

ANALISIS SENTIMEN DATA ULASAN APLIKASI RUANGGURU PADA SITUS GOOGLE PLAY MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER DENGAN NORMALISASI KATA LEVENSHTTEIN DISTANCE

Hindun Habibatul Mubaroroh¹, Hasbi Yasin², Agus Rusgiyono³

^{1, 2, 3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*Email: hindunhabibatul@gmail.com

ABSTRACT

One form of technological development in education is the increasing number of online based learning. More than that, during this period of Covid-19 pandemic distance education was tried by the government that requires learning are done online. The online learning application that is the implementation of this technological development continues to show its existence. Many non-formal educational companies are available, one of which is the Ruangguru, getting a nickname as a number one learning application requires the Ruangguru to continue and improve the performance. Users of the Ruangguru application can communicate a response to Ruangguru through the review feature available on the google play site. The reviews that have been written can be analyzed how the user sentiment is whether positive or negative using Multinomial Naïve Bayes. This method is used because it is easy to use with simple structures and gives high accuracy values. The model will be selected using 10-fold cross validation method to get the model with the best accuracy. The normalization phase of words was also perfected using Levenshtein Distance method that was proven to add accuracy value. Performance result using Multinomial Naïve Bayes by adding Levenshtein Distance method to fix the words gives an average accuracy value of 88,20% with the 8th fold as the fold with the best accuracy value of 94%.

Keywords: *Ruang guru, Sentiment Analysis, Multinomial Naïve Bayes, Levenshtein Distance*

1. PENDAHULUAN

Perkembangan teknologi terus mengalami peningkatan dari waktu ke waktu. Salah satu bidang yang dipengaruhi oleh perkembangan tersebut adalah bidang pendidikan. Menurut survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pada periode 2019 – kuartal II 2020 mayoritas konten media *online* yang diakses pengguna adalah konten pendidikan dan laman sekolah. Terlebih selama pemberlakuan Pembelajaran Jarak Jauh (PJJ) oleh pemerintah sebagai bentuk adaptasi dimasa pandemi *Covid-19* ini, semakin banyak pembelajaran berbasis *online* yang menunjukkan eksistensinya.

Salah satu perusahaan pendidikan non-formal yang ada adalah Ruangguru, mendapat julukan sebagai aplikasi bimbel No. 1 di Indonesia menjadikan Ruangguru sebagai aplikasi pembelajaran *online* yang banyak diminati. Banyaknya aplikasi pembelajaran *online* yang tersedia mengharuskan Ruangguru untuk terus meningkatkan dan memperbaiki performanya agar dapat mencapai kepuasan pengguna dan aplikasinya dapat bermanfaat secara optimal bagi pengguna. Aplikasi Ruangguru tersedia dalam situs *Google Play* yang dapat diunduh secara bebas. Situs *Google Play* memiliki fitur ulasan tempat pengguna memberikan *feedback*

kepada Ruangguru. Analisis sentimen dapat digunakan untuk melihat bagaimana respon pengguna yang dapat dijadikan sebagai bahan referensi oleh Ruangguru.

Terdapat banyak metode pengklasifikasian untuk analisis sentimen, salah satunya metode dengan menggunakan algoritma *Naïve Bayes Classifier*. Penggunaannya yang mudah dengan struktur yang sederhana dan tingkat keakurasiannya tinggi membuat *Naïve Bayes Classifier* menjadi salah satu metode yang populer untuk keperluan *text mining* (Hall, 2006). Salah satu model dari *Naïve Bayes* yang sering digunakan dalam klasifikasi teks adalah *Multinomial Naïve Bayes*. Penelitian ini akan membahas bagaimana analisis sentimen data ulasan pada aplikasi Ruangguru menggunakan algoritma *Multinomial Naïve Bayes Classifier*. Serta akan digunakan metode jarak *Levenshtein Distance* sebagai tahap penyempurnaan normalisasi kata yang ada. Ulasan pengguna aplikasi Ruangguru akan diklasifikasikan menjadi dua kelas yaitu sentimen positif dan sentimen negatif.

2. TINJAUAN PUSTAKA

Berdiri pada tahun 2014, Ruangguru merupakan perusahaan teknologi terbesar di Indonesia yang berfokus pada layanan berbasis pendidikan dan telah digunakan lebih dari 22 juta pengguna serta mengelola 300 ribu guru yang menawarkan jasa di lebih dari 100 bidang pelajaran (Ruangguru, 2021). Ruangguru terus menghadirkan berbagai inovasi secara proaktif dan responsif disesuaikan dengan keadaan yang sedang terjadi guna memperluas akses pendidikan berkualitas melalui teknologi baik bagi siswa maupun guru melalui fitur-fitur yang terus diperbarui dan dikembangkan.

Google Play merupakan layanan konten online milik *Google* berupa toko digital yang menyediakan aplikasi serta konten digital meliputi produk-produk seperti *game*, film, buku, dan musik yang dapat membantu pengguna. Selain terdapat banyak aplikasi yang bisa diunduh, *Google Play* juga memiliki fitur ulasan atau *review* tempat pengguna menyampaikan *feedback* terhadap suatu aplikasi.

Menurut Hasan dalam Nasution, (2017) statistik deskriptif mempelajari cara pengumpulan data dan penyajian data agar mudah dipahami. Penggunaan analisis deskriptif diharapkan dapat memberikan gambaran secara umum tentang aplikasi Ruangguru yang disajikan dalam bentuk diagram grafis yang mudah dipahami.

Text mining merupakan proses analisis teks untuk menggali informasi yang bermanfaat dari peramalan pola dan kecenderungan melalui pola statistik sebuah dokumen yang berisi rangkaian teks. Prosesnya dilakukan otomatis dengan bantuan komputer (Prilianti & Wijaya, 2014). *Text mining* merupakan bagian dari keilmuan *data mining* yang berfokus pada data tekstual yang mampu menganalisis data teks tidak terstruktur maupun semi terstruktur.

Salah satu bentuk implementasi *text mining* adalah analisis sentimen. Menurut Liu (2012) analisis sentimen merupakan bidang ilmu yang menganalisis pendapat seseorang, sentimen, evaluasi, penilaian, sikap, dan perasaan terhadap suatu objek berupa produk, pelayanan, organisasi, seseorang, peristiwa, atau topik tertentu. Analisis sentimen bertujuan untuk mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah teks tersebut bersentimen positif atau negatif.

Pre-processing merupakan tahap awal dari *text mining* untuk mengubah data sesuai dengan format yang dibutuhkan agar data siap untuk dianalisis. Tahap *pre-processing data* yang akan dilakukan dalam penelitian ini meliputi proses *case folding*, *cleaning data*,

normalisasi kata dengan perbaikan kata tidak baku, dan menggunakan metode jarak *Levenshetein Distance*.

Menurut Freeman (2006) *Levenshtein distance* digunakan untuk menghitung jarak kedekatan antara dua buah *string* melalui penambahan karakter, pengubahan karakter, dan penghapusan karakter. Algoritma ini menentukan *distance* berdasarkan jumlah minimum perubahan yang terjadi saat dilakukan transformasi dari *string* awal ke *string* target.

Pelabelan data dilakukan untuk menentukan suatu kalimat opini pada ulasan termasuk ke dalam kelas sentimen positif atau negatif. Pelabelan data menggunakan kamus kumpulan kata positif dan negatif untuk menentukan skor ulasan. Skor jumlah kata positif dihitung kemudian dikurangi skor jumlah kata negatif. Jika nilai skor > 0 maka akan diklasifikasikan ke dalam kelas sentimen positif, sedangkan jika skor < 0 maka akan diklasifikasikan ke dalam kelas negatif dan ketika skor $= 0$ maka akan diklasifikasikan ke dalam kelas netral (Santoso & Nugroho, 2019). Ulasan dengan label netral akan dihilangkan karena dianggap tidak memiliki sentimen.

Tahap *feature selection* dilakukan untuk memilih fitur yang dianggap berpengaruh terhadap isi ulasan. Tahap *feature selection* terdiri dari tiga proses meliputi *stopwords removal* untuk menghapus kata umum, tahap *stemming* untuk mengubah bentuk kata ke bentuk kata dasar dan tahap *tokenizing* untuk memecah teks menjadi kata tunggal.

Tahap pembobotan kata yang digunakan adalah metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk melakukan transformasi data dari data tekstual ke dalam data numerik. *Term Frequency* (TF) merupakan proses untuk menghitung jumlah kemunculan kata dalam tiap data ulasan. *Inverse Document Frequency* (IDF) digunakan untuk menghitung kata yang muncul di berbagai data ulasan yang dianggap sebagai kata umum, yang dinilai tidak penting. Hasil dari pembobotan kata dengan TF-IDF merupakan hasil perkalian nilai TF dan IDF yang dapat dituliskan sebagai berikut:

$$W_{j,i} = \frac{n_{j,i}}{\sum_k n_{k,i}} \cdot \log_2 \frac{D}{d_j} \quad (1)$$

$W_{j,i}$ merupakan pembobotan TF-IDF untuk kata ke- j pada ulasan ke- i , $n_{j,i}$ merupakan jumlah kemunculan kata ke- j pada ulasan ke- i , $\sum_k n_{k,i}$ merupakan total kata pada ulasan ke- i , D merupakan total ulasan dalam korpus dan d_j adalah banyaknya ulasan yang mengandung kata ke- j .

Teorema Bayes merupakan teorema yang mengacu pada probabilitas bersyarat. Metode *Naïve Bayes Classifier* dapat digunakan untuk mengklasifikasikan teks dan merupakan salah satu metode yang populer untuk keperluan *data mining* karena penggunaannya yang mudah (Hall, 2006). Model dari *Naïve Bayes* yang sering digunakan dalam klasifikasi teks adalah *Multinomial Naïve Bayes* yang memperhitungkan frekuensi dari setiap kata yang muncul pada setiap ulasan. Klasifikasi dokumen bertujuan untuk menentukan kelas terbaik pada suatu ulasan yang didasarkan pada nilai *Maximum A Posteriori* (MAP) kelas c_{map} dengan persamaan sebagai berikut:

$$c_{map} = \underset{c \in C}{\operatorname{argmax}} \hat{P}(c) \prod_{k=1}^n \hat{P}(t_k | c) \quad (2)$$

Nilai $\hat{P}(c)$ dan $\hat{P}(t_k|c)$ dapat dihitung dengan persamaan:

$$\hat{P}(c) = \frac{N_c}{N} \quad (3)$$

$$\hat{P}(t|c) = \frac{W_{ct} + 1}{\sum_{W' \in V} W_{ct'} + B'} \quad (4)$$

N_c merupakan jumlah ulasan dengan kategori kelas c , N merupakan jumlah seluruh ulasan, W_{ct} merupakan bobot TF-IDF *term* t pada ulasan dengan kategori c , $\sum_{W' \in V} W_{ct'}$ merupakan jumlah bobot TF-IDF seluruh *term* pada kategori c dan B' merupakan jumlah IDF seluruh *term* pada *vocabulary*.

K-fold cross validation merupakan teknik validasi model dengan memecah data menjadi k bagian set data dengan ukuran yang sama. Pelatihan dan pengujian dilakukan sebanyak k kali dengan pembagian masing-masing segmen berisi $(k-1)$ *fold* data latih dan 1 *fold* untuk data uji. *10-fold cross validation* merupakan salah satu *k-fold cross validation* yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang bias atau bias paling kecil dibanding *fold cross validation* lainnya.

Pengukuran terhadap kinerja suatu sistem klasifikasi digunakan untuk menggambarkan seberapa baik dan akuratnya sistem dalam mengklasifikasikan data dari prediksi klasifikasi yang sudah dihasilkan. Evaluasi hasil klasifikasi dapat dijelaskan dengan tabel *confusion matrix* yang merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi yang terlihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

Prediksi	Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Positive</i>
<i>Negative</i>	<i>False Negative</i>	<i>True Negative</i>

Performa dari model yang telah dihasilkan akan diuji menggunakan *Precision*, *Recall*, *F-Measure*, dan *Accuracy* dengan persamaan sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F - Measure = \frac{2PR}{P + R} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Word cloud merupakan salah satu teknik visualisasi yang terdiri dari kumpulan kata-kata yang paling banyak muncul saat dokumen teks dianalisis (Agrani & Rikumahu, 2020). Besar kecilnya huruf ditentukan oleh intensitas kemunculan tiap kata yang digunakan.

3. METODOLOGI PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data primer yang berupa data kualitatif. Data diperoleh dari hasil *scraping data* berupa data ulasan pengguna aplikasi Ruangguru yang diambil pada 9 Mei 2021 sebanyak 1500 data ulasan paling relevan teratas. Data ulasan yang berlabel netral pada tahap pelabelan akan dihilangkan dan kemudian diambil 1000 data ulasan teratas. Data ulasan yang digunakan merupakan ulasan yang ditulis pada tanggal 5 Februari hingga 7 Mei 2021.

Terdapat 3 macam variabel yang digunakan dalam penelitian ini, meliputi: tanggal dibuatnya ulasan, skor bintang yang diberi oleh pengguna dan ulasan pengguna terhadap aplikasi Ruangguru.

Metode pengumpulan data yang dilakukan dalam penelitian ini menggunakan teknik *web scraping* dengan bantuan aplikasi bawaan dari *Google Chrome* yaitu *Data Miner*. *Web Scraping* merupakan proses pengambilan data pada halaman sebuah *website* dengan bahasa HTML atau XHTML yang datanya dapat disimpan ke dalam format csv, txt, xls, dan sebagainya.

Analisis data yang dilakukan dalam penelitian ini menggunakan bantuan *software* RStudio 4.0.3 dan Microsoft Excel 2016. Berikut tahapan metode analisis data yang dilakukan:

1. *Scraping data*
2. Analisis deskriptif untuk mendeskripsikan gambaran umum dari aplikasi Ruangguru
3. *Pre-processing data* meliputi tahap *case folding*, *cleaning data*, normalisasi kata dengan perbaikan kata tidak baku, dan metode jarak *levenshtein distance*.
4. Pelabelan data ke dalam kelas positif dan kelas negatif
5. *Feature selection* melalui tahap *stopwords removal*, *stemming*, dan *tokenizing*
6. Pembobotan kata TF-IDF
7. Klasifikasi *Multinomial Naïve Bayes*
 - a. Membagi data menjadi data latih dan data uji dengan perbandingan sebesar 90% : 10%
 - b. Menghitung probabilitas prior ($P(c)$) dari tiap kelas sentimen
 - c. Menghitung probabilitas setiap *term* dari seluruh dokumen ulasan
 - d. Menghitung probabilitas setiap *term* terhadap kelas sentimen
 - e. Menghitung probabilitas dokumen *testing* ($P(c|d)$)
 - f. Menentukan kelas sentimen berdasarkan probabilitas kelas terbesar
 - g. Menyeleksi model klasifikasi menggunakan *10-fold cross validation*
 - h. Mengevaluasi model menggunakan *confusion matrix*
8. Visualisasi pengklasifikasian data menggunakan *word cloud*

4. HASIL DAN PEMBAHASAN

Scraping data dilakukan untuk mengambil dokumen teks berupa data ulasan pada aplikasi Ruangguru yang terdapat di *website google play* menggunakan aplikasi bawaan dari *google chrome* yaitu *data miner*. *Scraping data* dilakukan pada tanggal 9 Mei 2021 sebanyak 1500 ulasan pengguna dengan atribut yang diambil meliputi nama, ulasan, skor bintang, dan tanggal dibuatnya ulasan tersebut. Ulasan yang termasuk *spam* akan dihilangkan sehingga diperoleh 1484 data ulasan (5 Februari 2021 - 7 Mei 2021) yang dapat digunakan.

Mayoritas pengguna menilai aplikasi Ruangguru dengan bintang 5 yaitu sebesar 77% diikuti skor bintang 1 sebesar 8%, skor bintang 4 sebesar 7%, skor bintang 3 sebesar 5%, dan skor bintang 2 sebesar 3%. Semakin besar skor bintang yang diberikan mengindikasikan bahwa pengguna semakin puas terhadap aplikasi Ruangguru. Sebanyak 402 ulasan ditulis pada bulan Februari dan meningkat menjadi 628 ulasan pada bulan Maret. Peningkatan ini disebabkan adanya pelaksanaan ujian baik Ujian Tengah Semester maupun Ujian Tulis Berbasis Komputer (UTBK) yang merupakan ujian masuk perguruan tinggi pada bulan April, sehingga siswa perlu menyiapkan pemahamannya terhadap materi salah satunya dengan menggunakan aplikasi belajar *online* yaitu Ruangguru. Kemudian terdapat 388 ulasan yang ditulis pengguna pada bulan April dan 66 ulasan pada minggu pertama bulan Mei.

Tahap ini dilakukan untuk mengubah data sesuai dengan format yang dibutuhkan agar data siap untuk dianalisis. Berikut tahapan dari *pre-processing data*:

1. *Case folding* bertujuan untuk mengubah setiap bentuk kata menjadi huruf kecil guna menyamakan arti yang akan dibaca oleh mesin.
2. *Remove number* bertujuan untuk menghapus angka yang ada pada dokumen teks.
3. *Remove punctuation* bertujuan untuk menghapus tanda baca yang ada pada dokumen teks.
4. *Remove emoticon* bertujuan untuk menghapus *emoticon* yang ada pada dokumen karena terkadang penggunaannya tidak sesuai dengan sentimen yang ditulis oleh pengguna.
5. *Strip white spaces* dilakukan untuk menghapus spasi berlebih yang ada pada dokumen.
6. Normalisasi kata menggunakan kamus kata tidak baku bertujuan untuk mengubah kata tidak baku termasuk *slang* menjadi kata baku sesuai Kamus Besar Bahasa Indonesia (KBBI).
7. Normalisasi kata menggunakan metode jarak *Levenshtein Distance* dilakukan untuk menyempurnakan tahap normalisasi kata sebelumnya dengan menentukan jarak berdasarkan jumlah minimum perubahan dari *string* sumber ke *string* target melalui penambahan, penghapusan, atau pengubahan karakter. Salah satu contoh pembenaran kata menggunakan metode jarak *Levenshtein Distance* dapat dilihat pada Gambar 1 sebagai berikut:

		<i>string target (j)</i>					
			L	E	B	I	H
<i>string sumber (i)</i>		0	1	2	3	4	5
	L	1	0	1	2	3	4
	E	2	1	0	1	2	3
	B	3	2	1	0	1	2
	E	4	3	2	1	1	2
	H	5	4	3	2	2	1

Gambar 1. Normalisasi Kata Menggunakan Metode *Levenshtein Distance*

Hasil dari perhitungan *Levenshtein Distance* kata ‘LEBEH’ dengan kata ‘LEBIH’ menghasilkan nilai *edit distance* sebesar 1 yaitu pengubahan karakter ‘E’ pada kata ‘LEBEH’ menjadi karakter ‘I’ sehingga menjadi kata ‘LEBIH’ sebagai kata perbaikan.

Pelabelan data menggunakan kamus positif dan kamus negatif untuk menentukan skor sentimen ulasan data, diperoleh sebanyak 1071 ulasan dikategorikan ke dalam kelas positif, 256 ulasan dikategorikan ke dalam kelas negatif, dan sisanya sebanyak 157 ulasan dikategorikan ke dalam kelas netral. Ulasan dengan label netral akan dihilangkan karena dianggap tidak memiliki kecenderungan sentimen. Hasil pelabelan data ulasan menggunakan *sentiment scoring* tidak semuanya benar jika dibandingkan dengan pelabelan manual, dari 1484 ulasan yang dilakukan pelabelan terdapat 330 ulasan atau sebesar 22,24% data memiliki label yang berbeda. Hal ini dikarenakan terdapat ulasan yang berisi ungkapan kepuasan disertai dengan saran perbaikan sehingga penilaian manusia dengan mesin terkadang berbeda. Selanjutnya diambil 1000 data ulasan dengan label positif atau negatif paling relevan teratas dengan rincian sebanyak 789 ulasan berlabel positif dan 211 ulasan berlabel negatif.

Tahap *feature selection* atau seleksi fitur kata yang berpengaruh terhadap proses pengklasifikasian dengan menghapus kata umum melalui *stopwords removal*, menyamakan format setiap kata melalui proses *stemming* dan memecah dokumen teks menjadi kata tunggal melalui proses *tokenizing*.

Pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengetahui seberapa penting sebuah kata dalam dokumen yang hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Pembobotan Kata dengan TF-IDF

	academy	ada	adakan	adik	admin	agama	...	zenius	zoom
Ulasan ke-1	0	0	0	0	0	0	...	0	0
Ulasan ke-2	0	0	0	0	0	0	...	0	0
Ulasan ke-3	0	0	0	0	0,211	0	...	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ulasan ke-1000	0	0	0	0	0	0	...	0	0

Data ulasan yang telah diberi label dalam tahap pelabelan dibagi menjadi data latih dan data uji dengan perbandingan 90% : 10% yang berarti 900 ulasan digunakan sebagai data latih dan sisanya 100 ulasan digunakan sebagai data uji.

Metode *10-fold cross validation* digunakan untuk memperkirakan seberapa akurat model klasifikasi yang telah terbentuk dengan membagi dataset menjadi 10 *fold* berukuran sama, dimana 1 *fold* akan digunakan sebagai data uji dan sisanya 9 *fold* akan digunakan sebagai data latih. Dari 1000 data ulasan digunakan 100 data ulasan sebagai data uji dan 900 data ulasan sebagai data latih. Pengulangan dilakukan sebanyak 10 kali sehingga setiap ulasan tepat sekali menjadi data uji. Kinerja sistem *Multinomial Naïve Bayes* diukur dari rata-rata nilai akurasi seluruh *fold* yang dihasilkan. Model *Multinomial Naïve Bayes* menghasilkan rata-rata akurasi sebesar 88,20%. Tabel 3 merupakan tabel *confusion matrix* pada *fold* ke-8 sebagai *fold* dengan nilai akurasi tertinggi.

Tabel 3. *Confusion Matrix Fold ke-8*

Kelas Prediksi	Kelas Aktual	
	Negatif	Positif
Negatif	6	0
Positif	6	88

Penggunaan metode *Levenshtein Distance* untuk penyempurnaan tahap normalisasi kata mampu menambah nilai akurasi meskipun tidak signifikan. Tabel 4 menunjukkan hasil kinerja *Multinomial Naïve Bayes* dengan normalisasi kata menggunakan perbaikan kata tidak baku dan metode *Levenshtein Distance* dibandingkan dengan normalisasi kata menggunakan perbaikan kata tidak baku saja.

Tabel 4. Kinerja *Multinomial Naïve Bayes* untuk Normalisasi Kata Menggunakan dan Tanpa Menggunakan *Levenshtein Distance*

Evaluasi Kinerja Model	Normalisasi Kata	
	Perbaikan Kata Tidak Baku dan Metode <i>Levenshtein Distance</i>	
	Perbaikan Kata Tidak Baku	Perbaikan Kata Tidak Baku
Accuracy	88,20%	87,60%
Sensitivitas/Recall	96,22%	96,05%
Specifity	58,18%	57,18%
PPV/Precision	89,33%	88,66%
NPV	81,49%	79,70%

Terdapat peningkatan nilai rata-rata akurasi sebesar 0,60% dari yang sebelumnya sebesar 87,60% menjadi 88,20%.

Word cloud digunakan sebagai visualisasi data ulasan untuk mengetahui kata-kata apa saja yang sering digunakan oleh pengguna. Gambar 2 menunjukkan bahwa kata yang sering muncul dalam ulasan cenderung memiliki sentimen positif seperti kata “bagus”, “paham”, “bantu”, “mudah”, “baik”, dan “suka” yang berarti pengguna merasa terbantu dengan aplikasi Ruangguru karena materi yang disampaikan oleh pengajar berupa video mudah dipahami. Meskipun tidak sebanyak kemunculan kata positif, di dalam *word cloud* terdapat beberapa kata bersentimen negatif seperti kata “error”, “lambat”, dan “lama” yang mengindikasikan bahwa aplikasi Ruangguru terkadang masih *error*, pengaksesannya lama dan lambat.

- Priianti, K. R., & Wijaya, H. (2014). Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. *Jurnal Cybermatika*, 2(1), 1–6.
- Ruangguru. (2021). *Tentang Ruangguru*. <https://www.ruangguru.com/about-us> [Diakses tanggal 23 Februari 2021].
- Santoso, E. B., & Nugroho, A. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook. *Jurnal Eksplora Informatika*, 9(1), 60–69.