_{i=1}^t x_i^2 \sum_{i=1}^t y_i^2}} \quad (7)$$

- 1) Tentukan hasil kali skalar Q dan lima tinjauan sisanya. Berdasarkan pembilang pada rumus di atas, hasil perkalian setiap *review* dengan Q dijumlahkan seperti yang terlihat pada Tabel 4.

Tabel 4 Perkalian Skalar

| WD*Wdi   |     |      |     |     |     |
|----------|-----|------|-----|-----|-----|
|          | UL1 | UL2  | UL3 | UL4 | UL5 |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0,23 | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
|          | 0   | 0    | 0   | 0   | 0   |
| $\Sigma$ | 0   | 0,23 | 0   | 0   | 0   |

- 2) Tentukan durasi setiap tinjauan, dengan memasukkan Q. Kuadratkan bobot  $W_{ij}$ , tambahkan nilai kuadrat, lalu akarkan hasilnya seperti yang ditunjukkan pada Tabel 5.

Tabel 5 Panjang Vektor

| Panjang Vektor |      |      |      |      |      |      |
|----------------|------|------|------|------|------|------|
|                | Q    | B1   | B2   | B3   | B4   | B5   |
|                | 0,61 |      |      |      |      |      |
|                | 0,61 |      |      |      |      |      |
|                | 0,61 |      |      |      |      |      |
|                | 0,22 |      | 0,23 |      |      |      |
|                | 0,61 |      |      |      |      |      |
|                |      | 0,61 |      |      |      |      |
|                |      | 0,61 |      |      |      |      |
|                |      | 0,23 |      |      |      | 0,23 |
|                |      | 0,23 |      |      |      | 0,23 |
|                |      | 0,23 |      |      |      | 0,23 |
|                |      |      |      |      | 0,61 |      |
|                |      |      |      |      | 0,61 |      |
|                |      |      |      |      | 0,61 |      |
|                |      |      | 0,61 |      |      |      |
|                |      |      | 0,23 | 0,23 |      |      |
| $\Sigma$       | 2,66 | 1,91 | 1,07 | 0,23 | 1,83 | 0,69 |
| $\sqrt{\quad}$ | 1,63 | 1,38 | 1,03 | 0,48 | 1,35 | 0,83 |

- 3) Gunakan rumus Persamaan (7). Tentukan kemiripannya dengan UL1, UL2, dan seterusnya, sampai dengan UL5.

$$Sim(Q, UL1) = 0,2702$$

$$Sim(Q, UL2) = 0,3682$$

$$Sim(Q, UL3) = 0,779$$

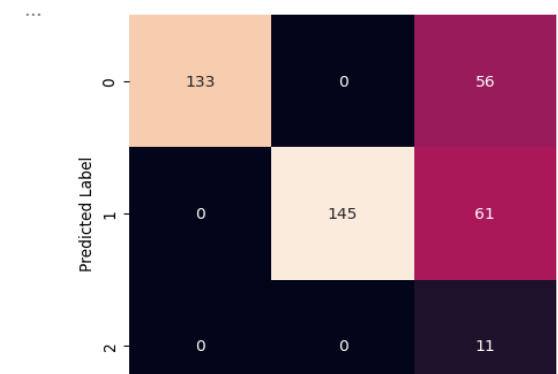
$$Sim(Q, UL4) = 0,2760$$

$$Sim(Q, UL5) = 0,449$$

- d. Susun temuan perhitungan *cosine-similarity* (UL3, UL5, UL2, UL4, dan UL1) dalam urutannya.  
 e. Pilih sebanyak K yaitu tiga *review* yang paling mirip dengan kategori UL3 (Negatif), UL5 (Positif), dan UL2 (Positif).  
 f. Karena hasil lebih banyak mengandung ulasan positif maka ulasan “terima kasih ruangguru, yang sangat sangat membantu saya dalam belajar” dapat diklasifikasikan sebagai ulasan positif.

Hasil klasifikasi ulasan yang lain dalam aplikasi Ruangguru menggunakan metode KNN diperlihatkan seperti Gambar 6.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| NEGATIF      | 0.70      | 1.00   | 0.83     | 133     |
| NETRAL       | 0.70      | 1.00   | 0.83     | 145     |
| POSITIF      | 1.00      | 0.09   | 0.16     | 128     |
| accuracy     |           |        | 0.71     | 406     |
| macro avg    | 0.80      | 0.70   | 0.60     | 406     |
| weighted avg | 0.80      | 0.71   | 0.62     | 406     |



Gambar 6. Hasil Klasifikasi KNN

Dari Gambar 6 diketahui bahwa ulasan negatif (0) yang diklasifikasikan negatif (0) adalah sebanyak 133 ulasan dari 406 ulasan, ulasan netral (1) yang diklasifikasikan netral (1) sebanyak 145 ulasan dari 406 ulasan dan ulasan positif (2) yang diklasifikasikan positif (2) sebanyak 11 ulasan dari 406 ulasan. Hasil perhitungan *Confusion Matrix* selengkapnya diperlihatkan seperti Tabel 6.

Tabel 6 Confusion Matrix KNN

| Prediksi   |                |                |            |
|------------|----------------|----------------|------------|
| True Class | 0              | 1              | 2          |
| 0          | 133<br>TNegNeg | 0<br>NegFNet   | 0<br>NetFP |
| 1          | 0<br>NetFNeg   | 145<br>TNetNet | 0<br>NegFP |
| 2          | 56<br>PFNeg    | 61<br>PFNet    | 11<br>TPP  |

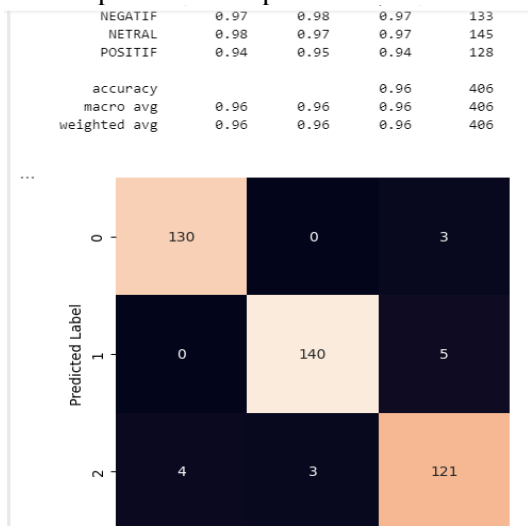
Tabel 6 terdapat kelas prediksi dan kelas sebenarnya. Kelas prediksi adalah kelas hasil prediksi dari proses klasifikasi menggunakan KNN, sedangkan kelas sebenarnya adalah kelas yang sebelumnya sudah ditentukan sentimennya. Nilai perhitungan akurasi dari metode KNN adalah :

$$\frac{11+145+133}{11+133+145+56+0+61+0+0+0} = \frac{289}{406} = 0,71.$$

Jadi pada pengujian akurasi yang diperoleh dari metode KNN dengan menggunakan 406 data ulasan menghasilkan akurasi 0,71 atau 71 %.

### 3.3. Klasifikasi *Random Forest*

Hasil klasifikasi ulasan aplikasi Ruangguru menggunakan metode *Random Forest* diperlihatkan seperti Gambar 7.



Gambar 7 Hasil Klasifikasi *Random Forest*

Dari Gambar 7 diketahui bahwa ulasan negatif (0) yang diklasifikasikan negatif (0) adalah sebanyak 130 ulasan dari 406 ulasan, ulasan netral (1) yang diklasifikasikan netral (1) adalah sebanyak 140 ulasan dari 406 ulasan dan ulasan positif (2) yang diklasifikasikan positif (2) adalah sebanyak 121 ulasan dari 406 ulasan. Hasil perhitungan *Confusion Matrix* selengkapnya diperlihatkan seperti Tabel 7.

Tabel 7 *Confusion Matrix Random Forest*

| True Class | Prediksi       |                |            |
|------------|----------------|----------------|------------|
|            | 0              | 1              | 2          |
| 0          | 130<br>TNegNeg | 0<br>NegFNet   | 3<br>NetFP |
| 1          | 0<br>NetFNeg   | 140<br>TNetNet | 5<br>NegFP |
| 2          | 4<br>PFNeg     | 3<br>PFNet     | 121<br>TPP |

Kelas sebenarnya dan kelas prediksi disajikan pada Tabel 6. Kelas sebenarnya adalah kelas yang sentimennya telah dipastikan, sedangkan kelas prediksi adalah kelas yang diprediksi dengan metode klasifikasi hutan acak (*Random Forest*). Nilai perhitungan akurasi dari metode *random forest* adalah :

$$\frac{(TPP + TNegNeg + TNetNet)}{(TPP + TNegNeg + TNetNet + PFNeg + NegFP + PFNet + NetFP + NetFNeg + NegFNet)}$$

$$\frac{121+140+130}{121+130+140+4+5+4+3+0+0} = \frac{391}{406} = 0,96.$$

Jadi pada pengujian akurasi yang diperoleh dari metode *Random Forest* dengan menggunakan 406 data ulasan menghasilkan akurasi 0,96 atau 96 %. Berdasarkan hasil klasifikasi dari algoritma *Random Forest* dan KNN disimpulkan bahwa algoritma *Random Forest* lebih baik daripada algoritma KNN dalam melakukan klasifikasi ulasan aplikasi Ruangguru ke dalam kelas bersentimen positif, negatif atau netral dengan akurasi mencapai 96%.

## 4. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, maka dapat diambil kesimpulan yaitu:

- Dengan menggunakan algoritma *Random Forest* dan KNN, penelitian ini berhasil mengklasifikasikan evaluasi kepuasan pengguna terhadap aplikasi Ruangguru, sehingga memungkinkan seseorang untuk menentukan apakah ulasan tersebut memiliki sikap positif, netral, atau negatif.
- Akurasi algoritma *Random Forest* sebesar 96% dibandingkan dengan algoritma KNN sebesar 71%. Artinya, performa algoritma *Random Forest* lebih baik dibandingkan algoritma KNN dalam mengklasifikasikan ulasan aplikasi Ruangguru ke dalam kelas sentimen positif, negatif, atau netral, dengan akurasi sebesar 96%.

## 5. SARAN

Saran-saran untuk penelitian lebih lanjut yaitu dapat ditambahkan kelas sentimen (*multi-class*) dengan menerapkan tingkat emosi (*emotion*) sebagai *class*.

## DAFTAR PUSTAKA

- [1] B. Liu, *Sentiment Analysis And Opinion Mining*, Morgan dan Claypool Publisher, 2012.
- [2] M. W. Berry and J. Kogan, *Text Mining Application and Theory*, United Kingdom: Wiley, 2015.
- [3] S. Symeonidis, D. Effrosynidis and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis.," *Expert Systems With Applications*, p. 298–310, 2018.
- [4] B. B. Baskoro, S. Khomsah and I. Susanto, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus : Ulasan Pelanggan Pada Situs Tripadvisor)," *INISTA*, vol. XV, no. 2, pp. 21-29, 2021.
- [5] Kusriani and E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Andi, 2015.
- [6] Isman, A. Ahmad and A. Latief, "Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal," *RESTI*, vol. V, no. 3, pp. 557-564, 2021.
- [7] R. Leonardo, J. Pratama and Chrisnatalis, "Perbandingan Metode Random Forest dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing," *JUTIKOMP*, vol. III, no. 2, pp. 455-459, 2020.
- [8] D. Alita and A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *Jurnal Komputasi*, vol. VIII, no. 2, pp. 50-58, 2020.
- [9] M. Y. Aldean , Paradise and N. A. S. Nugraha , "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," *J. OF INISTA*, vol. IV, no. 2, pp. 64-72, 2022.
- [10] S. Kasau, D. M. Togubu, Hasmah, M. I. Nur, Rosdiana and Sandrawati, "Model Klasifikasi Kanker Payudara Berdasarkan Status Keganasan Menggunakan Algoritma Random Forest Dan Decision Tree," *semantik*, vol. VIII, no. 2, pp. 167-174, 2022.
- [11] W. Gata and Purnomo, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," *Format*, vol. VI, no. 2, pp. 1-13, 2017.
- [12] C. Aksan, B. Pramono and A. M. Sajjah, "Analisis Sentimen Pembangunan Mass Rapid Transit (MRT) Jakarta Pada Twitter Menggunakan Metode Improved K-Nearest Neighbor," *semantik*, vol. VIII, no. 1, pp. 53-60, 2022.
- [13] A. R. Fauzi, A. H. Saputra and M. H. Abdillah, "Analisis Tingkat Kepuasan Pengguna Apex Mobile Berdasarkan Rating Dan Ulasan Google Play Store Menggunakan Naïve Bayes," *JRITI*, vol. I, no. 1, pp. 17-21, 2023.
- [14] R. W. Adanendra, "Analisis Sentimen Pada Aplikasi Duolingo Menggunakan Metode Naïve Bayes Multinomial Dengan Fitur Ekstraksi Countvectorizer," UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH, Jakarta, 2024.
- [15] Tupari, S. Abdullah and Chairani, "Visualisasi Data Analisa Sentimen RUU Omnibus law Kesehatan Menggunakan KNN dengan Software RapidMiner," *Jurnal Informatika: Jurnal pengembangan IT*, vol. VIII, no. 3, pp. 261-268, September 2023.
- [16] I. Afdhal, R. Kurniawan, I. Iskandar and R. Salambue, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. V, no. 1, pp. 122-130, 2022.
- [17] A. D. A. Putra and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. VIII, no. 2, pp. 636-646, 2021.
- [18] R. G. Ramli and Y. Sibaroni, "Klasifikasi Topik Twitter menggunakan Metode Random Forest dan Fitur Ekspansi Word2Vec," *e-Proceeding of Engineering*, vol. IX, no. 1, pp. 79-92, 2022.