Sentimen Analisis Komentar Instagram Untuk Deteksi Cyberbullying Dengan Model Gradient Boosting

Bagus Putu Eka Wijaya[1)], Dandy Pramana Hostiadi[2)], Putu Desiana Wulaning Ayu[3)]
[1] Magister Program, Departement of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
[2] Departement of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia e-mail: 222012011@stikom-bali.ac.id[1)], dandy@stikom-bali.ac.id[2)], wulaning_ayu@stikom-bali.ac.id[3)]

Abstrak

Pertumbuhan media sosial yang seiring dengan peningkatan jumlah pengguna internet di Indonesia menunjukkan bahwa di era komunikasi terbuka yang semakin meluas, masyarakat dapat lebih mudah mengakses pembicaraan. Cyberbullying atau perundungan di media digital merupakan salah satu bentuk kekerasan yang dilakukan dengan menggunakan teknologi komunikasi dan informasi, dengan hasil penelitian menunjukkan Model Gradient Boosting menunjukkan kinerja terbaik pada 10-fold Cross Validation dengan kombinasi nilai evaluasi tertinggi (AUC = 0,903, CA = 0,818, F1 = 0,818, Precision = 0,822, Recall = 0,818, MCC = 0,641). Penambahan jumlah fold menjadi 20-fold sedikit menurunkan performa, menunjukkan bahwa 10-fold Cross Validation adalah pilihan yang optimal untuk menjaga keseimbangan antara kinerja dan efisiensi model. Hasil ini menunjukkan bahwa model Gradient Boosting sangat menjanjikan untuk tugas klasifikasi Bullying. Validasi silang 10 lipatan memberikan hasil terbaik secara konsisten di berbagai metrik. Model ini dapat diandalkan untuk mengidentifikasi kasus Bullying dengan akurasi yang baik.

Kata kunci: *Gradient Boosting, Cyberbullying, Machine Learning*

1. Pendahuluan

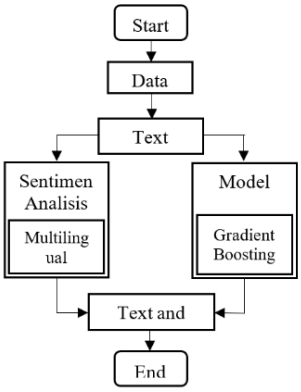
Pesatnya perkembangan teknologi informasi dan komunikasi (TIK) dapat meningkatkan pekerjaan menjadi lebih efisien dan efektif di berbagai sektor salah satunya media sosial, dan merupakan alat komunikasi yang mudah diakses diseluruh dunia[1]. Salah satu *platform* jejaring sosial yang cukup terkenal adalah *Instagram*[2]. Pertumbuhan media sosial yang seiring dengan peningkatan jumlah pengguna internet di Indonesia menunjukkan bahwa di era komunikasi terbuka yang semakin meluas, masyarakat dapat lebih mudah mengakses pembicaraan[3].

Cyberbullying atau perundungan di media digital merupakan salah satu bentuk kekerasan yang dilakukan dengan menggunakan teknologi komunikasi dan informasi, khususnya internet dan media sosial[1] melalui jejak digital yang ditinggalkan, seperti tulisan, gambar, atau video[4]. *Cyberbullying* menurut Taskin Tanrikulu (2015) Perilaku agresif dan berulang yang dilakukan melalui media elektronik, seperti pesan teks, email, atau media sosial. *Cyberbullying* dapat mencakup tindakan mempermalukan, mengancam, atau menyebarkan gosip tentang korban[5].

Penelitian terkait tentang *cyberbullying* yang dilakukan oleh Nadya Lestari dkk (2023) tentang "Penerapan Seleksi *Fitur Particle Swarm Optimizition* pada Klasifikasi Texs Komentar *Cyberbullying Instagram*"[6], menunjukan bahwa penambahan seleksi fitur dapat mengurangi jumlah fitur kata hingga 40%. Penerapan seleksi fitur menggunakan *Particle Swarm Optimization (PSO)* telah berhasil meningkatkan nilai *True Positif* dan *False Positif*, nilai *recall* dan *f1-score* masing-masing sebesar 5,36% dan 0,57%. Namun, peningkatan tersebut juga menyebabkan penurunan pada *True Negatif* dan *False Negatif*, yang berdampak pada penurunan akurasi, presisi, dan *AUC*, masing-masing sebesar 1,25%, 4,25%, dan 1,09%. Penelitian Nur Fajriyani dkk (2023) dimana penelitian berjudul "Optimasi *Hyperparameter* pada *Neural Network* (Studi Kasus: Identifikasi Komentar *Cyberbullying Instagram)*"[7] Model *Neural Network* dengan default *hyperparameter*. Hasil evaluasi dari empat parameter, yaitu *precision, recall, F1-score*, dan akurasi, menunjukkan konsistensi pada kisaran 80% di seluruh model yang dikembangkan. Selain itu, *Bayesian Optimization* mampu melakukan optimasi *hyperparameter* pada *Neural Network*.

2. Metode Penelitian

Berikut ini adalah alur penelitian yang dilakukan:



Gambar 1. Alur Penelitian

Alur penelitian pada gambar 1 terdapat proses diantaranya dataset *Cyberbullying* dari *Instagram* yang diinput menggunakan aplikasi *Orange 3 Vesri 3.37.0*, kemudian dataset masuk ke proses *Text Mining* diantaranya melalui *Corpus* untuk melatih model pembelajaran mesin, terutama dalam tugas-tugas seperti klasifikasi teks, analisis sentimen, penerjemahan mesin, dan *chatbot*. *Preprocess Text* merupakan langkah awal pada analisis sentimen teks. Teknik *pre-processing* yang sesuai juga dapat meningkatkan performa model *classifier*-nya, Tahap ini meliputi proses *Transformation, Tokenizing, Filtering* dan *Bag of Word* Digunakan dalam pengolahan bahasa alami (*Natural Language Processing*, NLP).

2.1 Gradient Boosting

Gradient Boosting adalah teknik ensemble dalam pembelajaran mesin yang digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan beberapa model prediksi yang lebih sederhana, biasanya berupa *decision trees* (pohon keputusan), untuk membentuk model prediksi yang lebih kuat. Metode pembelajaran mesin yang populer untuk masalah regresi dan klasifikasi[1], serta terdapat beberapa hyperparameter yang dapat disesuaikan guna meningkatkan kinerjanya[8].

2.2 Text and Score

Digunakan dalam Natural Language Processing (NLP) dan Machine Learning untuk merepresentasikan hubungan antara teks (text) dengan skor tertentu (score). Keduanya sering digunakan untuk memberikan nilai atau evaluasi terhadap teks berdasarkan kriteria tertentu.

2.3 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi akurasi dan kinerja algoritma dalam klasifikasi serta prediksi menunjukan hasil identifikasi[9] atribut pada data pengujian. Dalam *confusion matrix*, kita menemukan empat komponen penting, yaitu *False Negative (FN)*, *False Positive (FP)*, *True Negative (TN)*, dan *True Positive (TP)*. Berikut ini adalah tabel yang menggambarkan confusion matrix tersebut.

Tabel 1. Asumsi *Confusion Matrix*

Kelas Prediksi	*Kelas Aktual*	
	Positive	*Negative*
Positive	TP	FP
Negative	FN	TN

TP adalah kondisi di mana baik prediksi maupun nilai aktualnya benar. Sementara itu, FN merujuk pada kasus di mana prediksi tidak tepat meskipun nilai aktualnya benar. Di sisi lain, FP terjadi saat prediksi benar tetapi nilai aktualnya tidak sesuai. Untuk mengevaluasi kinerja model, terdapat berbagai metrik performa, antara lain akurasi, *recall*, dan presisi. Ketiga nilai tersebut dapat diperoleh melalui rumus-rumus yang tertera dalam Tabel 2. 2 di bawah ini.

Tabel 2. Rumus Evaluasi Peforma Metode

Performance Metrics	Rumus
Akurasi	$\frac{TP+TN}{TP+FP+FN+TN}$
Presisi	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$

3. Metode Penelitian

Dalam bagian hasil dan pembahasan, kami akan menjelaskan secara rinci mengenai proses-proses yang telah dilakukan untuk membandingkan parameter *Cross Validation* pada model *Gradient boosting*. Proses yang dilaksanakan mencakup beberapa langkah berikut:

3.1 Text Mining

Tahap awal meninputkan dataset https://www.kaggle.com/datasets?search=komentar+cyberbullying pada aplikasi aplikasi *Orange 3 Vesri 3.37.0* Data ini terdiri dari 650 komentar yang berlabel positif dan negatif, diantaranya 325 komentar bullying dan 325 komentar non bullying. Selanjutnya ke *Text Mining* yang terdiri dari *Corpus* menyediakan data yang dibutuhkan untuk melatih model pembelajaran mesin, terutama dalam tugas-tugas seperti klasifikasi teks, analisis sentimen, penerjemahan mesin, dan chatbot. Berikutnya teknik *pre-processing* yang sesuai juga dapat meningkatkan performa model *classifier*-nya[10] meliputi proses Transformation, *Tokenizing, Filtering, Normalization.*

Pada tahap transformation, dilakukan perubahan mengubah huruf kapital menjadi huruf kecil[11]. Pada tahap *Tokenizing*, dilakukan pola *Regexp* yang digunakan untuk mencocokkan teks dalam *string* seperti memecah kalimat menjadi kata-kata berdasarkan spasi atau tanda baca. *Filtering*, dilakukan *Stopwords* proses menghilangkan kata-kata yang tidak memiliki dampak penting terhadap performa. *Regexp* adalah alat yang sangat berguna dalam proses *filtering* atau penyaringan data teks. Dalam konteks *filtering*, *RegExp* digunakan untuk mengidentifikasi, memilih, atau mengecualikan teks yang sesuai dengan pola tertentu.

Pada tahap *Normalization* dilakukan *Porter Stemmer* adalah proses mengubah kata ke bentuk dasarnya (*stem/root form*) dengan menghapus akhiran atau *prefiks* yang umum.. Berikutnya *BoW* mengubah teks menjadi representasi berbasis vektor yang menggambarkan frekuensi kemunculan kata-kata dalam sebuah dokumen tanpa memperhatikan urutan kata.

3.2 Analisis Sentimen

Proses analisis ini melibatkan penggunaan algoritma analisis sentimen multilingual sentiment untuk mengklasifikasikan pernyataan atau komentar ke dalam kategori Non Bullying atau Bullying. Analisis sentimen dapat dilakukan secara lebih holistik dan dapat diterapkan pada beragam konten teks dari berbagai sumber dan bahasa[12].

3.3 Evaluasi

Pada tahap evaluasi dataset yang telah diinputkan memasuki proses klasifikasi menggunakan model *gradient boosting* dan menggunakan *matrix* evaluasi seperti *accuracy, precision, recall, f1-score*, dan parameter *Cross Validation* yang berbeda dimulai dari lipatan 2, 3, 5,10 dan 20. Secara garis besar evaluasi dapat terlihat pada tabel 3, 4, 5, 6, dan 7.

Tabel 3 Perbandingan *Cross Validation* 2 pada *Text and Score*

Cross Validation 2

Scores						
Model	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Gradient Boosting	0,880	0,798	0,798	0,804	0,798	0,602

Pada proses validasi silang dengan 2 lipatan yang terlihat pada tabel 3, bahwa model *Gradient Boosting* menunjukkan performa yang mengesankan dengan nilai *AUC* mencapai 0,880. Selain itu, model ini juga mencatat akurasi sebesar 0,798, *F1-Score* 0,798, presisi 0,804, *recall* 0,798, dan *MCC* 0,602. Hasil ini mencerminkan kemampuan model yang cukup baik dalam membedakan antara kelas *bullying* dan *non-bullying*.

Tabel 4 Perbandingan *Cross Validation* 3 pada *Text and Score*

Cross Validation 3

Scores						
Model	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Gradient Boosting	0,900	0,809	0,808	0,814	0,809	0,624

Pada proses validasi silang dengan 3 lipatan yang terlihat pada tabel 4, bahwa model *Gradient Boosting* menunjukkan hasil yang lebih unggul. Nilai *AUC* yang diperoleh mencapai 0,900, sementara akurasi tercatat sebesar 0,809. Selain itu, *F1-Score* mencapai 0,808, presisi berada di angka 0,814, *recall* mencapai 0,809, dan nilai *MCC* adalah 0,624. Semua indikator ini menunjukkan peningkatan kemampuan model dalam mengklasifikasikan antara *bullying* dan *non-bullying*.

Tabel 5 Perbandingan *Cross Validation* 5 pada *Text and Score*

Cross Validation 5

Scores						
Model	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Gradient Boosting	0,901	0,806	0,805	0,811	0,806	0,617

Pada proses validasi silang dengan 5 lipatan yang terlihat pada tabel 5, bahwa model *Gradient Boosting* mencatatkan performa dengan nilai *AUC* sebesar 0,901, akurasi 0,806, *F1-Score* 0,805, presisi 0,811, *recall* 0,806, dan MCC 0,617, yang menunjukkan kemampuan klasifikasi yang cukup baik dan konsisten.

Tabel 6 Perbandingan *Cross Validation* 10 pada *Text and Score*

Cross Validation 10

Scores						
Model	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Gradient Boosting	0,903	0,818	0,818	0,822	0,818	0,641

Pada proses validasi silang dengan 10 lipatan yang terlihat pada tabel 6, bahwa model *Gradient Boosting* menunjukkan performa terbaiknya. Model ini mencatatkan nilai *AUC* sebesar 0,903, akurasi mencapai 0,818, *F1-Score* 0,818, presisi 0,822, *recall* 0,818, dan nilai *MCC* sebesar 0,641. Hasil ini mencerminkan keseimbangan dan efektivitas klasifikasi yang optimal.

Tabel 7 Perbandingan *Cross Validation* 20 pada *Text and Score*

Cross Validation 20

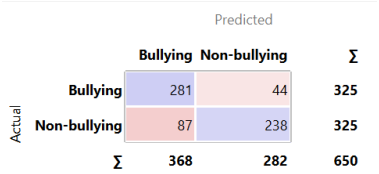
Scores						
Model	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Gradient Boosting	0,902	0,815	0,815	0,819	0,815	0,635

Pada proses validasi silang dengan 20 lipatan yang terlihat pada tabel 7, bahwa model *Gradient Boosting* menunjukkan performa yang sangat baik dengan nilai *AUC* sebesar 0,902, akurasi 0,815, *F1-Score* 0,815, presisi 0,819, *recall* 0,815, dan *MCC* 0,635, mencerminkan stabilitas performa klasifikasi meskipun sedikit di bawah hasil pada 10 lipatan.

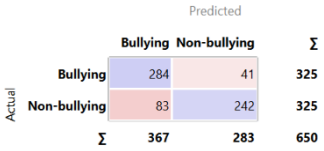
3.4 Perhitungan Confusion Matrix

Matriks kebingungan memberikan pemahaman yang lebih mendalam tentang kekuatan dan kelemahan model dalam memprediksi setiap kelas. Dengan menggabungkan berbagai metrik yang

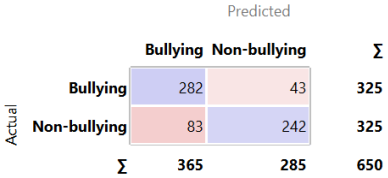
dihasilkan, seperti presisi, *recall*, *F1-score*, dan *MCC*, kita dapat mengevaluasi performa model secara komprehensif, termasuk *trade-off* antara jenis kesalahan yang terjadi (*false positive* dan *false negative*).



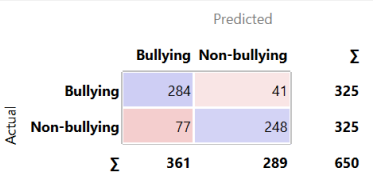
Gambar 2 *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 2



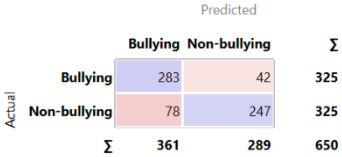
Gambar 3 *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 3



Gambar 4 *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 5



Gambar 5 *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 10



Gambar 6 *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 20

Pada Gambar 2 merupakan *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 2 Hasil ini menunjukkan bahwa model memiliki kinerja yang cukup baik, terutama dalam mendeteksi komentar Bullying (Recall = 86.5%), meskipun ada kesalahan prediksi untuk komentar Non-Bullying (FP = 87).

Sedangkan Gambar 3 merupakan *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 3, komentar Bullying (Recall = 87,4%) dan memiliki tingkat kesalahan prediksi yang relatif kecil untuk komentar Bullying (FN = 41). Namun, model masih menghasilkan beberapa kesalahan dalam memprediksi Non-Bullying sebagai Bullying (FP = 83), yang menunjukkan perlu adanya optimasi lebih lanjut untuk mengurangi false positive.

Kemudian Gambar 4 merupakan *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 5, menunjukkan performa yang cukup baik dalam mendeteksi komentar Bullying (Recall = 86,8%) dan memiliki tingkat akurasi keseluruhan sebesar 80,5%. Namun, model menghasilkan False Positive (FP) sebanyak 83, menunjukkan perlunya optimasi lebih lanjut untuk mengurangi kesalahan dalam memprediksi komentar Non-Bullying sebagai Bullying.

Berikutnya Gambar 5 merupakan *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 10, untuk tingkat akurasi yang cukup baik sebesar 81,8%. Model juga mampu mendeteksi komentar Bullying dengan baik (Recall = 87,4%) namun menghasilkan False Positive sebanyak 77, yang berarti ada data Non-Bullying yang salah diklasifikasikan sebagai Bullying. Precision untuk Bullying adalah 78,7%, yang menunjukkan ruang untuk perbaikan dalam mengurangi prediksi yang salah.

Dan Gambar 6 merupakan *Confusion Matrix* pada hasil prediksi *Gradient Boosting Cross Validation* 20, secara keseluruhan, model ini menunjukkan kinerja yang cukup baik dalam mengklasifikasikan kasus Bullying. Namun, masih ada ruang untuk perbaikan, terutama dalam mengurangi jumlah False Positive (kasus Non-Bullying yang salah diprediksi sebagai Bullying) dan False Negative (kasus Bullying yang tidak terdeteksi).

4. Kesimpulan

Model *Gradient Boosting* menunjukkan performa terbaik pada 10-*fold Cross Validation*, dengan kombinasi nilai evaluasi tertinggi yaitu *AUC* = 0,903, *CA* = 0,818, *F1* = 0,818, *Precision* = 0,822, *Recall* = 0,818, dan *MCC* = 0,641. Ketika jumlah fold ditingkatkan menjadi 20-*fold*, terjadi penurunan sedikit dalam kinerja, yang mengindikasikan bahwa 10-*fold Cross Validation* merupakan pilihan optimal untuk menjaga keseimbangan antara efisiensi dan efektivitas model.

Hasil ini menunjukkan bahwa model *Gradient Boosting* sangat menjanjikan untuk tugas klasifikasi *Bullying*. Validasi silang dengan 10 lipatan *consistently* menghasilkan hasil terbaik di berbagai metrik. Oleh karena itu, model ini bisa diandalkan untuk mengidentifikasi kasus *Bullying* dengan akurasi yang baik, meskipun masih ada kemungkinan kesalahan klasifikasi yang perlu diperhatikan.

Daftar Pustaka

[1] I. P. Ramayasa, I. G. Ayu, D. Saryanti, and I. K. Dharmendra, "Perbandingan Metode Vektorisasi Pada Analisa Sentiment, Studi Kasus : Cyberbullying Pada Komentar Instagram," *J. Teknol. Inf. dan Komput.*, vol. 9, pp. 505–512, 2023.

[2] N. Hardi, Y. Alkahfi, P. Handayani, W. Gata, and M. R. Firdaus, "Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier," *Sistemasi*, vol. 10, no. 1, p. 131, 2021, doi: 10.32520/stmsi.v10i1.1118.

[3] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[4] Y. Setiawan, N. U. Maulidevi, and K. Surendro, "Deteksi Cyberbullying dengan Mesin Pembelajaran Klasifikasi (Supervised Learning): Peluang dan Tantangan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, p. 1577, 2022, doi: 10.25126/jtiik.2022976747.

[5] P. Widiandana, Imam Riadi, and Sunardi, "Implementasi Metode Jaccard pada Analisis Investigasi Cyberbullying WhatsApp Messenger Menggunakan Kerangka Kerja National Institute of Standards and Technology," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, 2020, doi: 10.29207/resti.v4i6.2635.

[6] N. Lestari, Tursina, and E. E. Pratama, "Penerapan Seleksi Fitur Particle Swarm," *Penerapan Sel. Fitur Part. Swarm Optim. pada Klasifikasi Teks (Studi Kasus Komentar Cyberbullying Instagram)*, vol. 9, no. 2, pp. 323–330, 2023.

[7] N. Fajriyani, E. E. Pratama, and R. Septiriana, "Optimasi Hyperparameter pada Neural Network (Studi Kasus: Identifikasi Komentar Cyberbullying Instagram)," *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 2, p. 339, 2023, doi: 10.26418/jp.v9i2.68319.

[8] V. Ayumi, D. Ramayanti, H. Noprisson, Y. Jumaryadi, and U. Salamah, "Model Extreme Gradient Boosting Berbasis Term Frequency (TFXGBoost) Untuk Pengolahan Laporan Pengaduan Masyarakat," *JSAI (Journal Sci. Appl. Informatics)*, vol. 6, no. 1, pp. 65–70, 2023.

[9] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.

[10] N. Hafidz and D. Yanti Liliana, "Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 213–219, 2021, doi: 10.29207/resti.v5i2.2960.

[11] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.

[12] Ipan Hasmadi, Rudiman Rudiman, Khoirul Huda Dwi Putra, and Muhammad Farhat jundullah, "Analisis Sentimen Terhadap Kualitas Layanan Driver Gojek Di Aplikasi Play Store Menggunakan Algoritma Naïve Bayes Dan Aplikasi Orange," *SABER J. Tek. Inform. Sains dan Ilmu Komun.*, vol. 2, no. 1, pp. 138–151, 2023, doi: 10.59841/saber.v2i1.673.