

# Sistem Identifikasi Cyberbullying dan Analisis Kualitas Konten Youtube menggunakan Single-Layer Neural Network

Juan Peter Timothy Yuune<sup>1</sup>, Semmy Wellem Taju<sup>2</sup>

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Klabat  
Manado, Indonesia

e-mail: [1juanyuune@gmail.com](mailto:juanyuune@gmail.com), [2semmy@unklab.ac.id](mailto:semmy@unklab.ac.id)

## Abstrak

*Cyberbullying merupakan tindakan penyalahgunaan internet yang mengancam, mengejek dan mempermalukan orang lain. Cyberbullying sering dilakukan melalui media sosial, salah satunya adalah Youtube. Tindakan bullying yang dilakukan dalam media sosial Youtube berupa komentar-komentar yang ada di kolom komentar. Dalam sebuah konten video memiliki komentar yang sangat banyak maka dari itu untuk melakukan identifikasi komentar bullying dan non-bullying tidaklah mudah jika dilakukan secara manual. Oleh sebab itu, penelitian ini mengembangkan sebuah sistem yang dapat mengidentifikasi komentar bullying secara otomatis dan melakukan analisis terhadap kualitas dari konten tersebut. Penilaian sebuah konten dapat ditinjau dari presentasi jumlah komentar bullying dan non-bullying. Dalam proses klasifikasi, peneliti menggunakan metode Single-Layer Neural Network. Metode ini hanya terdiri dari satu hidden layer yang memungkinkan proses klasifikasi tidak memakan waktu yang lama. Performa model classifier yang dikembangkan menghasilkan performa accuracy 84%, sensitivity 90.5% dan specificity 77.6%. Hasil dari identifikasi komentar ditampilkan dalam bentuk pie chart, word-cloud dan tabel yang menunjukkan probability setiap komentar.*

**Kata Kunci:** Cyberbullying, Machine Learning, Sistem, Youtube.

## Abstract

*Cyberbullying is a form of abuse online where other people are threatened, ridiculed, and humiliated. Cyberbullying happens on social media YouTube. Bullying on YouTube social media comes in the form of comments in the comments section. Video content in social media Youtube has a lot of comments, therefore identifying comments on bullying and non-bullying can be challenging if done manually. Therefore, this research develops a system that can automatically identify comments bullying or non-bullying and analyze the content's quality. Assessment of a content can be reviewed from the presentation of the number of comments on bullying and non-bullying. In the classification process, the researcher uses the method Single-Layer Neural Network. This method contains only one hidden layer, so the classification process does not take much time. The performance of the classifier model gave an accuracy of 84%, a sensitivity of 90.5%, and a specificity of 77.6%. The results of the identification of comments are displayed in the form of a pie chart, word-cloud and a table showing the probability of each comment.*

**Keywords:** Cyberbullying, Machine Learning, System, Youtube.

## 1. Pendahuluan

Cyberbullying merupakan aktivitas mengancam, mempermalukan dan mengejek orang lain melalui internet [1]. Dalam melakukan identifikasi komentar *bullying* dengan jumlah yang sedikit pada sebuah konten di *platform* tertentu memang cukup mudah. Tetapi jika terdapat ribuan komentar pada konten tersebut maka bukanlah hal yang mudah untuk mengidentifikasi setiap komentar dan memakan waktu yang cukup lama[2]. Oleh karena itu, diperlukan sebuah sistem yang secara otomatis dapat membantu dalam mengidentifikasi setiap komentar *bullying* dari ribuan komentar. Ribuan komentar tersebut menjadi *dataset* dalam penelitian ini dan dari data yang ada dilakukan *training* dan *testing data*.

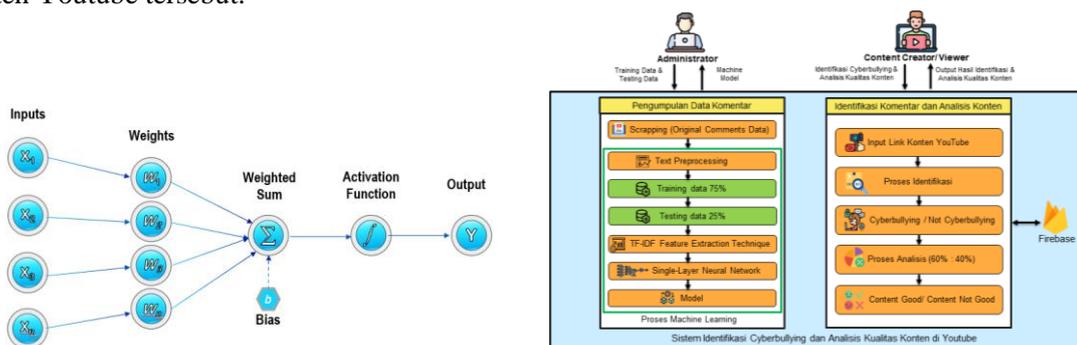
Sistem dirancang bukan hanya untuk dapat melakukan identifikasi saja, tetapi sistem juga dapat melakukan analisis kualitas konten di media sosial Youtube dengan memanfaatkan *dataset* yang ada dalam hal ini komentar-komentar dari konten tersebut. Untuk mengembangkan sebuah sistem dan proses analisis kualitas dari sebuah konten tidak lepas dari penggunaan pendekatan atau metode yang tepat. Penelitian ini menggunakan salah satu model yang ada dalam *Software Development Life Cycle (SDLC)* yaitu model *Prototype* [3]. Sedangkan *Single-Layer Neural Network* menjadi pilihan untuk mengklasifikasi komentar di media sosial Youtube.

*Single-Layer Neural Network (SLNN)* merupakan salah satu jenis pemodelan jaringan pada *Artificial Neural Network (ANN)* yang termasuk ke dalam jenis jaringan *feedforward* [4]. SLNN memiliki cara kerja yang lebih sederhana dibandingkan dengan yang lain yaitu hanya memiliki satu *layer* pada bagian *hidden layer*. Setiap komentar merupakan input layer dan komentar yang diidentifikasi diklasifikasikan menjadi dua kelompok yaitu komentar yang mengandung unsur negatif *bullying* dan yang mengandung unsur positif atau bukan *bullying*. Kedua hasil klasifikasi merupakan output yang dihasilkan.

Penelitian sebelumnya oleh [2] menggunakan metode *Naïve Bayes Classifier* yang digunakan untuk mendeteksi dan menghapus komentar *bullying* yang ada pada media sosial Instagram. Ada dua tingkat klasifikasi terhadap komentar postingan sosial media Instagram. Klasifikasi pertama, aplikasi mengklasifikasikan kriteria sentimen positif dan negatif. Jika klasifikasi pertama menunjukkan komentar adalah sentimen positif, maka tidak dilanjutkan untuk ke tingkat selanjutnya. Jika hasil klasifikasi komentar adalah sentimen negatif, maka dilanjutkan ke klasifikasi kedua yang merupakan klasifikasi kelas *bullying* dan bukan *bullying* [2]. Penelitian ini juga melakukan pengembangan aplikasi berbasis *web* yang melakukan penghapusan komentar yang mengandung *bullying* secara massal. Hasil performa model yang memiliki nilai akurasi sebesar 98,5%, dengan nilai presisi sebesar 96%, nilai *recall* sebesar 96% dan nilai kerusakan atau biasa disebut sebagai *error rate* sebesar 1,5%.

Penelitian lain yang telah dilakukan sebelumnya oleh [5] berfokus pada peningkatan kualitas konten oleh seorang *content creator*. Penelitian ini melakukan pengembangan aplikasi yang hanya berfokus pada pengelompokan setiap komentar yang ada. Pengelompokan tersebut diantaranya *class* sentimen positif, sentimen negatif, dan data bukan bahasa Indonesia. Data testing yang diproses pada model didapat dari API Youtube. Proses klasifikasi *Naïve Bayes* dan perhitungan *Confusion Matrix* menggunakan *Library Varunon* [5]. Hasil akurasi tertinggi didapati pada proses *stemming* dengan nilai akurasi 90,4%, nilai presisi 77,35, nilai *recall* 82,3% dan nilai F-score 79,7%.

Berdasarkan penelitian terakit diatas, peneliti mengembangkan sistem yang menggabungkan kedua ide dengan menambahkan beberapa fitur ke dalam sistem. Peneliti juga melakukan pembuatan *dataset* dengan melakukan pengambilan data komentar dari Youtube, melakukan *training data* dan *testing data*, membuat model *machine learning* dengan algoritma SLNN dan melakukan pengembangan sistem prediksi berbasis *web*. Adapun tujuan penelitian ini adalah untuk menghasilkan sebuah sistem yang dapat melakukan identifikasi komentar *bullying* dalam sebuah konten video di media sosial Youtube dan juga menghasilkan sistem yang dapat menganalisis kualitas konten secara otomatis melalui komentar yang telah diidentifikasi pada konten Youtube tersebut.



**Gambar 1.** *Single-Layer Neural Network***Gambar 2.** Kerangka Konseptual Sistem

## 2. Metode Penelitian

Ruang lingkup dari penelitian ini adalah sistem yang dikembangkan berbasis *web*, dapat mendukung identifikasi komentar yang memiliki *keywords bullying* Bahasa Indonesia, Bahasa lokal (Manado dan Jawa) dan Bahasa *Slang*. Sistem juga secara otomatis melakukan analisis kualitas konten dari hasil presentasi identifikasi komentar *bullying/non-bullying*. Pada dasarnya penelitian ini tidak membahas tentang bagaimana sistem melakukan pencegahan terhadap Tindakan *cyberbullying*. Sistem ini tidak mendukung dalam identifikasi komentar *bullying* yang menggunakan Bahasa Inggris dan komentar yang mengandung sarkasme.

Pengklasifikasian data dalam penelitian ini menggunakan metode *Single-Layer Neural Network*. Dataset yang digunakan diambil dari komentar-komentar yang ada dalam sebuah konten di media sosial Youtube. Hasil performa model untuk algoritma haruslah mencapai ambang batas pengambilan keputusan (*decision threshold*) sebesar 75%. Sistem ini memiliki menu *login* bagi *user* dan *admin*. *User* dapat melakukan identifikasi dan *admin* dapat mengelola data akun *user*.

Gambar 1 dan Gambar 2 secara berurutan adalah gambar dari *Single-Layer Neural Network* [6] dan kerangka konseptual dari sistem yang dikembangkan. Pengembangan sistem dalam penelitian ini menggunakan metode *System Development Life Cycle* (SLDC). Peneliti menggunakan model *Prototype* untuk membantu dalam pengembangan web untuk sistem yang dibuat oleh peneliti [7]. Berikut ini adalah tahapan-tahapan yang dilakukan oleh peneliti: a) *Requirements Gathering*, b) *Quick Design*, c) *Build Prototype*, d) *User Evaluation*, e) *Refining Prototype* dan f) *Implement and Maintain*.

## 3. Hasil dan Pembahasan

### 3.1 Data Collection

Data komentar yang ada di media sosial Youtube dikumpulkan dengan menggunakan teknik *scraping*. Peneliti mengunduh sejumlah komentar dari Youtube dengan menginstall *library package youtube-comment-downloader*. Setelah itu melakukan *scraping code* pada *anaconda prompt shell* menggunakan Bahasa pemrograman *python* dengan menambahkan *link* dari konten video di Youtube. Jumlah data komentar yang digunakan sebanyak 8000 komentar yang terdiri dari 4000 komentar *bullying* dan 4000 komentar *non-bullying*.

### 3.2 Text Preprocessing

Data komentar yang ada diproses dengan beberapa tahapan mulai dari *drop missing values*, *remove number*, *change comments to lowercase*, *remove punctuations*, *remove repeated characters*, *remove emoji*, dan sampai pada *filtering* (*stop-word removal*).

### 3.3 Split Data

Dari keseluruhan data komentar yang digunakan dibagi ke dalam *data training* dan *data testing* dengan perbandingan 75:25. *Data training* 6000 dan *testing* 2000 dimana keduanya terdiri dari klasifikasi *bullying* dan *non-bullying*.

### 3.4 Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF mengolah seluruh data berupa komentar yang telah diuji sebelumnya untuk diubah menjadi data numerik dengan bentuk vector [8]. *N-gram range* yang digunakan dalam menganalisa kata adalah 2,2. Jadi pembobotan kata dilakukan disetiap dua kata.

### 3.5 Pembentukan Model Klasifikasi

Pada pembentukan model klasifikasi, peneliti menerapkan algoritma *machine learning* yaitu *single-layer neural network*. Algoritma ini memberikan *voting* terhadap komentar sesuai dengan *label* yang telah ditentukan. Angka mewakili *label bullying* adalah 1 dan *label non-bullying* adalah 0.

Pengukuran performa model yang dilakukan menggunakan empat pengukuran (*metric*) yaitu *sensitivity*, *specificity*, *accuracy* dan *Matthews Correlation Coefficient* (MCC) Berikut adalah perhitungan dari masing-masing *metric*.

1. *Accuracy* akan mengukur seberapa akurat model dalam melakukan klasifikasi dengan benar dari keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% = \frac{905 + 776}{905 + 224 + 776 + 95} \times 100\% = 84\%$$

2. *Sensitivity* akan mengukur seberapa baik model melakukan klasifikasi dari kelas *bullying* dari keseluruhan data yang benar *bullying*.

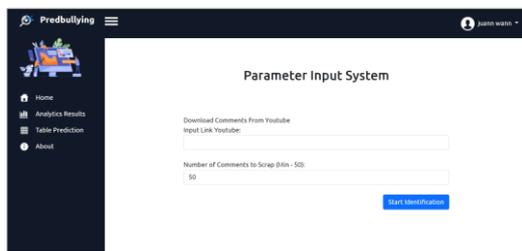
$$Sensitivity = \frac{TP}{TP + FN} \times 100\% = \frac{905}{905 + 95} \times 100\% = 90.5\%$$

3. *Specificity* akan mengukur seberapa baik model melakukan klasifikasi dari kelas *non-bullying* dari keseluruhan data yang benar *non-bullying*.

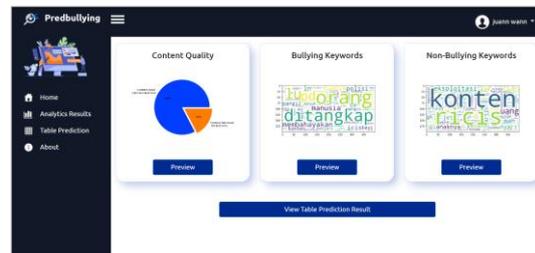
$$Specificity = \frac{TN}{TN + FP} \times 100\% = \frac{776}{776 + 224} \times 100\% = 77.6\%$$

4. *MCC* akan mengukur seberapa baik model melakukan klasifikasi dari kedua kelas.

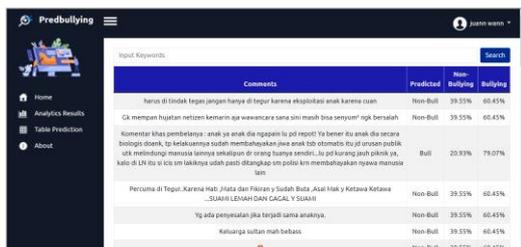
$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} = \frac{(905 \times 776) - (224 \times 95)}{\sqrt{(905 + 224)(905 + 95)(776 + 224)(776 + 95)}} = 0.687$$



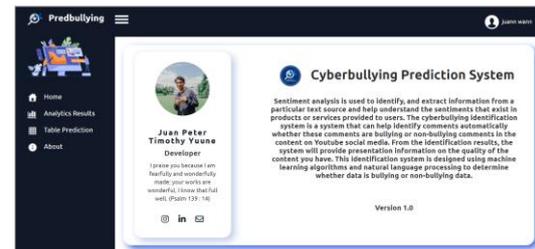
Gambar 3. Halaman Home User



Gambar 4. Halaman Analytics Results



Gambar 5. Halaman Table Prediction



Gambar 6. Halaman About

### 3.7 Analisis Sistem

Pada pengembangan sistem yang dilakukan berorientasi objek. Maka dari itu, peneliti menggunakan *Unified Modelling Language* (UML) [9]. *Use case diagram* menunjukkan interaksi antara *user* terhadap sistem dan juga admin terhadap sistem. Diagram ini juga menunjukkan setiap fitur yang dapat diakses atau dilakukan oleh sistem terhadap *user* [10].

### 3.8 Implementasi Antarmuka

Setelah melakukan tahapan perancangan antarmuka dari sistem yang dikembangkan, pada bagian ini menunjukkan implementasi dari setiap perancangan antarmuka yang telah dilakukan sebelumnya. Implementasi dari user interface dapat dilihat pada Gambar 3, Gambar 4, Gambar 5 dan Gambar 6.

Tabel 1. Pengujian Sistem dengan Link Konten Youtube

Input Parameter (Link, Jumlah Komentar)	Hasil Identifikasi	Hasil Analisis
<a href="https://www.youtube.com/watch?v=XYEDlh_JbbM">https://www.youtube.com/watch?v=XYEDlh_JbbM</a> , 900 komentar	Presentasi komentar <i>bullying</i> 6% (50 komentar) dan <i>non-bullying</i> 94% (849 komentar)	<i>Content Good</i>

<a href="https://www.youtube.com/watch?v=STGvnZZVjK4">https://www.youtube.com/watch?v=STGvnZZVjK4</a> , 1000 komentar	Presentasi komentar <i>bullying</i> 10% (101 komentar) dan <i>non-bullying</i> 90% (898 komentar)	Content Good
<a href="https://www.youtube.com/watch?v=F0dUsBmFvjI">https://www.youtube.com/watch?v=F0dUsBmFvjI</a> , 1200 komentar	Presentasi komentar <i>bullying</i> 9% (105 komentar) dan <i>non-bullying</i> 91% (1094 komentar)	Content Good
<a href="https://www.youtube.com/watch?v=Yze7F317zZY">https://www.youtube.com/watch?v=Yze7F317zZY</a> , 1600 komentar	Presentasi komentar <i>bullying</i> 9% (141 komentar) dan <i>non-bullying</i> 91% (1458 komentar)	Content Good
<a href="https://www.youtube.com/watch?v=of6NqyvAG2Q">https://www.youtube.com/watch?v=of6NqyvAG2Q</a> , 2000 komentar	Presentasi komentar <i>bullying</i> 12% (240 komentar) dan <i>non-bullying</i> 88% (1759 komentar)	Content Good

### 3.9 Implementasi Database

Seperti yang telah disebutkan sebelumnya bahwa pengembangan sistem ini memanfaatkan *Firestore* sebagai *real-time database*. Database digunakan untuk menyimpan data autentikasi *login* dari *user* berupa *email*, *first name*, *last name* dan *password*.

### 3.10 Pengujian

Peneliti juga melakukan pengujian dengan melakukan identifikasi sebanyak 15 kali (5 link yang dipilih dan hasil pengujian ditampilkan pada Tabel 1) dengan Youtube link dari konten yang berbeda-beda. Bukan hanya itu, jumlah komentar yang diidentifikasi juga bervariasi.

### 3.11 Pembahasan

Berdasarkan rumusan masalah dalam penelitian ini tentang bagaimana sistem dapat mengidentifikasi adanya komentar *bullying* dalam suatu konten dan bagaimana sistem secara otomatis dapat mengukur kualitas konten melalui komentar dari konten tersebut di media sosial Youtube telah dikembangkan. Peneliti dapat menyimpulkan bahwa sistem yang dikembangkan dapat membantu dalam melakukan identifikasi komentar *bullying* dan *non-bullying* bagi para pengguna Youtube baik sebagai *content creator* ataupun *viewer*. Pada penggunaan sistem ini juga dapat membantu mengetahui kualitas konten dari sebuah video di media sosial Youtube berdasarkan komentar-komentar klasifikasi dari setiap komentar yang ada.

Tabel 2. Parameter Algoritma *Machine Learning*

Algoritma	Parameter
<i>k-NN</i>	<i>n_neighbors=5, leaf_size=30, metric_params=None</i>
<i>Random Forest</i>	<i>n_estimators=100, max_samples=None, min_samples_split=2</i>
<i>Gaussian Naïve Bayes</i>	<i>GaussianNB = ()</i>
<i>SLNN</i>	<i>max_iter = 10, eta0 = 0.3, random_state = 0</i>

Hasil identifikasi dan klasifikasi yang diberikan mudah dipahami oleh pengguna sistem. Hal tersebut dikarenakan visualisasi hasil identifikasi berupa *pie chart*, *word-cloud*, dan tabel yang berisi *probability bullying* dan *non-bullying* dari setiap komentar. Dimana *admin* dapat melihat daftar pengguna sistem, melakukan penambahan pengguna baru, mengedit atau melakukan perubahan data pengguna, dan menghapus akun pengguna.

Tabel 3. Perbandingan Performa Model

Model	Sensitivity	Specificity	Accuracy	MCC
<i>k-NN</i>	35.3%	95.8%	65.6%	0.391
<i>Gaussian Naïve Bayes</i>	61.4%	93.8%	77.6%	0.583
<i>Random Forest</i>	99%	65.2%	82.1%	0.682
<i>SLNN</i>	90.5%	77.6%	84%	0.687

Tabel 2 menunjukkan penggunaan parameter dari setiap algoritma sebagai perbandingan pada penelitian ini. Pada penggunaan parameter kali ini, peneliti menggunakan *default parameter* terhadap seluruh algoritma selain algoritma *SLNN*. Hal ini dilakukan untuk dapat meningkatkan akurasi model yang ada. Dari hasil penggunaan algoritma dengan parameter yang ada maka peneliti mendapatkan performa model sebagai berikut.

Dari Tabel 3 bisa dilihat bahwa model *SLNN* memiliki hasil yang baik dibandingkan dengan ketiga algoritma yang lain. Maka dari itu, algoritma *SLNN* dapat digunakan dalam proses identifikasi *cyberbullying*. Hasil pada tabel menunjukkan bahwa *sensitivity*, *specificity*, *accuracy*, dan *MCC* mencapai 90.5%, 77.6%, 84%, dan 68.7%.

#### 4. Kesimpulan

Sistem berhasil dikembangkan dengan menggunakan metode *single-layer neural network* dalam mengklasifikasi setiap komentar apakah termasuk dalam kategori *bullying* atau *non-bullying*. Hasil identifikasi ditampilkan dalam bentuk *pie chart* serta tabel yang menunjukkan *probability* dari setiap komentar. Sistem dapat melakukan analisis kualitas konten berdasarkan komentar yang identifikasi oleh *user*. Hasil analisis juga ditampilkan bersamaan dengan hasil identifikasi dalam bentuk *pie chart*. Sistem yang dikembangkan telah memenuhi cakupan yang ada dalam penelitian ini. Dimana setiap fitur dan fungsi dalam sistem telah dilakukan pengujian dengan menggunakan metode *black box* sehingga dapat dilihat bahwa fungsionalitas dari sistem dapat berjalan dengan baik dan memenuhi kebutuhan pengguna. Hasil akurasi *model machine learning* untuk klasifikasi yang telah dikembangkan memiliki akurasi sebesar 84%. Maka dari itu, hasil akurasi yang didapat telah melebihi *threshold* yang telah ditentukan yakni 75%.

Saran yang diberikan yaitu dapat menambahkan fitur sistem bagi admin untuk melakukan *training model* secara otomatis dan peneliti lain dapat melakukan *tuning* untuk mendapatkan parameter yang terbaik sehingga dapat meningkatkan performa model untuk akurasi prediksi yang dilakukan nantinya.

#### Daftar Pustaka

- [1] C. Destitus, W. Wella, and S. Suryasari, "Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia," *Ultima InfoSys : Jurnal Ilmu Sistem Informasi*, vol. 11, no. 2, pp. 107–111, 2020, doi: 10.31937/si.v11i2.1740.
- [2] A. S. Hutagalung, A. B. P. Negara, and E. E. Pratama, "Aplikasi Pendeteksi Cyberbullying Terhadap Komentar Postingan Media Sosial Instagram dengan Metode Naïve Bayes Classifier Berbasis Website," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 9, no. 3, p. 364, 2021, doi: 10.26418/justin.v9i3.44843.
- [3] T. Pricillia and Zufachmi, "Perbandingan Metode Pengembangan Perangkat Lunak (Waterfall, Prototype, RAD)," *Jurnal Bangkit Indonesia*, vol. 10, no. 1, pp. 6–12, 2021, doi: 10.52771/bangkitindonesia.v10i1.153.
- [4] M. Xie and U. Roshan, "Exploring Classification, Clustering, and Its Limits in a Compressed Hidden Space of a Single Layer Neural Network with Random Weights," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11506 LNCS, no. May, pp. 507–516, 2019, doi: 10.1007/978-3-030-20521-8\_42.
- [5] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8.1, pp. 255–259, 2020.
- [6] S. W. Taju, E. Y. Putra, and G. F. Mandias, "Sentiment Identification System for E-Commerce Mobile App Reviews Using Single Layer Neural Network," *2022 4th International Conference on Cybernetics and Intelligent System, ICORIS 2022*, 2022, doi: 10.1109/ICORIS56080.2022.10031580.
- [7] F. Supandi, W. Desta P, Y. Ambar S, and M. Sudir, "Analisis Resiko Pada Pengembangan Perangkat Lunak Yang Menggunakan Metode Waterfall Dan Prototyping," *Prosiding Seminar Nasional Dinamika Informatika 2018 (SENADI 2018)*, vol. 2, no. 1, pp. 83–86, 2019, [Online]. Available: <http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/86>.
- [8] R. A. Sasmita and A. Z. Falani, "Pemanfaatan Algoritma TF/IDF Pada Sistem Informasi Ecomplaint Handling," *Jurnal Link*, vol. 27, no. 1, pp. 27–33, 2018.
- [9] R. Leonardo, I. Arwani, and D. E. Ratnawati, "Pemanfaatan Teknologi Firebase dalam Pengembangan Aplikasi Pengelolaan Stok Barang Berbasis Mobile pada Rumah Makan Nakamse Malang," *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, vol. 1, no. 1, pp. 1–11, 2020, doi: 10.25126/justsi.v1i1.1.
- [10] R. K. Ngantung and M. A. I. Pakereng, "Model Pengembangan Sistem Informasi Akademik Berbasis User Centered Design Menerapkan Framework Flask Python," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 1052, 2021, doi: 10.30865/mib.v5i3.3054.