CoMBInES
CONFERENCE ON MANAGEMENT, BUSINESS, INNOVATION, EDUCATION AND SOCIAL SCIENCES

Volume 4 No 1 (2024)

No DOI:

Received : February 01, 2021
Accepted : February 05, 2021
Published : February 24, 2021

Conference on Management, Business, Innovation, Education and Social Science
https://journal.uib.ac.id/index.php/combines

Fılm Serıes Revıew Analysıs Usıng Twıtter Apı And Text Mınıng Model

Herman[1], Franky[2]
Email of author correspondence: tanfranky0@gmail.com

[1]Faculty Of Computer Science, Universitas International Batam, Batam, Indonesia
[2]Faculty Of Computer Science, Universitas International Batam, Batam, Indonesia

Abstract
The development of information technology has brought about many changes to society in the industrial, economic, social, and even digital realms, including in the world of filmmaking. As the world of filmmaking has evolved, genres within film series have proliferated. This has created difficulties in navigating through recommendations for currently airing series due to the lengthy duration of episodes. Consequently, every enthusiast intending to watch a series must possess a high level of attraction to follow each episode. With the abundance of reviews, this can serve as a consideration for every viewer regarding a film series, allowing conclusions to be drawn about it. This research utilizes data from Twitter processed through machine learning and calculated using algorithms, comparing the results of analysis and experiments that demonstrate fairly accurate accuracy. The above research findings prove that the use of the Naïve Bayes machine learning algorithm provides a comprehensive overview of the assessment of film series. The predictions from the Naïve Bayes algorithm demonstrate sufficiently accurate accuracy compared to the experimental method and the analysis results from Rotten Tomatoes.

Keywords: Film Series, Twitter, Machine Learning, Naïve Bayes

Introduction

The development of information technology has made many changes to society in the industrial, economic, social, and even digitalization worlds. One of the developments that is quite rapid to date is the era of digitalization in the world of film. Film is a medium that has several perspectives, including as a medium for art, education and industrial development (Vivia, 2020). Film itself has several branches, including Film Series or commonly referred to as Series and Film Movies. In the current era of digitalization, films not only display images and audio, but are further developed in video visualization and text so that films are easy to understand and more interesting.

As the world of cinema develops, film series themselves have many genres, such as Drama, Comedy, Thriller, Horror, and so on. This makes it difficult to see every recommended series that is being broadcast, this is because in 1 series there are many episodes, at least 1 series has more than 12 episodes with an average duration of 45 minutes. So every fan who wants to watch the series must have a high level of interest in following the series every episode. Every year in Indonesia and internationally the newest Film Series will always be shown. However, the more film series that are shown, it makes it more difficult for people to choose which series to watch first. Film series viewing platforms that are well known to many audiences include Netflix, Wetv, Disney Plus, and so on, which are not uncommon. Film series fans not only watch but also provide reviews about the film series they are watching. With so many reviews submitted, this can be used as a consideration for every viewer regarding a film series, so that from this they can draw conclusions about the film series. With this, text mining is needed to display information data, which greatly influences the number of film series fans, this is because the reviews given are varied, such as positive, negative or neutral assessments, so that this can influence the influence of audience interest in each film series title.

One of the existing information search applications is Rotten Tomatoes, where information is obtained using web mining methods. Web mining is an application for searching based on keywords with certain techniques assisted by Data Mining. Data Mining is a method of analyzing data sets to find relationships and draw data conclusions, where the results can be understood using certain techniques or methods. The use of data mining techniques is expected to help speed up the decision-making process, allowing companies to manage the information contained in customer data and gain new insights to determine the necessary decisions (Wijaya, Herry Derajad, 2020). In the process of collecting and searching for data, an application is needed that can help collect and extract data in the form of comments from every viewer who has watched a film or series, so that from collecting these comments a conclusion can be made from a film or series. Social media is one of the platforms that produces the most data (Septia, Fitria, 2021), one of which is Twitter. The results of the data that have been obtained show that Twitter is a real-time information sharing social media that has quite large potential for digging up information that is currently trending or has already been published. occurred previously in Indonesia and outside Indonesia (Taofik Krisdiyanto, 2021). Twitter is also a social media which provides the opportunity for every individual to provide real-time criticism or suggestions regarding the world of films that are trending or have been broadcast for a long time, so that through the Twitter application the public can draw conclusions little by little regarding the assessment of films and series. which is being or will be broadcast.

In the data mining process, machine learning is needed which has algorithms to assist in carrying out calculations and data processing. The machine learning designed will use the KDD (Knowledge Discovery in Database Process) method to obtain large amounts of data and predict accuracy. The classification algorithm used is Naive Bays, and uses NLP (Natural Language Processing) technology, which is a machine learning technology to help the data search process. Naive Bayes is an algorithmic method used to predict probability in a class of data, which can be used to predict in various fields such as learning analysis (Christian, 2022), analysis predicting basic food prices, and so on. After data mining was carried out, the next step was for the author to carry out an experimental method with the help of a Google form on 500 people to obtain data and compare it with data mining that had been carried out previously as well as the results from the Rotten Tomatoes data mining website whether the accuracy of the comparison of the 3 methods had sufficient accuracy. in accordance.

Based on the increasing number of film series that exist every year, the author feels it is necessary to conduct research in terms of analyzing film series recommendations which aims to conclude whether a film series is good or bad to become recommendation material based on data that has been collected from the Twitter platform and by conducted an experiment with the help of a Google form on 500 respondents, and also compared the results from the Rotten Tomatoes web mining so that it could be concluded whether the film series could be used as a recommendation to the public. Therefore, the research title chosen by the author is "Analysis of Film Series Reviews Using the Twitter API and Text Mining Model".

Literature Review

The research entitled "Analysis of Film Series Reviews Using the Twitter API and Text Mining Model" is based on several previous studies, namely as follows:

Research conducted by Saputra, Aldi Rahmad & Sejati Waluyo (2022) is the research that is the main basis for this research. The research analyzes the increase in fuel oil. The aim of this research is to prove whether tweets made by people on Twitter regarding the increase in fuel prices have a big influence on people's lives so that they become a topic of conversation on Twitter for several days and chaos occurs in real life. This analysis was carried out using data crawling techniques on Twitter and the Naive Bayes algorithm method. The results of this research are that public respondents' opinions or tweets about complaints about fuel oil increases received positive value. In testing training data and test data with a ratio of 80 - 20, the researcher got a correct value of 81.00%, in a comparison of 70-30 the researcher got 83% correct value and in a comparison of 60 - 40 the researcher got a correct value of 77.50%.

The next research was conducted by Villarica, Mia V, et al (2022). This research designs a website that is useful for analyzing online crime incident reports. With crime reporting now being able to be done in real time via a website or application, therefore, researchers designed a website which includes crime features such as a location tracking feature to find out where crime often occurs, a crime analysis feature with artificial intelligence assistance using data mining. In developing the website, researchers used the SCRUM method, data mining assistance using the KDD method and to calculate data accuracy using three algorithms, namely KNN, Decision Tree, and Naive Bayes. The aim of this research is to design an online crime reporting system to reduce the increasing crime rate in Laguna (Philippines). The results of this

research, the design of the website and the data mining carried out to obtain data regarding crime analysis, have been successful because the website provides recommendations for the government to implement policies that must be implemented in order to reduce the value of crime.

The next research is research conducted by Ridwan, Achmad (2020). Researchers classify diabetes mellitus because diabetes mellitus is a disease that claims the largest number of lives in the world. So, researchers classify this disease based on the possibility of developing diabetes mellitus and the initial symptoms encountered when suffering from diabetes mellitus. This is because there are many possibilities that occur in diabetes mellitus, so researchers will classify diabetes mellitus. The method used by the research is quantitative and the Naive Bayes algorithm to obtain accuracy from the results of the evaluation process. The researcher's initial stage was to carry out data collection using machine learning methods, as well as creating a questionnaire, this aimed to obtain information from the public regarding diabetics or non-diabetics. After the data has been obtained, the data will then be evaluated using Naive Bayes to calculate how precise the accuracy obtained is. The aim of this research is to obtain data or information from the public about what causes people to get diabetes so that it will be evaluated again using the Naive Bayes algorithm to see whether it has sufficient accuracy. The results of this research are the results of the naive Bayes algorithm producing 16 classifications of risk predictions for diabetes. In these predictions the evaluation results were 90.20% so they had quite precise accuracy.

The next research is research conducted by Roihan Ahmad, et al (2020). This research aims to provide a more significant understanding of machine learning by conducting studies in various literature. Researchers collect data in the fields of traffic, industry, medicine and technology to solve problems and classify them using machine learning algorithms. The aim of this researcher is to define machine learning as having many benefits in various fields so that it can be used as a guideline that can be used by future researchers. The results of this researcher state that machine learning can be used to classify data and solve problems in various fields, using machine learning algorithms can speed up performance to obtain a high level of accuracy and precision.

The next research is research conducted by Meilani Nadya, et al (2023), in this research the accuracy of breast cancer prediction occurs. In this research, data mining was used using the KDD (Knowledge Discovery In Database) method to obtain large or extensive information, and also used the K-Nearest Neighbor (KNN) algorithm to predict whether the results were accurate enough. The aim of this research is to help hospitals predict the accuracy of diagnosing breast cancer. The results of this research are 279 data obtained, divided into 70% testing data and also 30% training data which shows good classification which produces an accurate accuracy value of 72.62% in predicting breast cancer patients where patients who suffer from breast cancer are caused by by fewer relapses than patients who did not relapse.

Table 1. Conclusion Bibliography

Author	Summary
(Saputra, Aldi Rahmad Waluyo, Sejati, 2022)	This research uses a data crawling technique to obtain results of public opinion on Twitter regarding the increase in fuel prices, and is calculated using Naïve Bayes.
(Villarica, Mia V Balahadia, Francis F Asor, Jonardo R Catedrilla, Gene Marck B, 2022)	In this research, we designed a website with SCRUM development to analyze online crime using data mining and calculating with 3 algorithms.
(Meilani Nadya & Odi Nurdiawan, 2023)	In this study, the accuracy of breast cancer classification was classified using machine learning with the KDD method and the KNN algorithm.
(Ridwan & Achmad, 2020)	In this study, we classify the risk of developing diabetes using machine learning and quantitative methods and calculating with the Naïve Bayes algorithm.
(Roihan, Ahmad Sunarya, Po Abas Rafika, Ageng Setiani, 2020)	In this research, we classify methods, solve problems and calculate problems in various fields using machine learning.

By referring to previous research, the author will design a website to detect film series reviews and develop the website which has been carried out by researchers Villarica, Mia V, et al (2022). In designing a film series review detection system, the author designed machine learning to help obtain data as was done by researchers Roihan, Ahmad, et al (2020). The author designed machine learning using the KDD method to predict accuracy carried out by Meilani Nadya, et al (2023). The machine learning that is designed will use crawling techniques to get opinions or views of the public on Twitter. After the data has been obtained, the author will then evaluate and classify the data using the Naive Bayes algorithm as done by Saputra, Aldi Rahmad (2022) and Ridwan Achmad (2020).

Research Methods

This picture below shows the author's steps in preparing the research method which was carried out in the form of a research flow

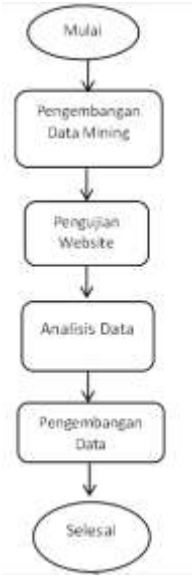

Figure 1. Research Flow

The initial stage carried out by the author is starting with a literature review, where he looks for references from various journals based on the topic being researched, then the author will identify the problems in the topic along with solving the problems that have been described. In the next stage the author creates and develops website-based data mining, where the designed website will display data mining results obtained by machine learning. The machine learning will be designed using the Knowledge Discovery In Database (KDD) development method. Next, the author will test website-based data mining that has been previously designed using the Knowledge Discovery In Database (KDD) method. Next, the author will collect data using an experimental method with the help of Google Form. The experiment will be carried out on 500 people and the experiment will include 10 film series titles. The next stage the author will carry out data analysis, where the data has been collected from website-based data mining results, through machine learning technology using the Naive Bays algorithm with the help of NLP (Natural Language Preprocessing), where the data collection is in the form of obtaining public opinion on Twitter so that will be reprocessed and calculated using the Naïve Bayes algorithm. The results of these calculations will be displayed on the designed website. Next, the author will compare the data resulting from the designed data mining (machine learning), experimental results, and the results of the Rotten Tomatoes website analysis, whether these three data have sufficient accuracy and then the author will write down the results of the data that has been obtained in the form of a report.

Data Mining Development

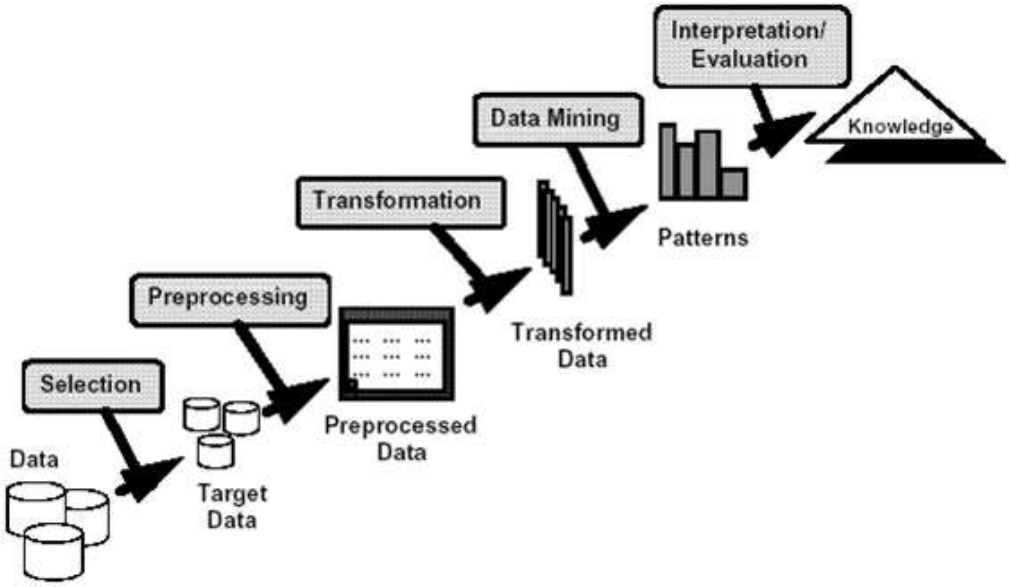

Figure 2. Knowledge Discovery In Database Flow

At the machine learning development stage, using the Knowledge Discovery in Database (KDD) stage, with the help of Natural Processing Language (NLP), namely:
1. Loading data / Data Selection: At the loading data or data selection stage, the process stage that will be carried out is reading the existing dataset using the Pandas library for CSV format as well as discarding data that has NA (null) followed by selecting the data that will be processed
2. Preprocessing: The next stage is preprocessing, where at this stage the sentiment dataset will be cleaned using the following stages:
a. Case folding: Case folding is the uniformization of data by changing the data to lowercase (lowercase)
b. Hashtag and Username removal: This is the stage of deleting hashtags and usernames on Twitter
c. Punctual removal: The stage of removing digits or some punctuation marks that are not necessary to use.
d. Digit removal
e. Negation Handling: at this stage, you will change the sentence containing the stop word negation into a complete word for negative sentiment. for example, not bad becomes good
f. Stopword Removal: Stage of removing irrelevant stopwords or auxiliary words, for example: be, too, so
g. Lemmatizing: The stage of changing affix verbs into basic words, for example going -> go
3. Transformation: The process stage of changing data into a form for processing to the next stage. In this research, data transformation will be carried out to convert the dataset into vector text data using TF-IDF (Term Frequency - Inverse Document Frequency)

4. Classification is a process at the data mining stage, where the process will be trained through kanggle data, and then the data will be trained using the Gaussian Naive Bayes method to get a classifier that will be used on other datasets as a model for classification, in this research to predict sentiment series review.
5. Evaluation: Next, after classification is carried out, the data will be used for comparison based on the classifier that has been trained with the confusion matrix measuring tool and obtain accuracy
6. Deploy: After further evaluation, the Vectorizer that has been created from TF-IDF and the Classifier that has been created from Gaussian Naive Bayes will be exported in .pkl form so that it can be used continuously without the need to go through text vectorization and classification repeatedly.

Testing Website

Website development is carried out using JavaScript language and using the UI library display as the final website display In the testing phase the author will use the Twitter API to obtain large amounts of data carried out by machine learning as data comparison material which will then be written into the report.

Data Collection Techniques

After the researcher obtains data from machine learning on the website, the next stage the researcher will obtain further data using an experimental method which will be carried out with the help of using a Google form which will be distributed to 500 people to show whether the 10 film series titles asked for are liked. by the community or not or neutral. This aims to obtain data and percentages that researchers will use as comparison material with the data results contained in machine learning.

Data Analysis

This data analysis stage is a continuation of the experimental data collection stage. After the researcher gets the results of the respondent data from the experimental method, the next stage is that the researcher will analyze the data contained in the Google Form, where in this analysis stage the data is obtained from the Google Form and also analyze the data on Rotten Tomatoes which will then be classified to see the positive percentage and negatives of the film series titles that have been included.

Data Comparison

The next researcher will compare the Naive Bayes algorithm machine learning data that has been processed and the results of the data using experimental methods and data analysis on Rotten Tomatoes to see whether the three data have sufficient accuracy so that they will be written in the report.

Results and Discussion

The following is a comparison table between the results of experimental methods and machine learning results with Rotten Tomatoes results.

Table 2. Comparison Table of Research Results

Film Series	Kategori Positif		Kategori Netral		Kategori Negatif		Rotten Tomatoes
	Metode Eskperimen	Machine Learning	Metode Eskperimen	Machine Learning	Metode Eskperimen	Machine Learning	
Manifest	51%	48%	32%	41%	17%	11%	68%
Naruto	44.9%	49%	19.1%	38%	36.1%	13%	81%
Ragnarok	66%	58%	11%	4%	23%	38%	63%
Boruto	44.9%	43%	19.2%	19%	35.9%	38%	54%
WandaVision	60.2%	60%	17.2%	21%	22.7%	19%	88%
Stranger Things	37%	50%	40.9%	12%	22%	38%	90%
Attack on Titan	52%	54%	24.4%	23%	23.6%	23%	94%
The Flash	52%	52%	21.9%	20%	26.1%	28%	59%
One Piece	46.9%	48%	28%	27%	25%	28%	95%
Loki	45.8%	46%	31%	29%	23.1%	25%	86%

Series "Manifest", the experimental method obtained results of 51% positive, 32% neutral, and 17% negative and machine learning results of 48% positive, 41% neutral, and 11% negative so it has positive reviews because the positive category has highest results. The Rotten Tomatoes results have an Average Audience Score of 68%, where in the Rotten Tomatoes results the film series Manifest gets the Full Popcorn Bucket logo so it is declared positive. Series "Manifest" has positive results.

"Naruto" series, based on the experimental method the results were 44.9% positive, 19.1% neutral, and 36.1% negative and the machine learning results were 49% positive, 38% neutral, and 13% negative so it had positive reviews because the positive category had highest results. The Rotten Tomatoes results have an Average Audience Score of 81%, where

the Naruto film series results get the Full Popcorn Bucket logo, so it can be stated that the Rotten Tomatoes results are positive. It can be concluded from the three data that "Naruto" series has positive results.

Series "Ragnarok", based on the experimental method the results were 66% positive, 11% neutral and 23% negative and machine learning obtained results of 58% positive, 4% neutral and 38% negative so it has positive reviews because of the positive category has the highest yield. The Rotten Tomatoes Average Audience Score result is 63%, where this series gets the Full Popcorn Bucket logo so it can be declared to have had positive results. It can be concluded that the Ragnarok series high positive result.

Series "Boruto", based on the experimental method the results were 44.9% positive, 19.2% neutral, and 35.9% negative and machine learning obtained results of 43% positive, 19% neutral, and 38% negative so it has positive reviews because of the positive category has the highest yield. The Rotten Tomatoes results which have an Average Audience Score of 54%, the Boruto film series received the Spilled Popcorn Bucket logo and were declared to have a Negative result. It can be concluded that the Twitter data collection and experiments obtained positive results that were higher than neutral and negative, but the negative values obtained were also quite high in line with the Rotten Tomatoes results.

"WandaVision" series, based on the experimental method the results were 60.2% positive, 17.2% neutral, and 22.7% negative and machine learning obtained results of 60% positive, 21% neutral, and 19% negative so it has positive reviews because the positive category has highest results. The Rotten Tomatoes results which have an Average Audience Score of 88%, where the WandaVision film series gets the Full Popcorn Bucket logo, can be concluded as a positive result. The WandaVision series received positive results.

Series "Stranger Things", based on the experimental method the results were 37% positive, 40.9% neutral and 22% negative and machine learning obtained results of 50% positive, 12% neutral and 38% negative so it had positive reviews because of the category positive has the highest yield. The Rotten Tomatoes results have an Average Audience Score of 90%, and get the Full Popcorn Bucket logo, so it had positive results. It can be concluded three data that series of Stranger Things can still interest to viewers

"Attack on Titan" series, based on experimental methods the results were 52% positive, 24.4% neutral and 23.6% negative and machine learning obtained results of 54% positive, 23% neutral and 23% negative so it had positive reviews because the positive category has the highest results. The Rotten Tomatoes results have an Average Audience Score of 94%, this result has the Full Popcorn Bucket logo, so it can be concluded as positive. It was concluded that Attack on Titan series got positif result.

Series "The Flash", the experimental method obtained results of 52% positive, 21.9% neutral, and 26.1% negative and machine learning obtained results of 52% positive, 20% neutral, and 28% negative so it has positive reviews because of the positive category has the highest yield. The Rotten Tomatoes results which have an Average Audience Score of 59%, the results of the film series The Flash getting the Spilled Popcorn Bucket logo, can be stated in

this result is Negative result. In collecting experimental and machine learning method data on the Twitter API, a number of viewers gave positive results to the series with quite small negative values. However, many viewers on Rotten Tomatoes gave negative comments, making the series The Flash less recommended for viewing.

"One Piece" series, based on the experimental method the results were 46.9% positive, 28% neutral and 25% negative and machine learning obtained results of 48% positive, 27% neutral and 28% negative so it had positive reviews because of the category positive has the highest yield. The Rotten Tomatoes results have an Average Audience Score of 95%, where the results get the Full Popcorn Bucket logo, so it is declared a positive result. It can be concluded that the One Piece series has positive results

"Loki" series, based on the experimental method the results were 45.8% positive, 31% neutral and 23.1% negative and machine learning obtained results of 46% positive, 29% neutral and 25% negative so it has positive reviews because of the positive category has the highest yield. The Rotten Tomatoes results have an Average Audience Score of 85%, where the results of this film series received the Full Popcorn Bucket logo, so it was declared positive. Loki series received high positive results from the three data that were obtained.

From the research results above, it can be proven that the use of the Naïve Bayes machine learning algorithm can provide a big picture of the assessment of film series because of the prediction results. The Naïve Bayes algorithm can provide quite precise accuracy by comparing the experimental methods obtained and the analysis results from Rotten Tomatoes so that it can be calculated so that the Naïve Bayes algorithm can predict well.

Conclusions

As the world of cinema develops, film series themselves have many genres, such as Drama, Comedy, Thriller, Horror, and so on. This makes it difficult to see every recommended series that is being broadcast. Twitter is a place for users to provide real-time criticism or suggestions regarding the world of film that is currently trending or has been broadcast for a

long time, so that they can draw conclusions little by little regarding the assessment of films or series that are currently or will be broadcast. With this, text mining is needed to display data, which has a big influence on the number of fans of film series, this is because reviews provide a variety of information, such as positive, negative or neutral assessments, so that this can influence the influence of audience interest in each film series title.

From the research results, it can be proven that machine learning can process quite a large amount of information data regarding film series with high accuracy results because it is processed using the Naïve Bayes algorithm. Then, the Naïve Bayes algorithm can predict the level of data accuracy that is accurate enough regarding film series so that it can classify sentences on Twitter that are positive, negative and neutral. In a comparison of three methods between experimental methods, rotten tomatoes and machine learning have a fairly accurate level of accuracy, so that the results from machine learning can be used well to conclude film series in data management in the Twitter API. Therefore, it is hoped that the existence of a Film Series Review Analysis website using the Twitter API and Text Mining Model can make it easier for users to carry out analysis or measurement of film series reviews, so that users can find out what film series are being widely discussed and what assessments of old film series are. whether the film series is recommended or not.

References

Ellysinta, Vivia, Wilson Vernando, Kelvin Kurniawan, and Junifer. 2020. "Pengaruh Illegal Movie Streaming Terhadap Popularitas Film Bagi Mahasiswa." *Jurnal Teknologi Informasi* Vol 6 no 1.

Erlianto, Reviansyah, and Hana Faridah. 2022. "PERLINDUNGAN HUKUM PEMBAJAKAN FILM DIGITAL (Studi Perbandingan Hukum Indonesia, Malaysia, dan Korea Selatan)." *AJUDIKASI (Jurnal Ilmu Hukum).*

Firdaus, Fakhriza , and Ali Mukhlis. 2020. "Implementasi Algoritma Naive Bayes Pada Data Set Kualitatif Prediksi Kebangkrutan." *JURIKOM (Jurnal Riset Komputer), Vol. 7 No. 1* 2715-7393.

Isnain, Auliya Rahman , Agus Sihabuddin, and Yohanes Suyanto. 2020. "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection." *IJCCS (Indonesian Journal of Computing and Cybernetics Systems) Vol.14, No.2* 2460-7258.

Krisdiyanto, Taofik, and Erry Maricha Oki Nurharyanto. 2021. "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers." *Jurnal CoreIT* Vol.7, No.1.

Kusuma, Purba Daru. 2020. *Machine Learning Teori, Program dan Studi Kasus.* Yogyakarta: CV Budi Utama.

Londjo, Muammar Farhan . 2021. "IMPLEMENTASI WHITE BOX TESTING DENGAN TEKNIK BASIS PATH PADA PENGUJIAN FORM LOGIN." *Jurnal Siliwangi Vol.7. No.2* 2615-4765.

Nurdin, Asep, Fauziah, and Ratih Titi Komala Sari. 2021. "White Box Testing Pada Sistem Manajemen Pengelolaan Surat Di Sekretariat Rektorat Berbasis Web." *Satuan Tulisan Riset dan Inovasi Teknologi (STRING).*

Ridwan, Achmad. 2020. "Penerapan Algoritma Naive Bayes Untuk Klasifiksi Penyakit Diabetes mellitus." *Jurnal Sistem Komputer dan Kecerdasan Buatan* Volume VI - Nomor 1 .

Rizaldi, Alexander , Evi Maria, Teguh Wahyono, Purwanto, and Kristoko Dwi Hartomo. 2022. "Analisis Penerapan Metode Scrum Pada Pengembangan Sistem Informasi Akuntansi Koperasi." *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 6, Nomor 1* 57-67.

Rizaldi, Alexander , Evi Maria, Teguh Wahyono, Purwanto, and Kristoko Dwi Hartomo. 2022. "Analisis Penerapan Metode Scrum Pada Pengembangan Sistem Informasi Akuntansi Koperasi." *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 6, Nomor 1* 57-67.

Rizaldi, Alexander , Evi Maria, Teguh Wahyono, Purwanto, and Kristoko Dwi Hartomo*. 2022. "Analisis Penerapan Metode Scrum Pada Pengembangan Sistem Informasi Akuntansi Koperasi." *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 6, Nomor 1* 57-67.

Roihan, Ahmad , Po Abas Sunarya, and Ageng Setiani Rafika. 2020. "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper." *IJCIT (Indonesian Journal on Computer and Information Technology) 5 (1)* 75-82.

Saputra, Rahmad Aldi, and Sejati Waluyo. 2022. "Penerapan Algoritma Naive Bayes Dalam Analisis Kenaikan Bahan Bakar Minyak Pada Twitter." *SENAFTI (Seminar Nasional Mahasiswa Fakultas Teknologi Informasi).*

Septianingrum, Fitria , Jajam Haerul Jaman, and Ultach Enri. 2021. "Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan." *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 5, Nomor 4* 1431-1437.

Setyaningtya, Sekar , Bangkit Indarmawan Nugroho, and Zaenul Arif. 2022. "TINJAUAN PUSTAKA SISTEMATIS PADA DATAMINING: STUDI KASUS ALGORITMA K-MEANS CLUSTERING." *Jurnal TEKNOIF Vol. 10 No. 2* 2598-9197.

Sonny, Sonny, and Sestri Novia Rizki. 2021. "PENGEMBANGAN SISTEM PRESENSI KARYAWAN DENGAN TEKNOLOGI GPS BERBASIS WEB PADA PT BPR DANA MAKMUR BATAM." *Jurnal Comasie - VOL. 04 NO. 04* 27156265.

Villarica , Mia V, Francis F Balahadia, Jonardo R Asor, and Gene Marck B Catedrilla. 2022. "Development of Crime Reporting System to Identify Patterns of Crime in Laguna." *International Journal Of Computing Sciences Research.*

Wijaya, Herry Derajad , and Saruni Dwiasnati. 2021. "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat." *JURNAL INFORMATIKA, Vol.7 No.1* 2528-2247.

Christian, Y. (2022). Analisis Performa Akademik Mahasiswa Menggunakan Distributed Random Forest. *Journal of Applied Informatics and Computing (JAIC)*, *6*(2), 180. http://jurnal.polibatam.ac.id/index.php/JAIC