Ceddi Journal of Information System and Technology (JST)
Vol. 2 No. 2 December (2023)
https://doi.org/10.56134/jst.v2i2.18

ISSN: 2829-808X (print)
ISSN: 2829-6575 (online)

Sentiment Analysis Regarding Kanjuruan Stadium Polemics Based on Public Opinion Through Twitter Social Media With SVM Classifier Method

Ghama Wellyandi[1*], Achmad Bayhaqy[2], Chandra[3], Efit Afandi[4], Rimah Abu Achmed[5]

[2-4]*Department of Informatics Technology, Universitas Nusa Mandiri, Jakarta, Indonesia*
[5]*Engineering and Computer Science, University of Sydney, Australia*

Abstract

Football fans are individuals who promote, motivate, and inspire football. Players of football clubs have both positive and negative fanaticism in both the real world and social media, especially on Twitter. Twitter is one of the communication media. Attracting people worldwide, Twitter saw a record increase in global users, with 313 million monthly active users in 2016 alone; the majority accessed Twitter through mobile devices, accounting for 82 percent of users. Due to the multitude of users tweeting, the latest news and comments become significant worldwide. What happens becomes the main topic, and comments received from many users trigger trending topics on Twitter. This research aims to develop a classification model to predict whether tweets from stadium events are positive or negative from fan perspectives. The classification model is based on a Twitter dataset, and sentiment analysis of tweets was conducted using the Support Vector Machine (SVM) algorithm. The next step involved preprocessing, including case-folding, cleansing, translation to English, and sentiment labeling using VADER. Subsequently, in the preprocessing step 2, tokenization, stopwords, and stemming were applied. For modeling, classic algorithms such as Naïve Bayes and Support Vector Machine were used. The highest accuracy, 87.77%, was achieved using the Support Vector Machine (SVM) algorithm.

Keywords: Twitter; Vader; Naïve Bayes; Rapid Minner; Kanjuruhan Stadium

Received: 28 December 2022 Revised: 26 June 2023 Accepted: 29 November 2023

Introduction

Twitter is a social media platform. Attract people all over the world. Emerging from the record increase in Twitter users worldwide, twitter had 313 active users Its monthly Million in 2016 alone most users Access Twitter via mobile devices passed 82 percent of users (Craig & Cunningham, 2019). Due to the large number of users tweeting from users - users then the latest news and comments become a big problem around the world (Wijanarto et al., 2020). This has become the main topic, and comments received from numerous users have triggered a trending topic on Twitter. Tweets may contain Opinions and impressions about the areas of social economy, entertainment, education, sports, and more (Astiningrum et al., 2020); (Mariam, 2023). One of the popular sports in Indonesia is football. In particular, Indonesia became the most Twitter user to comment and opinion on what was going on in the past regarding the riots at the kanjuruhan stadium in the unfortunate city over the Arema Fc vs Persebaya. A total of 135 people were recorded dead, and 583 others injured (Yahya, 2022).

The regular Liga 1 match between Arema and Persebaya Surabaya, two clubs that have long competed in the Super Jawa Timur Derby (Iswinarno, 2022), is scheduled to take place at Kanjuruhan Malang Stadium with a capacity of 42,000 people on October 1st (Alfajri et al., 2022). However, due to security concerns, the police have requested that the match start earlier at 15:30 WIB (08:30 UTC) instead of 20:00 (13:00 UTC), and only allow the presence of 38,000 spectators (Iswinarno, 2022). However, the officials of Liga 1 and the match organizers did not accept this request, resulting in all 42,000 tickets being printed (Suwiknyo, 2022). As a result, there are no tickets available for Persebaya supporters, following the advice of the police (Ismail, 2022). The disaster is the second deadliest in the history of worldwide football, after the 1964 Estadio Nacional tragedy in Peru, which claimed the lives of 328 people (Prasatya, 2022). Consequently, this disaster is the deadliest in Indonesia, Asia, and the Eastern Hemisphere (Febrianto & S,

*Corresponding author.
E-mail address: ghamawellyandi@gmail.com (Ghama Wellyandi)

2022). In the Opinions and impressions of supporters are diverse comments as the events unfold. Based on the explanation above, in this study, a Twitter sentiment analysis was carried out to classify opinion tweets and public comments (Valle-Cruz et al., 2020). The data will be processed using the SVM algorithm and generate tweets that have been classified as positive or negative (Iskandar et al., 2022; Suryandari et al., 2022). Of the 2 algorithms tested, namely Naïve Bayes and SVM as comparison SVM is the algorithm with the highest accuracy.

Method

In this study, the use of data mining methodology CRISP-DM as a common troubleshooter for business and research (Wibawa et al., 2018);(Iskandar et al., 2018). The methodology consists of six stages that is (i) Business Understanding, (ii) Data Understanding, (iii) Data Preparation, (iv) Modeling, (v) Evaluation, (vi) Deployment (Hasanah, Soim, & Handayani, 2021), Shown in Figure 1.

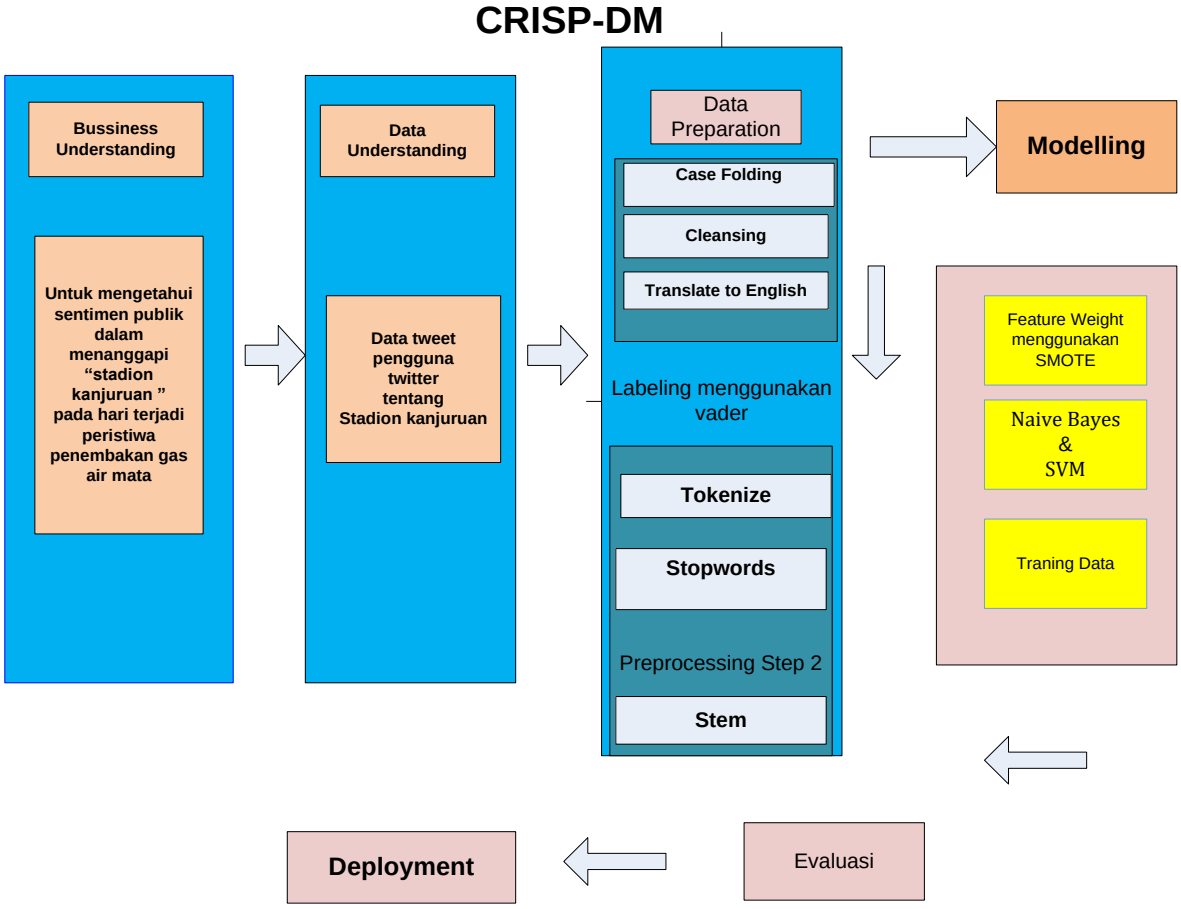

Figure 1. CRISP-DM of Process

1. Business Understanding

   This stage is to explain about what research will be done. And limitations of the problem to be in solve with python programs (Suyitno et al., 2017). In this case to know Public sentiment in response Kanjuruan Stadium Event on the day it happened Preaching tear gas firing (Silvia, 2022).

2. Data understanding

   The data comprehension stage begins with collection of preliminary data and results of activities in order to familiarize yourself with data, to identify problems data quality, for find the first insight into the data or to detect

a subset of interest for forming hypotheses for information that Hidden (Sabariah1 et al., 2021). In this case, the data collection technique uses Twitter data.

3. Data preparation

   The data preparation stage includes all Activities to build the final dataset (data to be entered into modeling) from the initial raw data (Garcia et al., 2015). Assignment Preparatory data is likely to be carried out multiple times and unpredictable Number (Pyle, 1999). Task results include tables, records and attribute selection and transformation and data cleansing for modeling. in this case the data preparation technique includes as (i) Case Folding, (ii) Cleansing, (iii) Tokenize, (iv) Stopwords, (v) Stem (Sari et al., 2019).

4. Modelling

   Modeling techniques selectable and applicable and parameters Model calibrated with values that Optimal. Typically, there are several data techniques mining that can solve the problem the same (Chatfield, 1995). Some techniques have Certain requirements on the formdata. By Therefore, stepping back to the stage Data preparation is often done (Hasanah, Soim, & Handayani, 2021). At this stage, it directly involves Machine Learning to determine data mining techniques, tools used using google colab and other additional features with a rapid minner as the result of the accuracy value later and the rest there is a program language that uses Python as well (Munawar et al., 2023). Then the data algorithm This research uses SVM and naïve bayes classification models as the highest accuracy comparison. selected algorithm Feature Weights using SMOTE (Safitri & Muslim, 2020).

5. Evaluation

   After the classification pattern is obtained in the SMOTE+ Cross Validation algorithm Furthermore, the evaluation stage of the algorithm comparison is carried out with the parameters used is the Confusion Matrix which is basically to provide information on the comparison of the results that have been carried out by the model with the actual classification results by looking at the accuracy (Bhakuni et al., 2022), Class Recall (Ariandi et al., 2023).

6. Deployment

   After the evaluation stage where to assess in detail the results from a model, the implementation of the the whole model that has been built (Kruchten, 2004). In addition, it is also adjustments are made to the model so that it can produce a result that matches the initial target this CRISP-DM stage. For this reason type with values Evaluation the best Into Application (Hasanah, Soim, Handayani, et al., 2021).

Results and Discussion

Result

The results of the research on Sentiment Analysis Regarding Kanjuruan Stadium Polemics Based on Public Opinion Through Twitter Social Media with the SVM Classification Method are as follows:

1. Valence-aware dictionary and Sentiment Reasoner (VADER) is a method used as a model for sentiment analysis and can determine the diversity of data through the intensity of emotional power that exists according to the available Lexicon data dictionary (Abimanyu, 2022). VADER was introduced in 2014 by C.J Hutto and Eric Gilbert whose formation method is based on a human-centric approach, combining qualitative analysis and empirical validation using human wisdom and judgment (Mustaqim et al., 2020).

   The lexicon dictionary can be used to assess the sentiment of phrases and sentences, without the need to look at the others. Sentiment can be categorized - such as {negative, neutral, positive} - or can be numerical - such as intensity range or score (Taboada et al., 2011). The lexical approach looks at the sentiment category or score of each word in a sentence and decides on the category or sentiment score of that whole sentence (Abimanyu et al., 2022). Labeling results using VADER, Shown in Figures 2, 3, and Table 1:

Total Positif	:	237
Total Negetive	:	658
Total Neutral	:	288
Total Data	:	1188

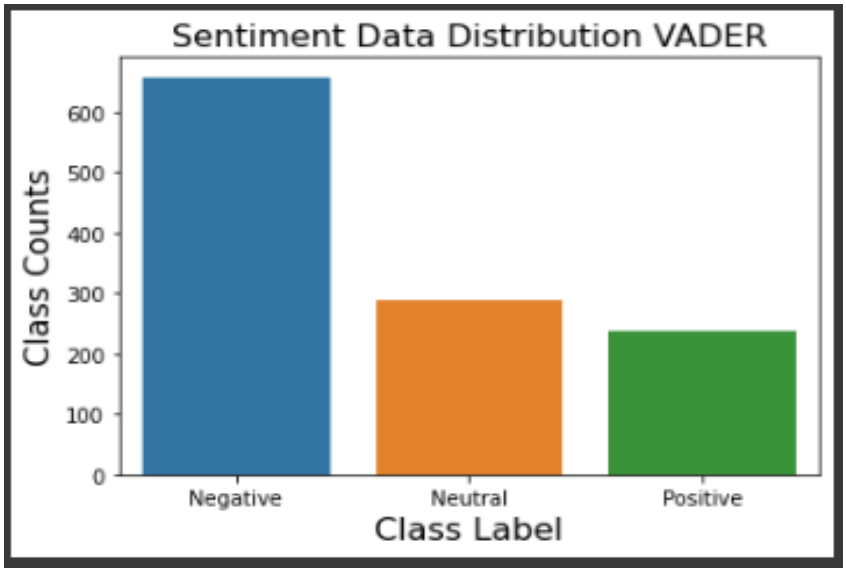

Figure 2. Labeling results using VADER

In this case, because the data taken is only positive and negative data, we will eliminate it first, so from the data taken from Twitter as many as 1183 for positive as many as 237, and negative as 658.

Total Positif	:	237
Total Negetive	:	658
Total Data	:	1188

Figure 3. Labeling results using VADER positive and negative

Table 1. Examples process-cleansing tweets

Index	Tweet	Translate	Sentiment	Cleansing_Tweet
0	The Ministry of Public Works and Public Housing will improve the security of the stadium buildings to make them more comfortable.https://t.co/353hORc4bF read also news from other media in Indonesia at https://t.co/GkwQxRUMd1 #stadionkanjuruhan #tragedikanjuruhan #kementerianpupr #vivabola #kanjuruhan #stadion	The PUPR Ministry will improve the safety of the stadium building to make it more comfortable https://t.co/353horc4bf also read news from other media in Indonesia at https://t.co/gkwqxrumd1 #stadionkankuruhan #tragedikanjuruhan #kembenianpupr #vivabola #kanjuruhan #stadion	Positive	"The Ministry of Public Works and Public Housing (PUPR) will improve the security of the stadium buildings to make them more comfortable. Also, read news from other media in Indonesia about the Kanjuruhan Stadium tragedy. The Kanjuruhan tragedy involves the Ministry of PUPR, Viva Bola, and the Kanjuruhan Stadium."
1	The police chief's classmates were unaware of FIFA's ban on tear gas. How come...????? Naturally, many people question the professionalism of the officials of this institution. #Arema #GasAirMata #TragediKanjuruhan #Kanjuruhan #KanjuruhanDisaster #Malang	The class of the police chief did not know the prohibition of FIFA for tear gas. How come...????? Naturally, if many people question the professionalism of this institutional official. #Arema #GasaIRMATA #TRAGEDIKANJURUHAN #KANJURUHAN #KANJURUHANDISASTER #Malang	Negative	In the same class as the police chief, not knowing the FIFA prohibition on tear gas is quite surprising, and, understandably, many people are questioning the professionalism of this institutional official. The Arema Gasairmata tragedy at Kanjuruhan, Malang, is indeed a disaster.

2. Based on Figure 4, it is a confusion matrix model in a rapid-miner application using the SVM algorithm.

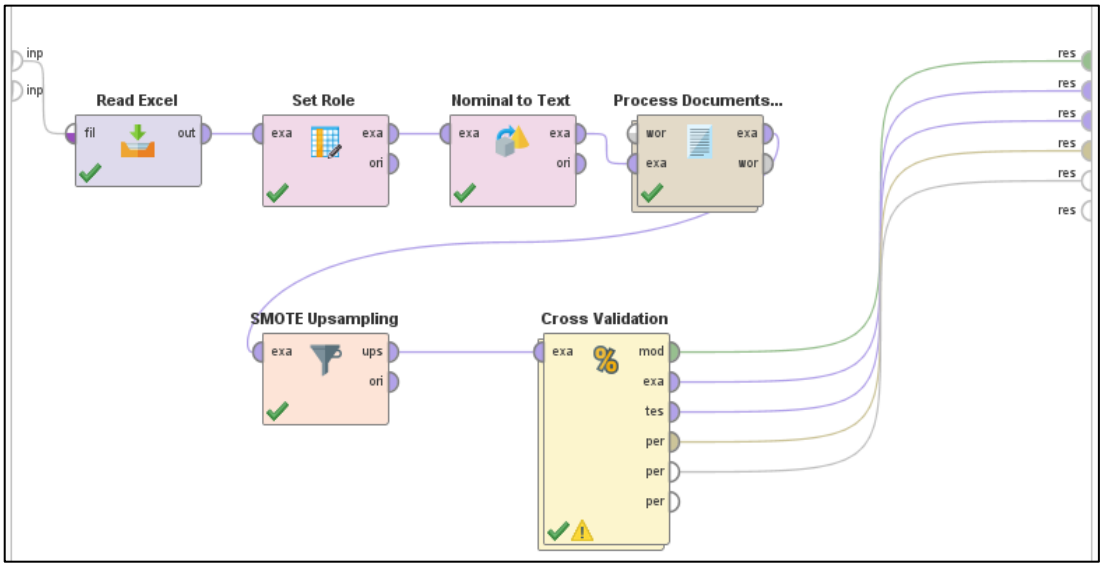

Figure 4. Proses Model

3. Shown in Figure 4, Text processing is done in the rapid miner application. The results of testing tweet data about Kanjuruhan stadiums from a total of positive and negative 895 tweet data using the Support Vector Machine (SVM) + SMOTE and Naïve Bayes + SMOTE methods as in the following table.

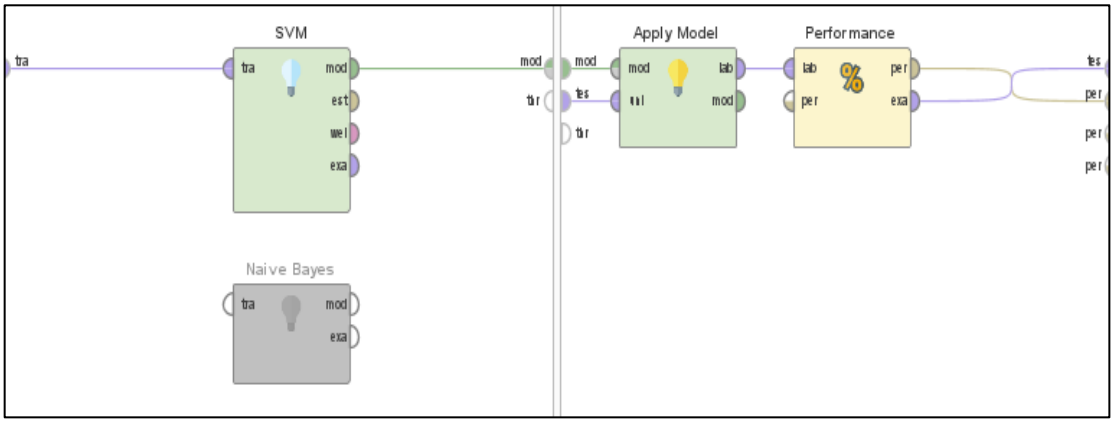

Figure 5. SVM and Naive Bayes algorithms

4. Table 2 explained that the accuracy result for the SVM method with the SMOTE technique was 87.77%, with a positive precision prediction class of 81.56% and a negative precision prediction class of 96.98%. The class recall true is positive at 97.57% and the class recall true is negative at 77.96%. The resulting tweet data shows that the classification for tweets is as predicted by Positive 642 data. The positive tweet data referred to in the negative prediction was 16 data. Negative tweet data with positive predictions of 145 data. Negative tweet data with negative predictions of 513 data.

Table 2. Confusion Matrix Metode SVM + SMOTE

Accuracy: 87.77% +/-2.97%			
	True Positive	True Negative	Class Precision
Positif	642	145	81.56%
Negatif	16	513	96.98%
Class Recall	97.57%	77.96%	

5. Table 3 explained that the accuracy result for the Naïve Bayes Method with the SMOTE technique was: 55.92%, with a positive precision prediction class of 53.16% and a negative precision prediction class of 97.56%. The class recall true is positive at 99.70% and the class recall true is negative at 12.16%. The resulting tweet data shows that the classification for tweets is as predicted by Positive 656 data. The positive tweet data referred to in the negative prediction was 2 data. Negative tweet data with positive predictions of 578 data. Negative tweet data with negative predictions of 80 data.

Tabl3 3. Confusion Matrix Metode Naïve Bayes + SMOTE

Accuracy: 55.92% +/-2.25%			
	True Positve	True Negative	Class Precision
Positif	656	578	53.16%
Negatif	2	80	97.56%
Class Recall	99.70%	12.16%	

It can be seen from the table above that the highest accuracy can be obtained by using the svm algorithm with an accuracy value of 87.77%.

Discussion

From the results above, it can be described from the beginning the results of data collection from twitter with a total of 1183 tweets from that many tweets were classified with positive, neutral, and negative labeling of 237 positive 658 negative and the remaining neutral 288. Then modeling with the naïve Bayes and svm algorithms from the modeling results in can be the highest accuracy value at 87.77% using the svm algorithm and for 55.92% naïve Bayes.

Conclusions and Suggestions

Conclusions

After conducting a series of research and experiments on analysts' sentiments regarding public opinions about events at Kanjuruan Stadium using the Support Vector Machine (SVM) algorithm, it can be concluded that the utilization of SVM achieves a superior level of accuracy. This algorithm precisely identifies and classifies sentiments, providing a deep understanding of societal responses. These findings have significant implications for understanding public dynamics, offering a robust foundation for decision-makers in the planning and execution of various activities at the stadium, and supporting the development of more effective and responsive strategies.

Suggestions

There are a few suggestions that can be given to develop this research, the advice given is as follows**:**

1. The data used can be added a lot so that the data of each class can have the same amount of data.
2. In solving problems SMOTE can be used to find out the influence of this concept against the degree of accuracy.

Acknowledgments

Thanks to Allah SWT, parents, wife, and family who pray so that this research can be completed on time.

References

Abimanyu, D. (2022). Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER. In *Jurnal Nasional Komputasi dan Teknologi Informasi* (Vol. 5, Issue 03). Jurnal Nasional Komputasi dan Teknologi Informasi.

Abimanyu, D., Budianita, E., Cynthia, E. P., Yanto, F., & Yuasra. (2022). Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, *5*(3), 423–431.

Alfajri, I., Aritonang, D. D., Sarwidaningrum, I., Werdiono, D., Irawati, D., & Hidayat, A. R. (2022). *Polda Jatim Bantah Kandungan Gas Air Mata Mematikan*. https://www.kompas.id/baca/investigasi/2022/11/10/polri-bantah-kandungan-zat-air-mata-mematikan

Ariandi, R., Pratiwi, O. N., & Fa'rifah, R. Y. (2023). Klasifikasi Soal Sejarah Tingkat SMA Berdasarkan Level Kognitif Revised Bloom's Taxonomy Menggunakan Algoritma K-Nearest Neighbour Manhattan. *EProceedings of Engineering*, *10*(2).

Astiningrum, M., Haniah, M., & Pradana, Y. rahmat yoga. (2020). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. *Seminar Informatika Aplikatif Polinema (Siap)*, 35— 39.

Bhakuni, M., Kumar, K., Iwendi, C., Singh, A., & others. (2022). Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment Analysis. *Journal of Sensors*, *2022*.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *158*(3), 419–444.

Craig, D., & Cunningham, S. (2019). *Social media entertainment: The new intersection of Hollywood and Silicon Valley*. NYU Press.

Febrianto, V., & S, R. (2022). Death count in Kanjuruhan tragedy climbs to 135. *ANTARA News*. https://en.antaranews.com/news/256465/death-count-in-kanjuruhan-tragedy-climbs-to-135

Garc\'\ia, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer.

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, *5*(2), 103–108. https://doi.org/10.30871/jaic.v5i2.3200

Hasanah, M. A., Soim, S., Handayani, A. S., & others. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing* (Vol. 5, Issue 2).

Iskandar, A., Fahlepi Tuasamu, M. R., Syamsu, S., Mansyur, M., Listyorini, T., Sallu, S., Supriyono, S., Saddhono, K., Napitupulu, D., & Rahim, R. (2018). Web based testing application security system using semantic comparison method. *IOP Conference Series: Materials Science and Engineering*, *420*(1). https://doi.org/10.1088/1757-899X/420/1/012122

Iskandar, A., Rahim, R., Matturungan, H., & others. (2022). Web-based STMIK AKBA Student Attendance Information System by Making QR Codes an Auxiliary Medium. *Ceddi Journal of Information System and Technology (JST)*, *1*(2), 24–29. https://doi.org/https://doi.org/10.56134/jst.v1i2.22

Ismail, H. C. (2022). Tragedi Kanjuruhan, Polri Akui Gunakan Gas Air Mata Kedaluwarsa. *Tempo.Co*. https://nasional.tempo.co/read/1643703/tragedi-kanjuruhan-polri-akui-gunakan-gas-air-mata-kedaluwarsa

Iswinarno, C. (2022). BRIN Periksa Gas Air Mata yang Picu Peristiwa Berdarah Tragedi Kanjuruhan. *Suara.Com*. https://www.suara.com/news/2022/10/14/183417/brin-periksa-gas-air-mata-yang-picu-peristiwa-berdarah-tragedi-kanjuruhan

Kruchten, P. (2004). *The rational unified process: an introduction*. Addison-Wesley Professional.

Mariam, M. (2023). Application of Customer Relationship Management in Maintaining Customer Loyalty (Case Study Hotel Melati). *Ceddi Journal of Information System and Technology (JST)*, *2*(1), 1–8. https://doi.org/https://doi.org/10.56134/jst.v2i1.31

Munawar, Z., Muliantara, A., Kmurawak, R. M. B., Reba, F., Sroyer, A., Sukmawan, D., Rahman, A., Insany, G. P., Mandowen, S. A., Toyib, W., & others. (2023). *Big Data Analytics: Konsep, Implementasi, dan Aplikasi Terkini*. Kaizen Media Publishing.

Mustaqim, T., Umam, K., & Muslim, M. A. (2020). Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. In *Conference Series* (Vol. 1567, Issue 3).

Prasatya, R. (2022). Tok! PSSI Putuskan Secepatnya Gelar KLB. *Detikcom*. https://sport.detik.com/sepakbola/liga-indonesia/d-6375792/tok-pssi-putuskan-secepatnya-gelar-klb

Pyle, D. (1999). *Data preparation for data mining*. morgan kaufmann.

Sabariah1, M. K., Adam Mukharil Bachtiar2, Dharmayanti3, D., & Perdana4, I. (2021). *BUSINESS DAN DATA UNDERSTANDING DALAM RANGKA PEMBENTUKAN RANGKA MENINGKATKAN LABA PENJUALAN MENGGUNAKAN METODE Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) BUSINESS DAN DATA UNDERSTANDING DALAM RANGKA PEMBENTUKAN MODEL TATA LETAK DAN TATA RUANG PASA*. *January*. https://doi.org/10.34010/komputa.v1i2.61

Safitri, A. R., & Muslim, M. A. (2020). Improved accuracy of naive bayes classifier for determination of customer churn uses smote and genetic algorithms. In *Journal of Soft Computing Exploration* (Vol. 1, Issue 1).

Sari, E. Y., Wierfi, A. D., & Setyanto, A. (2019). Sentiment Analysis of Customer Satisfaction on Transportation Network Company Using Naive Bayes Classifier. *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 1–6.

Silvia. (2022, October 11). *Komnas HAM Sebut Gas Air Mata Jadi Penyebab Utama Tragedi Kanjuruhan*. DetikNews. https://news.detik.com/berita/d-6342591/komnas-ham-sebut-gas-air-mata-jadi-penyebab-utama-tragedi-kanjuruhan

Suryandari, N., Giovani, D., & Madura, U. T. (2022). *EDUKATIF : JURNAL ILMU PENDIDIKAN*. *4*(3), 4154–4160.

Suwiknyo, E. (2022). Tim Pencari Fakta Serahkan Sampel Gas Air Mata Kedaluwarsa ke BRIN. *Bisnis.Com*. https://kabar24.bisnis.com/read/20221010/16/1586109/tim-pencari-fakta-serahkan-sampel-gas-air-mata-kedaluwarsa-ke-brin

Suyitno, P. P. W., Indrajit, R. E., & Fauzi, M. (2017). PENERAPAN DATA MINING DALAM MENANGANI KEMACETAN DI JAKARTA Popy. *Ikraith-Informatika*, *1*(2), 53–60.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. In *Computational linguistics* (Vol. 37, Issue 2). MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info~….

Valle-Cruz, D., López-Chau, A., & Sandoval-Almazán, R. (2020). Impression analysis of trending topics in Twitter with classification algorithms. *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, 430–441.

Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, *3*(1).

Wijanarto, W., Sari, A. P., & Rohmani, A. (2020). Tuning Model Analisis Sentimen Tweeter Sepakbola Pada Dataset Kecil dan Seimbang. *JOINS (Journal of Information System)*, *5*(1), 44–61. https://doi.org/10.33633/joins.v5i1.3275

Yahya, A. N. (2022). TGIPF Kanjuruhan: Sepatutnya Ketua Umum PSSI dan Jajaran Komite Eksekutif Mengundurkan Diri. *Kompas.Com*. https://nasional.kompas.com/read/2022/10/14/15573851/tgipf-kanjuruhan-sepatutnya-ketua-umum-pssi-dan-jajaran-komite-eksekutif