UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Ria

COMPARISON OF MACHINE LEARNING ALGORITHMS ON SENTIMENT ANALYSIS OF ELSAGATE CONTENT

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi

Oleh:

ZAIRA CINDYA DWYNNE
12050324080

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024

UIN SUSKA RIAU

LEMBAR PERSETUJUAN

COMPARISON OF MACHINE LEARNING ALGORITHMS ON SENTIMENT ANALYSIS OF ELSAGATE CONTENT

TUGAS AKHIR

Oleh:

ZAIRA CINDYA DWYNNE
12050324080

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 3 Juni 2024

Ketua Program Studi	Pembimbing
Eki Saputra, S.Kom., M.Kom.	**Mustakim, ST., M.Kom.**
NIP. 198307162011011008	**NIK. 130511023**

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
asalah.
im Ria

UIN SUSKA RIAU

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

COMPARISON OF MACHINE LEARNING ALGORITHMS ON SENTIMENT ANALYSIS OF ELSAGATE CONTENT

TUGAS AKHIR

Oleh:

ZAIRA CINDYA DWYNNE
12050324080

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 21 Mei 2024

Pekanbaru, 21 Mei 2024
Mengesahkan,

Ketua Program Studi

Dr. Hartono, M.Pd.
NIP. 196403011992031003

Eki Saputra, S.Kom., M.Kom.
NIP. 198307162011011008

DEWAN PENGUJI:

Ketua : **Nesdi Evrilyan Rozanda, S.Kom., M.Sc.**

Sekretaris : **Mustakim, ST., M.Kom.**

Anggota 1 : **Dr. Rice Novita, S.Kom., M.Kom.**

Anggota 2 : **M. Afdal, ST., M.Kom.**

asalah.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

UIN SUSKA RIAU

UIN SUSKA RIAU

Lampiran Surat :
Nomor : Nomor 25/2021
Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Zaira Cindya Dwynne
NIM : 12050324080
Tempat/Tgl. Lahir : Dumai, 8 Mei 2002
Fakultas/~~Pascasarjana~~ : Sains dan Teknologi
Prodi : Sistem Informasi
Judul Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*:
Comparison Of Machine Learning Algorithms On Sentiment Analysis Of Elsagate Content.

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan ~~Disertasi/Thesis~~/Skripsi/~~Karya Ilmiah lainnya~~* dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu ~~Disertasi/Thesis~~/Skripsi/~~Karya Ilmiah~~ lainnya* saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan ~~Disertasi/Thesis~~/Skripsi~~/(Karya Ilmiah lainnya)~~* saya tersebut, maka saya besedia menerima sanksi sesua peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 19 Juni 2024
Yang membuat pernyataan

METERAI TEMPEL
BDADCALX183849959

.....
NIM : 12050324080

**pilih salah satu sasuai jenis karya tulis*

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
asalah.
im Riau

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Dengan menyebut nama Allah yang maha pengasih lagi maha penyayang

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahi Rabbil 'Alamin, segala puji bagi *Allah Subhanahu Wa Ta'ala* sebagai bentuk rasa syukur atas segala nikmat yang telah diberikan tanpa ada kekurangan sedikitpun. Shalawat beserta salam tak lupa pula kita ucapkan kepada Nabi Muhammad *Shallallahu 'Alaihi Wa Sallam* dengan mengucapkan *Allahumma Sholli'ala Sayyidina Muhammad Wa'ala Ali Sayyidina Muhammad*. Semoga kita semua selalu senantiasa mendapat syafa'at-Nya di dunia maupun di akhirat, *aamiin ya rabbala'alaamiin*. Kupersembahkan hadiah istimewa karya kecil ini sebagai salah satu bentuk bakti, rasa terima kasih, dan hormatku kepadamu orang tuaku tercinta, Ayah dan Ibu.

Ayah dan Ibu tersayang, terima kasih untuk setiap perjuangan yang di usahakan, selalu mendoakan disetiap sujud panjang, membimbing, dan mendorong saya dalam kebaikan dan selalu ada saat keadaan tersulit sekalipun. Terima kasih untuk segala pengorbanan yang dilakukan. Sampai kapanpun tiada rasa, tiada cara yang dapat membalas semua yang telah kalian lakukan. Untuk itu saya anakmu ini selalu mendoakan yang terbaik agar Ayah dan Ibu bahagia dunia dan akhirat, serta diberikan tempat istimewa di sisi-Nya kelak sehingga kita bisa berkumpul kembali bersama-sama di Jannah-Nya.

Terima kasih saya ucapkan kepada Bapak Mustakim selaku pembimbing Tugas Akhir saya yang telah membimbing dari awal perkuliahan hingga saat ini dengan setulus hati. Terima kasih saya ucapkan kepada Bapak dan Ibu Dosen Sistem Informasi yang selama ini sudah mewariskan ilmu, motivasi, dan arahan untuk menyelesaikan studi di Program Studi Sistem Informasi ini. Terima kasih juga saya ucapkan kepada kakak-kakak tingkat dan teman-teman saya yang telah membantu dan memberikan semangat dalam menyelesaikan Tugas Akhir ini.

Wassalamu'alaikum Warahmatullahi Wabarakaatuh.

UIN SUSKA RIAU
Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahi Rabbil 'Alamin, bersyukur kehadirat *Allah Subhanahu Wa Ta'ala* atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Shalawat serta salam kita ucapkan kepada Nabi Muhammad *Shallallahu 'Alaihi Wa Sallam* dengan mengucapkan *Allahumma Sholli'Ala Sayyidina Muhammad Wa'Ala Ali Sayyidina Muhammad*. Tugas Akhir ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi Universitas Islam Negeri Sultan Syarif Kasim Riau. Pada penulisan Tugas Akhir ini, terdapat beberapa pihak yang sudah berkontribusi dan mendukung penulis baik berupa materi, moril, dan motivasi. Penulis ingin mengucapkan banyak terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd sebagai Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Eki Saputra, S.Kom., M.Kom sebagai Ketua Program Studi Sistem Informasi, Fakultas Sains dan teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Siti Monalisa, ST., M.Kom sebagai Sekretaris Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Bapak Mustakim, S.T., M.Kom sebagai Dosen pembimbing Proposal dan Tugas Akhir yang telah berkenan membimbing, mengarahkan dan meluangkan waktu, tenaga dan pikiran guna membantu penulisan dalam menyelesaikan Tugas Akhir ini.
6. Bapak Mustafa, S.Sos., M.Ikom sebagai Dosen pembimbing Proposal dan Tugas Akhir yang telah berkenan membimbing, mengarahkan dan meluangkan waktu, tenaga dan pikiran guna membantu penulisan dalam menyelesaikan Tugas Akhir ini.
7. Para Dosen Program Studi Sistem Informasi Fakultas Sains dan Teknologi Sultan Syarif Kasim Riau yang memberikan ilmu yang bermanfaat serta memberikan semangat untuk meraih cita-cita dan masa depan yang cerah.
8. Seluruh Pegawai dan Staf Fakultas Sains dan Teknologi yang telah membantu dan mempermudah proses administrasi selama perkuliahan ini.
9. Keluarga hebat dan membanggakan, keluarga *Puzzle Research Data Tech-*

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

nology (PREDATECH), yang terdiri dari dosen pembimbing terhebat (Pak Mustakim, Pak Inggih, Pak Afdal, Buk Rice dan dosen lainnya) dan seluruh alumni serta anggota aktif PREDATECH yang tidak dapat penulis sebutkan satu persatu.

10. Teristimewa untuk kedua orang tua penulis, yang selalu menjadi motivasi untuk penulis dalam menyelesaikan segala urusan dalam perjalanan karir penulis.
11. Kepada saudara kembar penulis yang selalu membantu dan memberikan motivasi selama perkuliahan hingga penyelesaian Tugas Akhir.
12. Kepada keluarga serta saudara-saudara penulis yang terus memberikan semangat dan motivasi kepada penulis sehingga penulis dapat menyelesaikan Tugas Akhir.
13. Para Dosen Program Studi Sistem Informasi Fakultas Sains dan Teknologi Sultan Syarif Kasim Riau yang memberikan ilmu yang bermanfaat serta memberikan semangat untuk meraih cita-cita dan masa depan yang cerah.
14. Kepada Kakak-kakak senior. Terima kasih atas motivasi dan arahannya.
15. Serta semua pihak yang namanya tidak dapat disebutkan satu persatu, yang telah banyak membantu dalam pelaksanaan penelitian Tugas Akhir dan menyelesaikan Tugas Akhir ini.

Semoga segala doa dan dorongan yang telah diberikan selama ini menjadi amal kebajikan dan mendapat balasan setimpal dari Allah *Subhanahu Wa Ta'ala.* Penulis menyadari bahwa penulisan Tugas Akhir ini masih banyak terdapat kekurangan dan jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangat diharapkan untuk kesempurnaan Tugas Akhir ini dan semoga laporan ini bermanfaat bagi kita semua. Akhir kata penulis ucapkan terima kasih.

Wassalamu'alaikum Warahmatullahi Wabarakaatuh.

Pekanbaru, 3 Juni 2024

Penulis,

UIN SUSKA RIAU

ZAIRA CINDYA DWYNNE

NIM. 12050324080

UIN SUSKA RIAU
Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau State Islamic University of Sultan Syarif Kasim Riau

Comparison of Machine Learning Algorithms on Sentiment Analysis of Elsagate Content

Zaira Cindya Dwynne
Department of Information System
Universitas Islam Negeri Sultan Syarif Kasim Riau
Pekanbaru, Indonesia
12050324080@students.uin-suska.ac.id

Mustakim
Department of Information System
Universitas Islam Negeri Sultan Syarif Kasim Riau
Pekanbaru, Indonesia
mustakim@uin-suska.ac.id

Mustafa
Department of Communication Science
Universitas Islam Negeri Sultan Syarif Kasim Riau
Pekanbaru, Indonesia
mustafa@uin-suska.ac.id

***Abstract*— This study responds to the increasing phenomenon of elsagate content on various platforms, especially on YouTube Kids and YouTube, which are often accessed by children. Elsagate content contains sensitive elements for children such as horror, sexuality, and violence. The community's response to this content varies, from positive to negative, to neutral, especially on platforms like YouTube. The main purpose of this research is to understand the opinions of the YouTube community regarding children's content with unclear meanings or containing elsagate elements. Using 2452 data, this study applies five machine learning algorithms to classify sentiment: Naive Bayes Classifier (NBC), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Logistic Regression (LR). The research results show that data division with a 90:10 ratio provides the best performance. The Support Vector Machine algorithm achieves the highest accuracy of 63%, with precision of 61%, recall of 63%, and an F1-score of 60%. On the other hand, the K-Nearest Neighbors algorithm shows the lowest performance with an accuracy of 56%, precision of 55%, recall of 56%, and an F1-score of 55%. Thus, besides aiming to provide insights into elsagate, this research also highlights the performance of Support Vector Machine in analyzing sentiment towards elsagate content.**

Keywords*—*Elsagate, Naive Bayes Classifier, Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression.

I. INTRODUCTION

The development of technology is becoming increasingly sophisticated and digital. This makes it easier for people to engage in various activities, obtain information, knowledge, and entertainment. One example of commonly used technology is gadgets. According to [1], gadgets are electronic devices used as communication tools by humans. Currently, not only adults but also children, even toddlers, can enjoy gadgets as educational and entertainment tools. During the COVID-19 pandemic, the use of gadgets among children has increased due to the implementation of the Work From Home (WFH) system, which requires children to attend school and learn online [2]. According to the results of a survey by the Indonesian Child Protection Commission (KPAI) in 2020, 71.3% of children already have their own gadgets, and 79% of children are also allowed to use gadgets for purposes other than learning.

The common platform used by parents to provide education and entertainment to children is YouTube. YouTube provides a lot of child-specific content and even has a dedicated platform for children called YouTube Kids. The purpose of YouTube Kids is to serve as an educational tool to help children discover new interests. However, it is not uncommon for children's content available on YouTube and YouTube Kids to contain elements that are not suitable for children. This phenomenon is called "Elsagate," which in recent years has spread to various platforms including YouTube and YouTube Kids. Elsagate is a phenomenon where content intended for children, such as animated videos using popular characters from cartoons and movies, often display inappropriate material for children to watch [3][4].

In the study titled "The Elsagate Phenomenon: Regulation and the Role of the State in Creating Child-Friendly Entertainment in Indonesia (2023)," the Elsagate phenomenon first emerged in 2016 in one of the videos available on YouTube. The study provides a detailed explanation of Elsagate from its inception, impact, and the role of the state. However, the term "Elsagate" itself is still not widely known by the public. Therefore, the purpose of this research is to provide knowledge to readers about Elsagate and to understand the public's perception of children's content that falls under Elsagate and similar content. In this study, the focus is on how the public responds to Elsagate content, whether positively, negatively, or neutrally. Data was collected through YouTube using text mining techniques. Text Mining is the process of data mining that consists of text, usually taken from documents, with the aim of finding words that represent the content of the document. The main goal is to analyze the relationship between documents.

This research applies five machine learning algorithms to classify data, namely Naïve Bayes Classifier (NBC), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Logistic Regression (LR). The application of multiple algorithms in this study aims to improve the accuracy and reliability of the analysis results. It also allows the analysis results to be validated by other algorithms, ensuring that the conclusions drawn are not solely dependent on one analysis method. Each algorithm certainly has unique characteristics that make it more suitable for certain types of data or situations.

The Naïve Bayes Classifier (NBC) algorithm is chosen because it has several advantages, such as being a text classification method with high processing speed and accuracy when applied to cases with large, diverse, and in-depth data. Naive Bayes is recognized as a simple and

effective algorithm for text classification [5]. Previous research on sentiment analysis of Jiniso product reviews by Efta Lam Herny (2022) conducted analysis using Naïve Bayes Classifier and achieved a relatively good accuracy of 94% [6].

Random Forest has a high level of accuracy, resilience to outliers and noise, and better speed compared to bagging and boosting techniques [7]. In a study conducted by Shoffan et al. (2021) on anxiety detection based on social media data by implementing K-NN, Bernoulli, Decision Tree Classifier, Support Vector Classifier, Random Forest, and XG-boost algorithms, the best accuracy is achieved by Random Forest at 84.95%[8].

Support Vector Machine (SVM) is a supervised learning approach that relies on the principle of Structural Risk Minimization (SRM) and is used to find the best hyperplane that can separate two classes. The goal of SVM is to find the optimal hyperplane that can separate two classes as effectively as possible [9]. In a study on sentiment analysis by Márcio Guia et al. (2019) on Amazon unlocked reviews comparing four algorithms Naïve Bayes Classifier, Support Vector Machine, Decision Tree, and Random Forest, the best accuracy was obtained by Support Vector Machine with a value of 89% on a linear kernel [7].

The K-Nearest Neighbor (K-NN) method is a simple approach to classifying objects based on their proximity in the training data. K-NN considers attributes and sample data to classify new data [13]. K-NN is particularly effective in handling noisy training data and performs well with large training datasets [14]. In a study conducted by Salma and Imelda (2024) on sentiment analysis of the second Covid-19 booster vaccination using the K-Nearest Neighbor and Support Vector Machine algorithms, the classification accuracy obtained by K-NN at k = 17 was 85.47%[10].

Logistic regression is a commonly used technique in analyzing data involving one or more predictor variables to predict a response variable, which typically has binary values of 1 (yes) and 0 (no). The response variable in logistic regression is related to the Bernoulli distribution [11]. In a study conducted by Audenza et al. (2023) on sentiment analysis of Gojek application reviews by implementing Logistic Regression, Multinomialnb, SVM, and K-NN, the best performance score by Logistic Regression sequentially is 82.45%, 82.49%, 82.45%, and 82.43%.

Based on previous research, it is outlined that the five algorithms to be used in this study yield fairly good accuracy. These five algorithms are known to perform very well and are proposed to be implemented in sentiment analysis. This research differs from previous studies as the topic used has not been extensively discussed and linked to the field of data mining. The aim of this research is to analyze public sentiment towards elsagate, whether it is positive, negative, or neutral.

II. Research Methodology

Several stages were conducted in this research, including data collection, labeling, preprocessing, TF-IDF weighting, and sentiment classification using machine learning algorithms. The stages in this research can be seen in Figure 1.

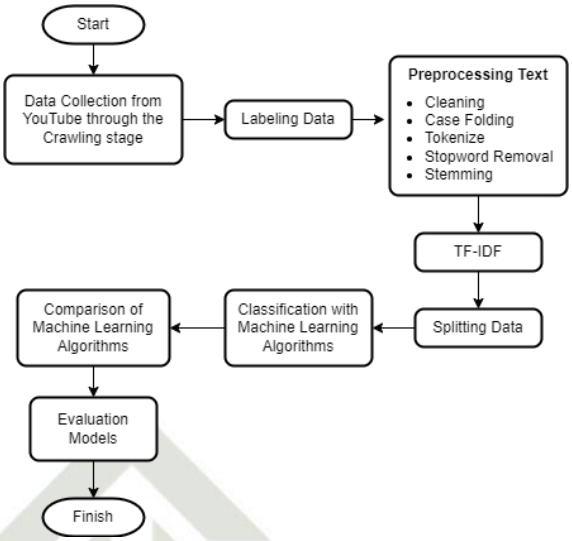

Fig. 1. Research Methodology

A. Data Collecting

The data used was taken from the YouTube platform, specifically the comments data on one of the uploads on the Duzzle account titled "Why is Youtube Kids Cringe?". The content has been identified as elsagate content after being examined by several researchers focusing on the elsagate phenomenon. Data collection was carried out using Crawling techniques with the YouTube Data API V3, resulting in a total of 2680 comment data obtained.

B. Pre-processing

Text preprocessing is the initial step in cleaning and preparing text before further analysis or processing. This process involves removing punctuation and special characters, converting text to lowercase, text tokenization, removing stop words, and standardizing text to improve data quality and facilitate the analysis process [12][13].

C. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF (term frequency-inverse document frequency) is a commonly used weighting method in natural language processing to assign weights to words in documents based on their importance or informativeness [19]. The TF-IDF calculation is performed by multiplying the Term Frequency (TF) value with the Inverse Document Frequency (IDF) value [14][15].

$$Wdt = tfdt * Idft \quad (1)$$

D. Confusion Matrix

Confusion matrix is a table that displays the number of test data classified correctly and the number of test data classified incorrectly [20]. For example, Table 1 is a confusion matrix for binary classification showing the distribution of correct and incorrect classification results.

TABLE I. Confusion Matrix

		Prediction Class	
		Postive	Negative
Actual Class	Positive	TP	FP
	Negative	FN	TN

UIN SUSKA RIAU
Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

UIN SUSKA RIAU
Hak Cipta dilindungi Undang-Undang
1. Dilarang menggutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

E. Naïve Bayes Classifier

Naïve Bayes is the most commonly used classification method for sentiment analysis, which calculates the probability of each factor and then selects the result with the highest probability [16]. Here is the formula for Naïve Bayes Classifier.

$$P(C|X) = \frac{P(X|C)\,P(c)}{P(x)} \quad (2)$$

F. Random Forest

Random Forest is a flexible and accurate method for regression and classification, including cases with multiple classes. It is fast in training and prediction, capable of handling large datasets well. This method provides insights into the importance of variables without requiring complex preprocessing. Additionally, it reduces variability and is more robust than single decision trees or boosting methods. Random Forest is also effective in addressing overfitting, implementing class weighting, detecting outliers, and handling missing values[17]. Here is the formula for Random Forest.

$$Fi_i \quad \frac{\sum j:\, node\ j\ splits\ on\ feature\ i\ ni_j}{\sum k \in all\ nodes\ ni_k} \quad (3)$$

G. Support Vector Machine

Support Vector Machine (SVM) is a common method in Supervised Learning used for classification and regression. In classification, SVM stands out due to its mature and clear mathematical concepts. SVM can handle classification and regression problems well in both linear and nonlinear contexts [18]. Here is the formula for Support Vector Machine.

$$w.x + b = 0 \quad (4)$$

H. K-Nearest Neighbors

K-Nearest Neighbor is used to classify new objects by comparing them to the most similar training data. The K-Nearest Neighbor algorithm is relatively easy to implement and allows users to determine the desired number of nearest neighbors [19]. Here is the formula for Euclidean Distance in K-Nearest Neighbors.

$$dis\,(x1, x2) = \sqrt{\sum}\,(x1i - x2i)2n \quad (5)$$

I. Logistic Regression

Logistic regression is a classification method commonly used to model the relationship between dependent and independent variables through linear logarithms [20]. Here is the formula for Logistic Regression.

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-x}} \quad (6)$$

J. K-Fold Cross Validation

K-Fold Cross Validation is one of the techniques used to partition data into training and testing data [21]. Additionally, K-Fold Cross Validation implements data splitting by dividing the total number into balanced k subsets of data. This process is repeated k times in iterative training and testing.

III. Results and Discussion

A. Initial Data

The initial data obtained from one of the YouTube contents discussing elsagate can be seen in Table 2.

TABLE II. INITIAL DATA

No	Username	Date	Text
1	@Duzzles	2022-10-02	Sorry kalo hasil Recordnya agak patah - patah, ternyata PC gua juga gak kuat nonton YouTube Kids.
2	@pluffs8	2022-10-02	acumalaka bgt beliau
3	@akuzheva1088	2022-10-02	Pc lu bayi
…	…	…	…
2680	@Duzzles	2022-10-02	Wih first nih..

B. Labeling

The collected comment data entered the Labeling stage, which was done manually with the assistance of language experts. Through the comment labeling process, public sentiment related to this research topic can be generated, covering positive, negative, and neutral values. The results of the labeled data can be seen in Table 3.

TABLE III. LABELING RESULT

No	Text	Sentiment
1	Sorry kalo hasil Recordnya agak patah - patah, ternyata PC gua juga gak kuat nonton YouTube Kids.	Netral
2	acumalaka bgt beliau	Negatif
3	Pc lu bayi	Negatif
…	…	…
2680	Wih first nih..	Netral

C. Cleaning Data

The processed data in text form will first be cleaned to remove duplicate words, symbols, hashtags, usernames, URLs, and others, so that the result will be as shown in Table 4.

TABLE IV. CLEANING RESULTS

No	Text	Sentiment
1	sorry kalo hasil recordnya agak patah - patah ternyata pc gua juga gak kuat nonton youtube kids	Netral
2	acumalaka bgt beliau	Negatif
3	pc lu bayi	Netral
…	…	…
2452	manuk akal	Netral

Based on Table 4, after cleaning the data, the comments are all written in lowercase and without emoticons, @ symbols, hashtags (#), and URLs. The number of data also changed, initially consisting of 2680 data, after data cleaning, it became 2452 data. From the manual labeling results by language experts after cleaning, the number of comments with positive sentiment is 1314 data, comments with negative sentiment are 570 data, and comments with neutral sentiment are 568 data.

UIN SUSKA RIAU

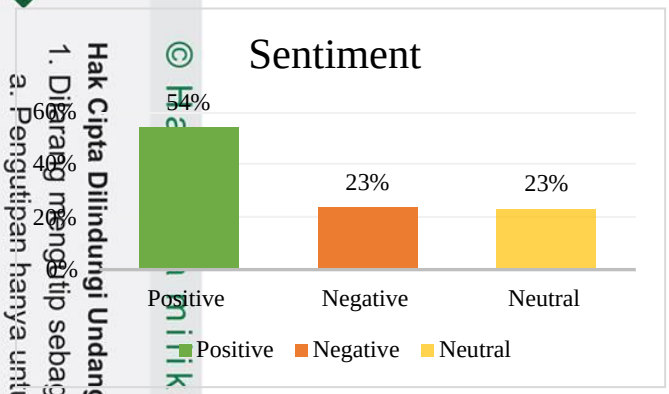

Fig. 2. Sentiment Percentage

D. Preprocessing

Pre-processing is a stage carried out by eliminating unsuitable data or transforming data into a form that is easier for the system to process. In the data pre-processing stage, Python programming language is used, and Visual Studio Code is used. The initial steps in data preparation for this study involve cleaning, filtering, and stemming. During the filtering process, significant terms are isolated from sentences to eliminate irrelevant elements like punctuation and stopwords. A stoplist is employed to identify these stopwords. Subsequently, words are transformed into their root forms, accomplished by identifying the base word for each term through literature modules in the Python programming language.

E. Term Frequency-Inverse Document Frequency (TF-IDF)

The data after the pre-processing stage must be in numeric form. To convert the data into numerical values, we use the TF-IDF weighting method. TF-IDF is a weighting system that assigns weights to each word in a document based on term frequency (tf) and inverse document frequency (idf). In the TF-IDF stage, calculations are performed using Python using the scikit-learn model. TF-IDF is divided based on the division of training data and test data. The TF-IDF results can be seen in Table 5.

TABLE V. TF-IDF RESULTS

No	Terms					
	aaaaaa	aaahh	aamiin	absurb	…	zuzuzu
1	0.00	0.00	0.00	0.00	…	0.00
2	0.00	0.00	0.00	0.00	…	0.00
3	0.00	0.00	0.00	0.00	…	0.00
4	0.00	0.00	0.00	0.00	…	0.00
…	…	…	…	…	…	…
2452	0.00	0.00	0.00	0.00	…	0.00

F. Splitting Training Data and Testing Data

Based on the described research methodology, this study divides the training and testing data taken from the initial dataset with a total of 2452 data into 3 ratios: 90:10, 80:20, 70:30 after going through several pre-processing stages and the TF-IDF stage.Kemudian data akan diuji menggunakan algoritma NBC, SVM, K-NN, RF, dan LR.

G. Machine Learning Algorithm Classification Results

In this study, data modeling has been conducted using several machine learning algorithms, namely Naïve Bayes Classifier, Support Vector Machine, K-Nearest Neighbors, Random Forest, and Logistic Regression to measure which algorithm performs better using 3 data separation ratios. The results of several modeling algorithms can be seen in Table 6.

TABLE VI. ALGORITHM PERFORMANCE RESULTS

Train-Test Splitt	Algorithm	Accuracy	Precision	Recall	F1-Score
90:10	NBC	62%	62%	62%	57%
	RF	59%	56%	59%	56%
	SVM	63%	61%	63%	60%
	K-NN	56%	55%	56%	55%
	LR	63%	56%	59%	56%
80:20	NBC	60%	61%	60%	65%
	RF	60%	58%	60%	56%
	SVM	61%	59%	61%	57%
	K-NN	56%	55%	56%	55%
	LR	61%	62%	62%	58%
70:30	NBC	62%	63%	62%	56%
	RF	60%	58%	60%	56%
	SVM	61%	60%	62%	58%
	K-NN	56%	57%	57%	56%
	LR	63%	62%	62%	58%

Based on the modeling results presented in Table 6, using 5 machine learning algorithms for data classification, the best experimental results were found with the 90:10 ratio. The SVM algorithm showed the best performance, with an accuracy of 63%, precision of 61%, recall of 63%, and an F1-score of 60%. Conversely, the K-NN algorithm showed the lowest performance, with an accuracy of 56%, precision of 55%, recall of 56%, and an F1-score of 55%.

H. K-Fold Cross Validation

The Support Vector Machine achieves the highest accuracy when evaluated based on the Confusion Matrix. Therefore, further testing was conducted by comparing k-fold cross-validation with 10 iterations. The results can be seen in Figure 3.

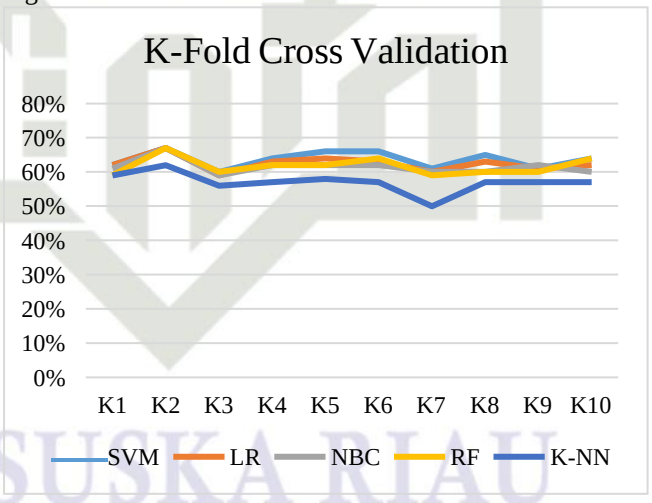

Fig. 3. K-Fold Cross Validation

Based on the diagram in Figure 3, it can be concluded that when evaluated using k-fold cross-validation with 10 iterations, the Support Vector Machine almost always produces the best performance compared to the other four algorithms.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

UIN SUSKA RIAU
Hak cipta dilindungi undang-undang
1. Dilarang menyalin atau mendistribusikan karya tulis ini tanpa mencantumkan atau menyebutkan sumber:
a. Penggunaan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Sultan Syarif Kasim Riau

I. *Word Visualization*

A Wordcloud is a visual representation of words used to observe how often words appear in a text document. The wordcloud application will display snippets of words in the comments found in the content, these snippets of words can be seen in Figure 4.

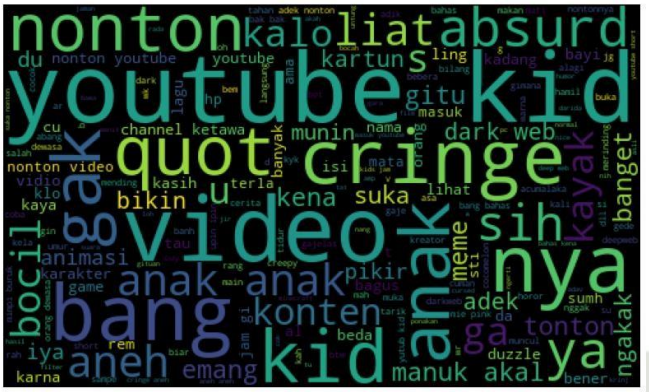

Fig. 4. Visualization Data

In Figure 4, it's a visualization of the word cloud in the dataset related to Elsagate content, where the frequently appearing words in the content on the YouTube platform are "YouTube Kid", "video", "cringe", "absurd", and others.

IV. Conclusion

The research results indicate that sentiment analysis using five machine learning algorithms (NBC, RF, SVM, K-NN, and LR) with data splits of 90:10, 80:20, and 70:30 yielded the best performance with the 90:10 split, where SVM achieved the highest accuracy of 63%, precision of 61%, recall of 63%, and an F1-score of 60%. A k-fold cross-validation test (10 iterations) was conducted to validate whether SVM outperformed the other four algorithms. The results showed that SVM almost always achieved the highest accuracy in each iteration. This concludes that Support Vector Machine is suitable for this study. Sentiment labeling by linguistic experts resulted in 54% positive sentiment, proving the public's lack of awareness about Elsagate. This study emphasizes the importance of government action, child protection agencies, and related stakeholders to reduce Elsagate-like content and raise public awareness in selecting children's content.

References

[1] H. Ahmad *et Al.*, "Impact Of Excessive Usage Of Smartphone In Kids," *International Journal Of Advanced Trends In Computer Science And Engineering*, Vol. 10, No. 3, Pp. 2351–2359, Jun. 2021, Doi: 10.30534/Ijatcse/2021/1201032021.

[2] A. Mohan, P. Sen, D. Mujumdar, C. Shah, And E. Jain, "Series Of Cases Of Acute Acquired Comitant Esotropia In Children Associated With Excessive Online Classes On Smartphone During Covid-19 Pandemic; Digital Eye Strain Among Kids (Desk) Study-3," *Strabismus*, Vol. 29, No. 3, Pp. 163–167, 2021, Doi: 10.1080/09273972.2021.1948072.

[3] W. Han And M. Ansingkar, *2020 Zooming Innovation In Consumer Technologies Conference (Zinc) : Online, 26-27 May 2020*. 2020. Doi: 10.1109/Zinc50678.2020.9161808.

[4] K. Papadamou *et Al.*, "Disturbed Youtube For Kids: Characterizing And Detecting Inappropriate Videos Targeting Young Children," Jan. 2019, [Online]. Available: Http://Arxiv.Org/Abs/1901.07046

[5] W. Dirgantara, F. Iqbal Maulana, S. Subairi, And R. Arifuddin, "The Performance Of Machine Learning Model Bernoulli Naïve Bayes, Support Vector Machine, And Logistic Regression On Covid-19 In Indonesia Using Sentiment Analysis."

[6] J. Mantik, E. R. Putri, And H. Februariyanti, "Product Review Sentiment Analysis At Online Store Jiniso Official Shop Using Naive Bayes Classifier (Nbc) Method," Online, 2022.

[7] M. Guia, R. R. Silva, And J. Bernardino, "Comparison Of Naive Bayes, Support Vector Machine, Decision Trees And Random Forest On Sentiment Analysis," In *Ic3k 2019 - Proceedings Of The 11th International Joint Conference On Knowledge Discovery, Knowledge Engineering And Knowledge Management*, Scitepress, 2019, Pp. 525–531. Doi: 10.5220/0008364105250531.

[8] Stephenie, B. Warsito, And A. Prahutama, "Sentiment Analysis On Tokopedia Product Online Reviews Using Random Forest Method," In *E3s Web Of Conferences*, Edp Sciences, Nov. 2020. Doi: 10.1051/E3sconf/202020216006.

[9] R. Singla, B. Chambayil, A. Khosla, And J. Santosh, "Comparison Of Svm And Ann For Classification Of Eye Events In Eeg," *J Biomed Sci Eng*, Vol. 04, No. 01, Pp. 62–69, 2011, Doi: 10.4236/Jbise.2011.41008.

[10] A. R. Isnain, J. Supriyanto, And M. P. Kharisma, "Implementation Of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis Of Online Learning," *Ijccs (Indonesian Journal Of Computing And Cybernetics Systems)*, Vol. 15, No. 2, P. 121, Apr. 2021, Doi: 10.22146/Ijccs.65176.

[11] W. Dirgantara, F. Iqbal Maulana, S. Subairi, And R. Arifuddin, "The Performance Of Machine Learning Model Bernoulli Naïve Bayes, Support Vector Machine, And Logistic Regression On Covid-19 In Indonesia Using Sentiment Analysis."

[12] G. A. Buntoro, R. Arifin, G. N. Syaifuddiin, A. Selamat, O. Krejcar, And H. Fujita, "Implementation Of A Machine Learning Algorithm For Sentiment Analysis Of Indonesia'S 2019 Presidential Election," *Ium Engineering Journal*, Vol. 22, No. 1, Pp. 78–92, 2021, Doi: 10.31436/Iiumej.V22i1.1532.

[13] H. Tsaniya, R. Rosadi, And A. S. Abdullah, "Sentiment Analysis Towards Jokowi`S Government Using Twitter Data With Convolutional Neural Network Method," In *Journal Of Physics: Conference Series*, Iop Publishing Ltd, Jan. 2021. Doi: 10.1088/1742-6596/1722/1/012017.

[14] R. Afrinanda, W. Tawa Bagus, And L. Efrizoni, "Comparison Of Machine Learning Algorithm Models In Bitcoin Price Sentiment Analysis," *Indonesian Journal Of Computer Science Attribution*, Vol. 12, No. 2, Pp. 2023–502.

[15] A. W. Sari, T. I. Hermanto, And M. Defriani, "Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine," *Sinkron*, Vol. 8, No. 3, Pp. 1366–1378, Jul. 2023, Doi: 10.33395/Sinkron.V8i3.12447.

[16] H. Wisnu, M. Afif, And Y. Ruldevyani, "Sentiment Analysis On Customer Satisfaction Of Digital Payment In Indonesia: A Comparative Study Using K-NN And Naïve Bayes," In *Journal Of Physics: Conference Series*, Institute Of Physics Publishing, Feb. 2020. Doi: 10.1088/1742-6596/1444/1/012034.

[17] F. Ichsani Rafif, M. Dwifebri Purbolaksono, And W. Astuti, "Sentiment Analysis Using Random Forest And Word2vec For Indonesian Language Movie Reviews," Vol. 7, No. 3, Pp. 1109–1116, 2023, Doi: 10.30865/Mib.V7i3.6299.

[18] A. Maulana, I. Afifah, A. Mubarrak, K. Fauzan, A. Dwintara, And B. Zen, "Comparison Of Logistic Regression, Multinomialnb, Svm, And K-Nn Methods On Sentiment Analysis Of Gojek App Reviews On The Google Play Store," *Jurnal Teknik Informatika (Jutif)*, Vol. 4, No. 6, Pp. 1487–1494, Dec. 2023, Doi: 10.52436/1.Jutif.2023.4.6.863.

[19] A. R. Isnain, J. Supriyanto, And M. P. Kharisma, "Implementation Of K-Nearest Neighbor (K-Nn) Algorithm For Public Sentiment Analysis Of Online Learning," *Ijccs (Indonesian Journal Of Computing And Cybernetics Systems)*, Vol. 15, No. 2, P. 121, Apr. 2021, Doi: 10.22146/Ijccs.65176.

[20] T. H. J. Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, And M. W. Adisaputra, "Sentiment Analysis Of Twitter Data Related To Rinca Island Development Using Doc2vec And Svm And Logistic Regression As Classifier," In *Procedia Computer Science*, Elsevier B.V., 2021, Pp. 660–667. Doi: 10.1016/J.Procs.2021.12.187.

[21] R. Syahputra, G. J. Yanris, And D. Irmayani, "SVM And Naïve Bayes Algorithm Comparison For User Sentiment Analysis On Twitter," *Sinkron*, Vol. 7, No. 2, Pp. 671–678, May 2022, Doi: 10.33395/Sinkron.V7i2.11430.

UIN SUSKA RIAU

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Ria

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A

DOKUMENTASI CONFERENCE

IEEE INDONESIA SECTION

SIME

Certificate

Awarded to

Zaira Cindya Dwynne

has **presented** the paper titled

Comparison of Machine Learning Algorithms on Sentiment Analysis of Elsagate Content

at the International Conference on Smart Computing, IoT, and Machine Learning (SIML) 2024

Surakarta, 6-7 June 2024

SIME

Dr. Gunawan Ariyanto
General Chair of SIML 2024

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

LAMPIRAN B

SURAT DARI JURNAL

B.1 Copyright Receipt

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

Comparison of Machine Learning Algorithms on Sentiment Analysis of Elsagate Content

Ms. Zaira Cindya Dwynne, Mr. Mustakim Mustakim and Mr. Mustafa Mustafa

2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)

COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the IEEE PSPB Operations Manual.
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

You have indicated that you DO wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."

CONSENT AND RELEASE

1. In the event the author makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the author, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Ria

permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.

2. In connection with the permission granted in Section 1, the author hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

ZAIRA CINDYA DWYNNE — 05-05-2024

Signature — Date (dd-mm-yyyy)

Information for Authors

AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from

UIN SUSKA RIAU

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Ria

B.2 E-Mail Accepted

ZAIRA CINDYA DWYNNE Sistem Informasi <12050324080@students.uin-suska.ac.id>

SIML 2024 Notification for Paper 1571009542

simi2024-chairs@edas.info <simi2024-chairs@edas.info> 26 April 2024 pukul 13.41
Kepada: Zaira Cindya Dwynne <12050324080@students.uin-suska.ac.id>, Mustakim Mustakim <mustakim@uin-suska.ac.id>, Mustafa Mustafa <mustafa@uin-suska.ac.id>

Dear Ms. Zaira Dwynne:

We are pleased to inform you that your paper #1571009542 titled 'Comparison of Machine Learning Algorithms on Sentiment Analysis of Elsagate Content' has been accepted for presentation at the 2024 International Conference on Smart Computing, IoT and Machine Learning (SIML).

Your submission has met the high standards required by our review committee and promises to contribute valuable insights to the field of smart computing, IoT, and machine learning. We are excited about the potential impact of your work and look forward to your presentation at the conference.

For the next step, please follow the below instructions:
1. Revise your paper(s) according to the reviewers' comments. The detailed review is listed below this email. Please be advised that failure to make the required revisions will result in the rejection of your paper and exclusion from the conference proceedings.
2. Each accepted paper has to have one author registered as a speaker (not a Participant). Fill out the registration form that can be accessed from the website (https://siml.ums.ac.id/2024/registration/) and submit it by including the proof of your payment. You can also upload proof of IEEE membership and proof of student status when it is relevant. See the registration fee link to make sure how much you have to pay.
3. For every manuscript in the SIML 2024 Proceedings, copyright consent must be given by the authors via EDAS. The signed copyright consent form of each accepted paper MUST be submitted before uploading your final manuscript. To submit your copyright consent, please log in to EDAS, go to your accepted paper details and click on the button close to “Copyright form”.
4. Your final submission must strictly follow the A4-IEEE format. Detailed guidelines for preparing your camera-ready paper can be found on the conference website (https://siml.ums.ac.id/2024/submission/). Please ensure that your submission complies with these formatting instructions. Papers that do not meet these formatting requirements will not be published in the conference proceedings.

Please be reminded that the camera-ready deadline is on 3 May 2024, and the registration deadline is on 10 May 2024. Please ensure that you adhere to these deadlines to facilitate the smooth organization of the conference.

The conference will take place by Hybrid from June 6-7, 2024 in Surakarta, Indonesia. The schedule will be posted on the conference website as soon as it is completed.

We look forward to your participation in SIML 2024 and are excited to see how your research contributes to our discussions. Should you have any questions or require further information, please do not hesitate to contact us at siml@ums.ac.id.

Warm Regards,
SIML 2024 Organizers

The reviews are below or can be found at https://edas.info/showPaper.php?m=1571009542.

======= Review 1 =======

> *** Detailed Comments: Please justify your recommendation and suggest improvements in technical content or presentation.

1. Introduction section
-must Explain the context of the study and state the precise objective
authors should give an idea of the state-of-the art of the field the report is about.
- please tell readers why they should proceed reading. Experience shows that for this part a few lines are often sufficient.
- Authors should place the paper in proper context by citing relevant papers to support this section.

2. Results and Discussion
- The presentation of results should be simple and straightforward in style.
- This section reports the most important findings and must be comparisons to other research results.

UIN SUSKA RIAU

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Conclusion
the authors must be summarize sentences the primary outcomes of the study in a paragraph (NOT in numbering).

======= Review 2 =======

> *** Detailed Comments: Please justify your recommendation and suggest improvements in technical content or presentation.

The paper compares several machine learning methods for sentiment analysis in Elsagate content on YouTube. The topic is interesting since the term Elsagate is yet to be familiar to the public. However, several areas require further revision.
1. No need to mention the title of the paper that we refer to in section 1
2. Authors should add literature about sentiment analysis on YouTube comments.
3. The authors should provide a clear explanation for their choice to obtain comments from the 'Mengapa Youtube Kids Cringe?' in the Duzzle account rather than crawling comments from YouTube videos containing Elsagate content.
4. Each variable in the equation needs to be explained.
5. There is a confusion matrix table in the sub-section confusion matrix, but it does not mention its use. It would be beneficial for the authors to include the equation used to calculate the matrix measurement for accuracy, precision, and recall in their experiment evaluation based on Table 1.
6. The authors should address the issue of dataset imbalance (the positive label is twice the number of negative and neutral labels), which may have contributed to the low performance of the model.
7. It would be more helpful if the author provided a word cloud for each sentiment, which would give more insight into the dominant words for each sentiment.
8. The word cloud still contains a lot of noise, such as "gw, yg, yt, gak, ya"; that term can be removed by adding them as additional terms in the stopword list.

======= Review 3 =======

> *** Detailed Comments: Please justify your recommendation and suggest improvements in technical content or presentation.

The paper is very interesting and valuable scientifically, however I have some comments.
Comment #1 abstract
Motivation of the manuscript should be clearly stated in the abstract.
Comment #2 introduction
Contribution of the manuscript should be stated at the end of Introduction in a brief description
Comment #3 research methodology
It is not define clearly about the explanation of methodology (ex : the authors need to provide explanation for each step, contain in fig 1)
Is it enough to use confusion matrix as an evaluation technique to classify data. Please explain the performance use.
It is necessary to justify why the author use the 5 algorithms among other method
Comment #4 result and discussion
It is necessary to describe more detail for preprocessing phase, which consist of 3 steps, and the example result of it.
For TF-IDF result, the authors should give the data that contain the top highest weight that the specific terms have.
Comment #5 references
Please check the bibliography writing.

UIN SUSKA RIAU

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

LAMPIRAN C

SURAT PERNYATAAN PELABELAN

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama	: Nuratika, S.Hum., M.Pd
Tempat Tanggal Lahir	: Sibiruang XIII Koto Kampar, 02 Desember 1988
Pekerjaan	: Dosen Pendidikan Bahasa dan Sastra Indonesia
Alamat	: RK Harapan Kelurahan Ujungbatu Kecamatan Ujungbatu Kabupaten Rokan Hulu

Menyatakan dengan sesungguhnya bahwa benar memvalidasi untuk pelabelan sentimen yang bersifat positif, negatif dan netral dalam komentar *Youtube* dengan unggahan yang membahas mengenai *Reaction* Elsagate/Konten Tidak Ramah Anak pada *Youtubekids* sebanyak 2860 data dalam Tugas Akhir dari:

Nama	: Zaira Cindya Dwynne
NIM	: 12050324080
Jurusan	: Sistem Informasi
Judul Tugas Akhir	: Comparison of Machine Learning on Sentiment Analysis of Elsagate Content

Demikian surat ini saya buat dengan sesungguhnya tanpa paksaan dari pihak manapun. Atas perhatiannya saya ucapkan terima kasih.

Pekanbaru, 23 Februari 2024

Yang membuat pernyataan

Nuratika, S.Hum., M.Pd

UIN SUSKA RIAU

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Ria

DAFTAR RIWAYAT HIDUP

Zaira Cindya Dwynne lahir di Kota Dumai, pada tanggal 08 Mei 2002. Penulis merupakan anak dari Bapak Ardi dan Ibu Betty Yuliane. Penulis adalah anak bungsu dari empat bersaudara. Pada tahun 2007 penulis memulai pendidikan di TK Melati Ujung Batu, Kabupaten Rokan Hulu dan lulus pada tahun 2008. Lalu melanjutkan pendidikan Sekolah Dasar di SD Negeri 001 Ujung Batu, kemudian pindah ke kota Pekanbaru dan melanjutkan kembali pendidikan Sekolah Dasar pada tahun 2011 di SD Negeri 048 Pekanbaru. Penulis menyelesaikan pendidikan Sekolah Dasar pada tahun 2014. Setelah menyelesaikan pendidikan Sekolah Dasar penulis melanjutkan pendidikan tingkat SLTP di SMP Negeri 35 Pekanbaru yang selesai pada tahun 2017. Penulis melanjutkan pendidikan tingkat SLTA di SMA Negeri 14 Pekanbaru. Setelah menyelesaikan pendidikan di SMA Negeri 14 Pekanbaru pada tahun 2020, penulis pun melanjutkan pendidikan dengan menjadi mahasiswa Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau dan penulis menyelesaikan kuliah Strata satu (S1) tersebut pada tahun 2024. Selama perkuliahan penulis aktif dalam mengikuti berbagai kegiatan yang diadakan oleh kampus maupun luar kampus. Selain itu penulis juga pernah menjadi anggota *Puzzle Research Data Technology* (PREDATECH) mulai dari tahun 2020 hingga 2024. Penulis juga pernah menjadi kepala dinas kewirausahaan Dewan Eksekutif Mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau periode 2022. Penulis juga pernah mengikuti kegiatan Kuliah Kerja Nyata (KKN) Nusantara Moderasi Beragama yang diselenggarakan di Kabupaten Tana Toraja, Provinsi Sulawesi Selatan pada tahun 2023. Pada penelitian Tugas Akhir ini penulis mengambil judul penelitian Tugas Akhir yaitu *"Comparison of Machine Learning Algorithms on Sentiment Analysis of Elsagate Content"*.

UIN SUSKA RIAU