

Perbandingan Algoritma Regresi Logistik, *Support Vector Machine*, dan *Gradient Boosting* Pada Analisis Sentimen Data Komentar Siswa

Novita Ranti Muntiari ^{1✉}, Kharis Hudaiby Hanif ²⁾, Indah Chairun Nisa³⁾

^{1,3)} *Sistem Informasi, Politeknik Bisnis Kaltara, Tarakan, Indonesia*

¹novitaranti@poltekbiskal.ac.id

³indahcn@poltekbiskal.ac.id

²⁾ *Teknik Komputer, Universitas Borneo Tarakan, Tarakan, Indonesia*

hudaiby21@borneo.ac.id

Abstract— Digital Teacher Assessment (DITA) application used in teaching evaluations. In carrying out evaluations, there is no classification in using the DITA application, so that schools know that students' comments on teachers are positive or negative. Using the DITA application, students provide comments to the teacher. Comments will be grouped into negative, positive and neutral. Sentiment analysis is used to group comments based on their groups. Next lexicon based and words are counted using TF-IDF and evaluated. In this research, an algorithm is used Support Vector Machine (SVM), gradient boosting, and logistic regression. So, the research results compare the accuracy of the three algorithms with the algorithm gradient boosting has an accuracy rate of 97.5% has high accuracy. So, it can be concluded that the algorithm gradient boosting in student comment data using sentiment analysis is better.

Keywords—Sentiment Analysis, Student, SVM, Gradient Boosting

Intisari— Digital Teacher Assessment (DITA) aplikasi yang digunakan dalam evaluasi pengajaran, aplikasi ini dalam melakukan evaluasi belum adanya klasifikasi dalam mengelompokkan dalam menggunakan aplikasi DITA, agar sekolah tau bahwa komentar siswa terhadap Guru termasuk positif atau negatif. Menggunakan aplikasi DITA, siswa memberikan komentar kepada Guru. Komentar akan di kelompokkan menjadi negatif, positif, dan netral. Analisis sentimen digunakan dalam mengelompokkan komentar berdasarkan kelompoknya. Selanjutnya *lexicon based* dan kata-kata dihitung menggunakan TF-IDF dan dievaluasi. Dalam penelitian ini menggunakan algoritma *Support Vector Machine* (SVM), *gradient boosting*, dan *regresi logistik*. Maka hasil penelitian perbandingan akurasi dari tiga algoritma dengan algoritma *gradient boosting* memiliki tingkat akurasi sebesar 97,5% memiliki akurasi tinggi. Maka dapat disimpulkan bahwa algoritma *gradient boosting* dalam data komentar siswa menggunakan analisis sentimen lebih baik.

Kata kunci—Analisis Sentimen, Siswa, SVM, Gradient Boosting

I. PENDAHULUAN

Sekolah dalam dunia pendidikan diharapkan semakin memiliki kontribusi yang besar dalam mendidikan dan memberikan layanan pendidikan yang baik untuk siswa. Agar mutu sekolah dan pengajar (Guru) semakin berkualitas maka dilakukan evaluasi pengajaran.

Evaluasi pengajaran dilakukan dalam menganalisis komentar siswa menggunakan aplikasi *Digital Teacher Assessment* (DITA), apakah dalam sistem pengajaran dianggap baik, kurang ataupun masih perlu ditambahkan dan Komentar akan dikelompokkan menjadi negatif, positif, dan netral. Berasaskan komentar selanjutnya menggunakan analisis sentimen Upaya untuk menentukan pendapat atau pendapat dalam kalimat dokumen memiliki sentuhan positif atau negatif pengertian dari analisis sentimen. [1],[2].

Machine learning yaitu bagaimana menggunakan komputer untuk belajar dengan mudah dan cepat. Metode dan algoritma khusus digunakan untuk memprediksi, mengenali pola, dan klasifikasikan objek. Algoritma dalam *machine learning* yaitu KNN, *decision tree*, *random forest* dan lain sebagainya [3].

Penelitian menggunakan algoritma seperti, *regresi logistik*, *Support Vector Machine* (SVM), dan *gradient boosting*. Algoritma *regresi logistik* adalah untuk melihat bagaimana variabel prediksi dan variabel respons berhubungan satu sama lain [4]. Salah satu algoritma *supervised learning* yang paling banyak digunakan untuk regresi dan klasifikasi adalah algoritma SVM. Dibandingkan dengan metode lainnya, dalam pemodelan klasifikasi, konsep *support vector machine* lebih maju dan lebih jelas secara matematis. Selain itu, algoritma ini dapat menangani masalah regresi dan klasifikasi menggunakan metode non linear maupun linear.

Algoritma *gradient boosting* sering digunakan karena skalabilitas dan efisiensi dalam menyelesaikan berbagai masalah regresi dan klasifikasi. [5]

Lexicon based adalah metode untuk menganalisis tanpa melakukan pelatihan sebelumnya [6]. Daftar kata telah ditentukan sebelumnya dalam menggunakan metode [7]. Praktis dan mudah digunakan, ini pilihan yang bagus untuk melihat perasaan orang dalam komentar. [8]. Kata diberi bobot. TF-IDF yaitu bagaimana mengubah data menjadi numerik setiap fitur atau kaya dari data teks [6].

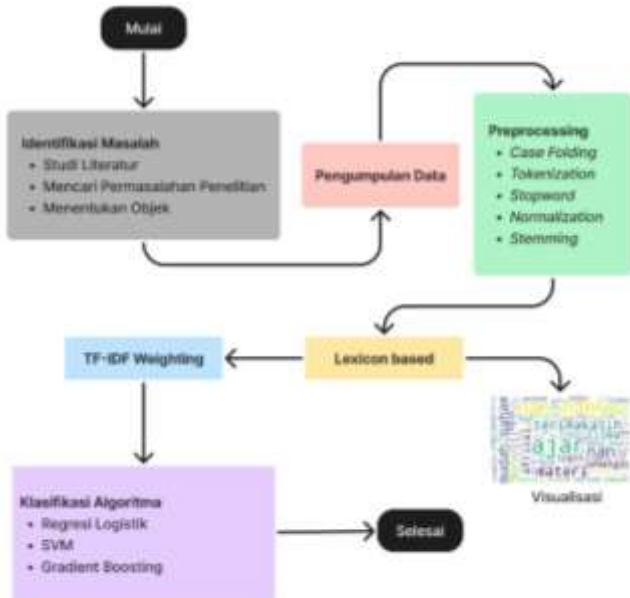
Penelitian terdahulu berjudul "Perbandingan Metode *Support Vector Machine* dan *Decision Tree* Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online" yang dilakukan oleh Khoirul, dkk [9] Tujuan penelitian adalah data ulasan Google Play di analisis sentiment dan mengevaluasi keakuratan metode *SVM* dalam kelompok ulasan, yaitu positif dan negatif. Lalu menggunakan metode *Decision Tree*. Hasil penelitian melalui metode *SVM* nilai akurasi 90,20%, dan metode *Decision Tree* nilai akurasi 89,80%. Oleh karena itu, *Decision Tree* nilai akurasinya lebih rendah dari pada *SVM*.

Penelitian akan menganalisis sentimen komentar siswa terhadap guru dengan algoritma *SVM*, *regresi logistik*, dan *gradient boosting*. Dibandingkan penelitian sebelumnya, penelitian ini menggunakan proses *preprocessing* untuk menganalisis sentimen komentar siswa dan kemudian menggunakan *lexicon* untuk menganalisis sentimen. Penelitian sebelumnya menggunakan penentuan sentimen dengan manual agar dapat mengetahui hasil pengelompokan sentimen negatif, positif, dan netral. Selanjutnya, penelitian ini membandingkan tingkat akurasi masing-masing algoritma dengan menggunakan pembobotan kata TF-IDF.

II. METODOLOGI PENELITIAN

A. Alur Penelitian

langkah-langkah yang tepat diperlukan. Alur penelitian di tunjukkan Gambar 1.



Gambar 1. Alur Penelitian

Alur penelitian Gambar 1 terdapat dari beberapa tahap, yang masing-masing dijelaskan :

1) Identifikasi Masalah

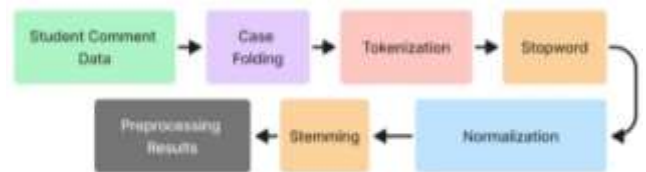
Penelitian ini awalnya melakukan identifikasi masalah. Ini melibatkan memeriksa literatur, menemukan permasalahan penelitian, dan menetapkan objek penelitian. Selanjutnya studi literatur yaitu menggunakan, mencari, dan belajar literatur tentang analisis sentimen, seperti buku, jurnal dan kertas, dilakukan pada tahap studi literatur. Objek penelitian ini yaitu analisis sentimen pada komentar siswa dengan menggunakan algoritma *regresi logistik*, *SVM*, *gradient boosting*, dan *lexicon based*. Data tanggapan siswa yang dikumpulkan data difokuskan. Data dikumpulkan akan dikategorikan menjadi netral, positif, dan negatif. Untuk mengelompokkan komentar dalam kelompok, analisis sentimen diperlukan.

2) Pengumpulan data

Tahapan ini dikumpulkan data komentar siswa. Kemudian, data ini dilabelkan secara manual.

3) *Preprocessing*

Diharapkan analisis sentimen menjadi lebih akurat, proses menghilangkan *noise* diperlukan karena data komentar siswa mengandung banyak *noise*. Gambar 2 menunjukkan rute *preprocessing* untuk penelitian ini.



Gambar 2. Alur Preprocessing

4) *Lexicon Based*

Setelah proses *preprocessing* selesai, data diklasifikasikan dengan *lexicon based*. Membagi analisis sentimen menjadi : netral, positif, dan negatif [10]. *Lexicon based* hasilnya dibandingkan sentimen manual. Kemudian, visualisasi dibuat dari hasil *lexicon based*.

5) *Pembobotan TF-IDF*

Nilai bobot kata TF-IDF dihitung dokumen. Proses yaitu IDF dan TF. IDF (*Inverse Document Frequency*) setiap kata yang kadang tampil dihitung jumlah dokumen. TF (*Term Frequency*) menghitung setiap kata tampil dokumen, dan mencari jumlah kemunculan terbesar yang memiliki nilai tertinggi. Persamaan (1) menunjukkan rumus pembobotan kata TD-IDF.

$$W_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \log \frac{N}{df_t} \tag{1}$$

dimana $W_{t,d}$ adalah bobot TF-IDF, $tf_{t,d}$ adalah jumlah frekuensi kata, idf_t adalah jumlah *inverse* frekuensi dokumen tiap kata, df_t adalah jumlah frekuensi dokumen tiap kata, dan N adalah jumlah total dokumen.

Perkalian nilai TF dan IDF yang dihasilkan dari TF-IDF. nilai TF-IDF lebih rendah jika, kata sering tampil pada setiap dokumen, dan nilai TF-IDF lebih tinggi jika kata tidak sering muncul pada setiap dokumen. Pembobotan TF-IDF tanpa normalisasi dalam penelitian ini.

6) *Klasifikasi Algoritma*

Klasifikasi algoritma pada dataset menggunakan *lexicon*, pembobotan TF-IDF. Kemudian, algoritma *regresi logistik*, SVM, dan peningkatan *gradient* diproses.

Hasil dari klasifikasi algoritma berdasarkan sentimen *lexicon* yaitu tingkat akurasi akan dibandingkan dari masing-masing algoritma yang menggunakan parameter kinerja algoritma, yaitu *recall*, *f1-measure*, *presisi* dan akurasi.

B. Pembagian Data

Dalam pembagian data terdapat data uji dan data latih adalah dua bagian dari data yang diolah oleh pengajaran mesin. Data latihan melatih model klasifikasi, dan data uji untuk pengujian. Gambar 3 menunjukkan 70% data klasifikasi dan 30% pengujian.



Gambar 3. Pembagian Data

C. Klasifikasi Model

1) Regresi Logistik

Dengan menggunakan regresi logistik, dapat menemukan bagaimana variabel respons dichotomus—yaitu variabel dengan skala nominal atau ordinal yang termasuk dalam dua kategori—berhubungan dengan satu atau lebih prediktor berskala kategori atau kontinu. [11]. Kepekaan peluang (fkp) untuk variabel distribusi logistik random seperti yang ditunjukkan pada persamaan (2).

$$f(x) = \frac{\exp\left\{\frac{x - \mu}{\tau}\right\}}{\tau[1 + \exp\left\{\frac{x - \mu}{\tau}\right\}]^2}, -\infty \leq x \leq \infty; \tau > 0 \quad (2)$$

dengan mean μ dan variansi $\sigma^2 = \frac{\pi^2\tau^2}{3}$

Regresi logistik biner mencakup regresi logistik dengan respon biner, dll. Teknik analisis data yang bertujuan untuk mengidentifikasi hubungan antara variabel respons (y) yang biner atau *dichotomous* dan variabel prediktor (x) yang kategorik atau kontinu.

2) *Support Vector Machine*

Akurasi pengenalan pola SVM sangat tinggi. Prinsip kerja SVM adalah memisahkan dua kategori di ruang input dan mencari titik maksimal. Sementara kekurangan SVM membuatnya sulit digunakan dalam masalah jumlah sample besar [12]. Penelitian ini akan berkonsentrasi pada penggunaan linier dalam kelompok dua kategori negatif dan positif. Fungsi mencari hyperplane digambarkan dalam persamaan (3), dan garis tepi hyperplane untuk membagi kelas positif dan negatif digambarkan dalam persamaan (4).

$$(w \cdot x) + b = 0 \quad (3)$$

$$(w \cdot x) + b = +1 \text{ (untuk kelas positif)} \quad (4)$$

$$(w \cdot x) + b = -1 \text{ (untuk kelas negatif)}$$

3) *Gradient Boosting*

Algoritma *gradient boosting* untuk digunakan dalam klasifikasi, algoritma peningkatan gradient termasuk pembelajaran yang diawasi berbasis pohon keputusan. Keputusan dari berbagai mesin pembelajaran digabungkan, dan ensemble memprediksi kelas mana yang akan menerima mayoritas "suara". Ini adalah gambaran sederhana dari konsep kelompok. Pokok peningkatan progresif dimulai dengan pohon klasifikasi awal, yang kemudian disesuaikan dengan mengurangi fungsi kerugian untuk pohon baru [13]. Persamaan (5) rumus algoritma *gradient boosting*.

$$-\log L1 = -\sum_{i=1}^N y_i \log(odds) + \log(1 + e^{\log(odds)}) \quad (5)$$

D. Evaluasi Model

Parameter *recall*, *f1-measure*, *presisi*, dan akurasi yaitu beberapa parameter pengukuran kinerja algoritma yang dapat diketahui berdasarkan matrix confusion pada Tabel 1 [14].

Tabel 1. Tabel Confusion Matrix

	Label atau kelas	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Positive</i>
Negatif	<i>False Negative</i>	<i>True Negative</i>

Rumus untuk menghitung ketepatan ditunjukkan dalam persamaan (6). Persamaan (7) rumus menghitung *recall*, dan persamaan (8) rumus menghitung *F-measure*. Rumus untuk mengukur ketepatan menggunakan persamaan (9).

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

$$F - measure = \frac{2 \times recall \times precision}{recall + precision} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

dimana FP (*False Positive*), TP (*True Positive*), TN (*True Negative*) dan FN (*False Negative*).

Nilai tertinggi adalah 1 dan nilai terendah adalah 0. Salah satu perhitungan yang sering digunakan untuk mengevaluasi kinerja sebuah algoritma adalah akurasi, yang dihitung dengan membandingkan rasio jumlah data yang diprediksi secara akurat oleh algoritma dengan jumlah total data yang ada dalam dataset.

III. HASIL DAN PEMBAHASAN

Berdasarkan analisis pada data siswa menggunakan regresi logistik, gradient boosting, dan SVM dibahas. Penelitian ini menggunakan aplikasi DITA dan mengolah data dengan Jupyter Notebook dan mengumpulkan data komentar siswa.

Setelah data komentar dikumpulkan, label sentimen digunakan secara manual untuk mengklasifikasikan sentimen yaitu netral, positif, dan negatif. Ini ditunjukkan pada Gambar 4 dalam format CSV.

No	A	B	C
1	Meneri tidak bisa	Komentar	Klasifikasi
1	3105200165	okeelah tugasnya jgn yg susah-susah ya pak :v	Positif
2	3105200178	Sangat baik terba	Positif
4	3105200184	Mungkin bisa di beri materi tertulis	Netral
3	3105200379	Its ok	Positif
6	3105200012	Sudah cukup baik	Positif
7	3105200034	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Netral
8	3105200019	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Positif
9	3105200084	Tidak ada	Netral
10	3105200181	Baik sekali	Positif
11	3105200117	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	Positif
12	3105200482	Menurut saya sudah baik	Positif
13	3105200189	baik	Positif
14	3105200949	Terimakasih pak semoga sehat selalu	Positif
15	3105200188	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	Positif
16	3105200181	Maaf	Positif
17	3105200188	semoga kelulusan lebih baik lagi dari sebelumnya	Negatif
18	3105200180	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	Positif
19	3105200682	kurang kerajinan tugas videonya pak	Negatif
20	3105200015	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	Negatif
21	3105200185	lakin baik lagi	Negatif
22	3105200188	Tetap belajar pak	Netral
23	3105200124	Terimakasih telah membimbing i lahnya	Positif
24	3105200042	semoga kelas 11 selanjutnya juga bisa tetap baik	Netral
25	3105200685	The best	Positif
26	3105200185	baik	Positif
27	3105200018	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	Positif
28	3105200181	terima kasih pak	Positif
29	3105200029	semoga bisa lebih baik untuk kedepannya	Netral

Gambar 4. Data Komentar Labelling Data Manual

Data 676 komentar akan diproses, dengan isi data berdasarkan nomor induk siswa, komentar, dan klasifikasi, Seperti yang ditunjukkan pada Gambar 5.

Nomor Induk Siswa	Komentar	Klasifikasi	
0	3105200165	okeelah tugasnya jgn yg susah-susah ya pak :v	Positif
1	3105200178	Sangat baik terba	Positif
2	3105200184	Mungkin bisa di beri materi tertulis	Netral
3	3105200379	Its ok	Positif
4	3105200012	Sudah cukup baik	Positif
...
671	3105190386	Baikk	Positif
672	3105190548	Materi mudah di pahami	Positif
673	3105200051	Pembelajaran yang disampaikan oleh Pak Agus cu...	Positif
674	3105190500	cukup baik	Positif
675	3105190485	Baik	Positif

Gambar 5. Tampilan Data Awal

Data banyak mengandung suara dan tidak memiliki informasi penting untuk proses klasifikasi sentimen, seperti simbol, tanda baca, angka, dan penggunaan bahasa asing, tahap *preprocessing* diperlukan. Tahapan *Preprocessing case folding, tokenization, stopword, normalization, dan stemming*.

Tahapan *case folding* huruf besar mengubah menjadi huruf kecil atau besar menjadi huruf kecil, Anda dapat menggunakan cara *case folding*. Contoh "Mungkin" diubah menjadi "mungkin", dan ini adalah contoh holding *case pseudocode*.

```

Deklasrasi variabel :
Komentar = "Mungkin"
Proses lower case
Cetak Komentar
Komentar_lower ="mungkin"
    
```

Tahapan *tokenization* menghapus karakter yang memengaruhi proses klasifikasi sentimen. Ini termasuk baca titik (.), koma (,), simbol, dan karakter huruf dan angka. Ini juga menghapus. Tokenisasi pseudocode

```

Deklasrasi variabel :
Komentar_lower = "mungkin bisa "
Proses tokenization
Cetak Komentar_token
Komentar_token ="[mungkin, bisa]"
    
```

Tahapan *stopword* untuk menghilangkan kata-kata yang tidak perlu. Berikut *pseudocode stopwords*.

```

Deklasrasi variabel :
Komentar_token = "[materi, mudah, di, pahami]"
Proses stopwords
Cetak Komentar_token_stopword
Komentar_token_stopword =[materi, mudah, pahami]
    
```

Tahapan *normalization* digunakan untuk mengubah kata yang memiliki *noise* menjadi kata baku. *Pseudocode normalization*.

```

Deklasrasi variabel :
Komentar_token_stopword = "[materi, yg, pahami]"
Proses normalization
Cetak Komentar_normalization
Komentar_normalization =[materi, yang, pahami]
    
```

Tahapan *stemming* digunakan untuk mengembalikan kata. Contoh "Tugasnya" menjadi "tugas". Berikut *pseudocode stemming*.

```

Deklasrasi variabel :
Komentar_normalization = "[Tugasnya]"
Proses stemming
Cetak Komentar_stemming
Komentar_stemming =[tugas]
    
```

Preprocessing pada Gambar 6 menunjukkan hasil dari tahapan dan akhirnya menjadi komentar baru setelah selesai.

Komentar Asli	Komentar Asli	Komentar Asli	Komentar Asli	Komentar Asli	Komentar Asli	Komentar Asli	Komentar Asli
okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v	okeelah tugasnya jgn yg susah-susah ya pak :v
Sangat baik terba	Sangat baik terba	Sangat baik terba	Sangat baik terba	Sangat baik terba	Sangat baik terba	Sangat baik terba	Sangat baik terba
Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis	Mungkin bisa di beri materi tertulis
Its ok	Its ok	Its ok	Its ok	Its ok	Its ok	Its ok	Its ok
Sudah cukup baik	Sudah cukup baik	Sudah cukup baik	Sudah cukup baik	Sudah cukup baik	Sudah cukup baik	Sudah cukup baik	Sudah cukup baik
Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan	Alasan awal sudah saat kegiatan pembelajaran saya melakukan kebutuhan
Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen	Belakangan saya sangat berminat atas wktu dan sumbernya dalam pen
Tidak ada	Tidak ada	Tidak ada	Tidak ada	Tidak ada	Tidak ada	Tidak ada	Tidak ada
Baik sekali	Baik sekali	Baik sekali	Baik sekali	Baik sekali	Baik sekali	Baik sekali	Baik sekali
terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi	terima kasih pak, sudah mengajar kami dengan baik dan sabar, sangat rapi
Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik	Menurut saya sudah baik
baik	baik	baik	baik	baik	baik	baik	baik
Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu	Terimakasih pak semoga sehat selalu
terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku	terima kasih pak sudah sabar dan berkeadilan kami sangat hami paku
Maaf	Maaf	Maaf	Maaf	Maaf	Maaf	Maaf	Maaf
semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya	semoga kelulusan lebih baik lagi dari sebelumnya
terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku	terimakasih pak sudah sabar dan berkeadilan kami sangat hami paku
kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak	kurang kerajinan tugas videonya pak
maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak	maafnya awal saja karena id banyak kebutuhan yang ditunggu semoga tidak
lakin baik lagi	lakin baik lagi	lakin baik lagi	lakin baik lagi	lakin baik lagi	lakin baik lagi	lakin baik lagi	lakin baik lagi
Tetap belajar pak	Tetap belajar pak	Tetap belajar pak	Tetap belajar pak	Tetap belajar pak	Tetap belajar pak	Tetap belajar pak	Tetap belajar pak
Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya	Terimakasih telah membimbing i lahnya
semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik	semoga kelas 11 selanjutnya juga bisa tetap baik
The best	The best	The best	The best	The best	The best	The best	The best
baik	baik	baik	baik	baik	baik	baik	baik
terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan	terima kasih sudah baik dalam mengajar, semoga saran dari saya materi yang disampaikan
terima kasih pak	terima kasih pak	terima kasih pak	terima kasih pak	terima kasih pak	terima kasih pak	terima kasih pak	terima kasih pak
semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya	semoga bisa lebih baik untuk kedepannya

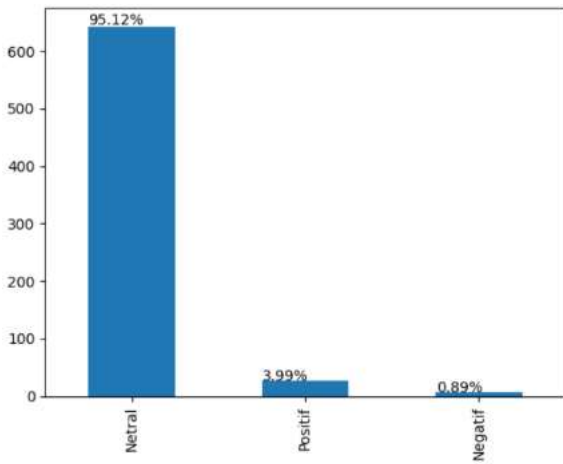
Gambar 6. Proses reprocessing

Lexicon based mengkategorikan sentimen dan membedakan sentimen netral, positif, dan negatif. Nilai polarity menunjukkan bahwa nilainya kurang dari 0 negatif, nilainya lebih dari 0 sentimen netral, dan jika nilainya lebih dari 0 positif, dilihat pada Gambar 7.

Komentar_tekun_stemmed	Komentar_baru	Compound_Score	Sentiments
[suka, tugas, jgn, yg, susah-susah, ya]	ohi tugas jgn yg susah-susah ya	0.000	Netral
[bakterial]	bakteria	0.000	Netral
[materi, tulis]	materi tulis	0.000	Netral
[ts, ok]	ts ok	0.296	Positif
[]		0.000	Netral

Gambar 7. Klasifikasi Proses *Lexicon Based*

Berdasarkan klasifikasi berbasis *lexicon*, didapat tiga kelompok, dengan komentar siswa yang paling banyak masuk ke kelompok netral (95,12%) dan kelompok negatif paling rendah 0,89%.



Gambar 8. Klasifikasi *Lexicon based*

Hasil klasifikasi yang didasarkan pada *lexicon* dilengkapi dengan hasil visualisasi disusun berdasarkan kelompok. Gambar 9 menunjukkan kata-kata yang termasuk dalam kelompokan sentimen [15] tersebut. Sentimen positif (a), negatif (b), dan netral (c)



Gambar 9. Kelompok sentimen

Menggabungkan hasil visualisasi yang ditampilkan berdasarkan hasil kelompok. Gambar 9 menunjukkan kata-kata yang berasal dari kelompokan sentimen [15] tersebut, yang disebut sebagai positif (a), negatif (b), dan netral (c).

Setelah proses klasifikasi berbasis *lexicon*, proses pembobotan TF-IDF melakukan komentar baru. Kode program dan tampilannya ditunjukkan pada Gambar 10.

```
#Jumlah kalimat dan kata
document.shape
(676, 424)

#Menjumlahkan TF-IDF untuk setiap kalimat
result = np.sum(document,axis=1)
result.shape
(676, )

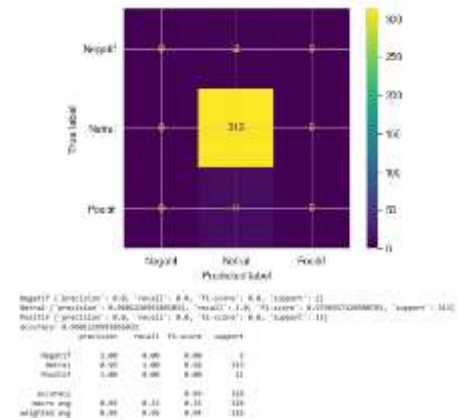
#Ditampilkan TF-IDF setiap kalimat dari kecil ke besar
sorted(result)
```

Gambar 10. Hasil Pembobotan TFIDF

Gambar 10 menunjukkan hasil pembobotan TF-IDF, yang menunjukkan bahwa ada 424 kata atau term dari 676 komentar. Nilai TF-IDF dari hasil kode program.

Hasil evaluasi algoritma regresi logistik, SVM, dan peningkatan *gradient* didapatkan dengan menggunakan 70% data uji dan 30% data uji sebagai berikut:

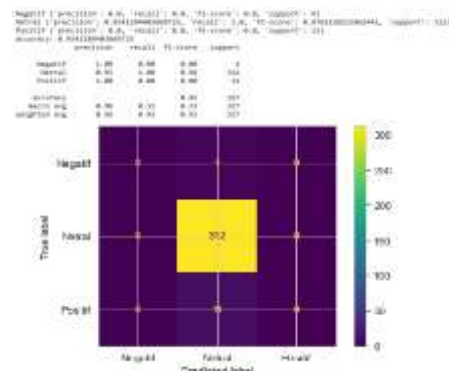
Gambar 11 menunjukkan hasil dari algoritma regresi logistik.



Gambar 11. Hasil Regresi Logistik

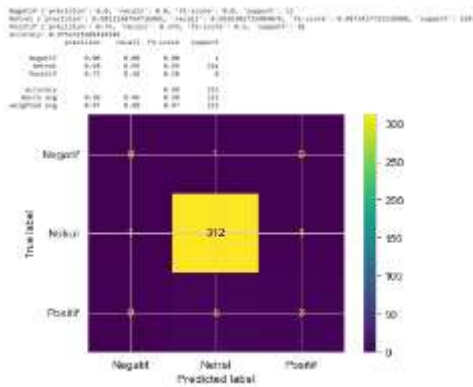
Gambar 11 menunjukkan hasil algoritma regresi logistik, yang menghasilkan nilai akurasi 0,96 persen dan menunjukkan *confusion matrix*.

Gambar 12 hasil SVM yaitu 0,95 % ditunjukkan dalam *confusion matrix*.



Gambar 12. Hasil SVM

Gambar 13 hasil gradient boosting dengan nilai akurasi yaitu 0,975 % ditunjukkan *confusion matrix*.



Gambar 13. Hasil Gradient Boosting

Dengan menggunakan 70 persen data uji dan 30 persen data uji, hasil algoritma regresi logistik, SVM, dan peningkatan gradient dapat dilihat. Ini juga memungkinkan untuk mengetahui seperti *presisi*, *recall*, *f1-measure*, dan akurasi. Perbandingan algoritma regresi logistik, SVM, dan peningkatan gradient dapat dilihat pada Tabel 2.

Tabel 2. Hasil Perbandingan

No	Nama Algoritma	Nilai Akurasi
1	Regresi Logistik	96 %.
2	SVM	95 %
3	Gradient Boosting	97,5 %

Seperti yang ditunjukkan dalam Tabel 2, ada hasil perbandingan akurasi dari algoritma regresi logistik, SVM, dan gradient boosting. Algoritma gradient boosting menunjukkan tingkat akurasi tertinggi, yaitu 97,5%, sehingga dapat disimpulkan bahwa menggunakan algoritma ini lebih baik untuk mengklasifikasikan data analisis sentimen komentar siswa.

IV. KESIMPULAN

Berdasarkan hasil pengujian menggunakan dengan pelatihan sebesar 70% dan data pengujian sebesar 30%, hasil pengujian menggunakan regresi logistik, SVM, dan peningkatan gradient ditemukan bahwa ketepatan klasifikasi menggunakan metode regresi logistik mencapai 96% ketepatan, 96% ketepatan, 96% recall, dan 96% skor f1. Secara keseluruhan, hasil ketepatan klasifikasi dengan metode gradient boosting menunjukkan hasil yang lebih baik daripada metode regresi logistik, SVM, dan gradient boosting. Hasilnya menunjukkan ketepatan sebesar 97,5%, keakuratan sebesar 97%, keakuratan sebesar 98%, dan skor f1 sebesar 97%.

REFERENSI

[1] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019.

[2] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear," *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017.

[3] N. R. Muntiari and K. H. Hanif, "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning," *J. Ilmu Komput. dan Teknol.*, vol. 3 (1), pp. 1–6, 2022.

[4] M. R. Fauzaan and T. A. Hermawan, "Klasifikasi Pengguna Internet berdasarkan Aktifitas Akses Internet Advertisement dengan Menggunakan Metode Regresi Logistik," *Annu. Res. Semin.* vol. 4, no. 1, pp. 978–979, 2019.

[5] I. Komputer, I. Pengetahuan, U. L. Mangkurat, J. Jenderal, A. Yani, and K. Selatan, "Perbandingan Metode Extreme Gradient Boosting Dan," pp. 373–382, 2019.

[6] Y. Wang, G. Huang, J. Li, H. Li, Y. Zhou, and H. Jiang, "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis," *IEEE Access*, vol. 9, no. 1, pp. 37075–37085, 2021, doi: 10.1109/ACCESS.2021.3062654.

[7] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Na'ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021.

[8] M. Alfi, R. Reynaldhi, and Y. Sibaroni, "Analisis Sentimen Review Film pada Twitter menggunakan Metode Klasifikasi Hybrid SVM, Naïve Bayes, dan Decision Tree," vol. 8, no. 5, pp. 10127–10137, 2021.

[9] K. A. Rokhman, B. Beriliana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021.

[10] N. N. Wilim and R. S. Oetama, "Sentiment Analysis About Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, And Decision Tree," *IJNMT (International J. New Media Technol.*, vol. 8, no. 1, pp. 50–56, 2021, doi: 10.31937/ijnmt.v8i1.1965.

[11] A. R. S. Darwanto, Taza Luzia Viarindita, and Yekti Widyaningsih, "Analisis Regresi Logistik Binomial dan Algoritma Random Forest pada Proses Pengklasifikasian Penyakit Ginjal Kronis," *J. Stat. dan Apl.*, vol. 5, no. 1, pp. 1–14, 2021.

[12] A. Budianto, R. Ariyuana, and D. Maryono, "Perbandingan K-Nearest Neighbor (Knn) Dan Support Vector Machine (Svm) Dalam Pengenalan Karakter Plat Kendaraan Bermotor," *J. Ilm. Pendidik. Tek. dan Kejuru.*, vol. 11, no. 1, p. 27, 2019, doi: 10.20961/jiptek.v11i1.18018.

[13] S. E. Suryana, B. Warsito, and S. Suparti, "Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank," *J. Gaussian*, vol. 10, no. 4, pp. 617–623, 2021, doi: 10.14710/j.gauss.v10i4.31335.

[14] W. Li, P. Liu, Q. Zhang, and W. Liu, "An improved approach for text sentiment classification based on a deep neural network via a sentiment attention mechanism," *Futur. Internet*, vol. 11, no. 4, 2019, doi: 10.3390/FI11040096.

[15] I. Hilmy Khairi Idris, Mochammad Ali Fauzi, "Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square", *Syst. Inf. Syst. Informatics J.*, vol. 3, no. 1 pp. 25–32, 2017.