JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALYSIS OF PUBLIC OPINION ON INDONESIAN TELEVISION SHOWS USING SUPPORT VECTOR MACHINE

Fidya Farasalsabila[1*], Ema Utami [1], Muhammad Hanafi[1]
[1]Master of Informatics, Universitas Amikom Yogyakarta
Email: fidya@students.amikom.ac.id

Abstract: There are a great number of academics that are now conducting research on sentiment analysis by employing supervised and machine learning techniques. The research can be carried out with the assistance of a variety of sources, including reviews of movies, reviews of Twitter, reviews of online products, blogs, discussion forums, and other social networks. With the progress of technology, individuals may now effortlessly utilize social media platforms to access and share information, as well as express their viewpoints to the general public, without any constraints of distance or time. Twitter is a social media network that serves as a repository for opinions. Diverse techniques are employed to provide optimal and realistically precise pressure detection. The analysis and discussion affirm that the Support Vector Machine (SVM) was effectively employed in this study, utilizing public opinion data on television program reviews in Indonesia. An SVM classifier is employed to examine the Twitter data set by utilizing various parameters. The study successfully completed the preprocessing process by collecting a total of 400 data points, consisting of 320 reviews from 4 television shows for training data and 80 reviews for testing. The data was filtered and classified using SVM, with 200 positive and 200 negative data points for comparison. The experiment utilized the SVM method using TF-IDF to achieve the most accurate test results. The test accuracy was 80%, while the training data accuracy reached 100%.

Keywords: Sentiment Analysis; Support Vector Machine; Television Shows Review, TF-IDF,

Abstrak: Saat ini, banyak akademisi sedang menyelidiki analisis sentimen melalui pemanfaatan teknik yang diawasi dan pembelajaran mesin. Kajian dapat dilakukan dengan menggunakan beberapa sumber seperti review film, review Twitter, review produk online, blog, forum diskusi, atau jejaring sosial lainnya. Dengan kemajuan teknologi, masyarakat kini dapat dengan mudah memanfaatkan platform media sosial untuk mengakses dan berbagi informasi, serta menyampaikan pandangan mereka kepada masyarakat umum, tanpa batasan jarak dan waktu. Twitter adalah jaringan media sosial yang berfungsi sebagai gudang opini. Beragam teknik digunakan untuk menghasilkan deteksi tekanan yang optimal dan presisi secara realistis. Analisis dan pembahasan menegaskan bahwa Support Vector Machine (SVM) efektif digunakan dalam penelitian ini, memanfaatkan data opini publik tentang review program televisi di Indonesia. Pengklasifikasi SVM digunakan untuk memeriksa kumpulan data Twitter dengan memanfaatkan berbagai parameter. Penelitian berhasil menyelesaikan proses preprocessing dengan mengumpulkan total 400 titik data yang terdiri dari 320 review dari 4 acara televisi untuk data pelatihan dan 80 review untuk pengujian. Data disaring dan diklasifikasikan menggunakan SVM, dengan 200 titik data positif dan 200 titik data negatif sebagai perbandingan. Percobaan ini menggunakan metode SVM dengan menggunakan TF-IDF untuk mencapai hasil pengujian yang paling akurat. Akurasi pengujiannya mencapai 80%, sedangkan akurasi data pelatihan mencapai 100%.

Kata kunci: Analisis Sentimen, Review Tayangan Televisi, TF-IDF, Support Vector Machine.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Television is an electronic medium that offers both entertainment and knowledge through television programs to its viewers. To determine television show ratings, one can analyze the popularity of programs among the general audience. Television viewers often express their opinions or remarks about their preferred shows on social media channels like Twitter [1]. The viewpoint is expressed as a tweet, which will then be featured as news on the Twitter timeline. The significance of Twitter public sentiment towards television shows lies in its potential for doing sentiment analysis to forecast individuals' assessment of a TV program, whether it is favorable or negative [2].

Several text mining techniques, such as Lexicon Based, Support Vector Machine, K-Nearest Neighbor, and Nave Bayes classifier, can be employed for sentiment analysis. The Support Vector Machine (SVM) method was selected for text mining due to its superior accuracy compared to other methods, The effectiveness of the K-Nearest Neighbour algorithm, the Naive Bayes algorithm classifier, and the Support Vector Machine was investigated in a study [3] that investigated the effectiveness of these three algorithms. Then, in subsequent research [4], obtain accuracy results of 98.33% using the SVM method, and [5] obtain the highest accuracy score by SVM of 89.70%. Based on the three studies dis-cussed above, the SVM method has the highest accuracy value.

Study on Public Sentiment Analysis on Twitter Regarding the Implementation of Simultaneous PILKADA Using Support Vector Machine Algorithm [6]. The classification procedure yields two distinct categories of tweets, good and negative. The classification method achieves a superior accuracy rate of 91%, whereas the l-means clustering method only achieves 82% accuracy. Twitter Implementation Sentiment Analysis For Review Movies [7] research Vector Engines Benefit from the Use of Algorithms. With the classification process using Algorithm Support Vector Machine, it is easy to see positive, negative, or neutral opinion, which is higher than the Naive Bayes algorithm only until 75% is the classifier stable. According to [8], the Naive Bayes algorithm was employed to assess the sentiment on Twitter on the effectiveness of the Corruption Eradication Commission (KPK) in Indonesia. With the aim of resolving concerns regarding KPK performance using Twitter data.

Based on the findings of the pre-ceding studies, the SVM method has the highest accuracy value. Furthermore, study found that using term weights with a more efficient training process and classification functions can improve accuracy [9]. The goal of weighting words (terms) is to give equal weight to each word (term) contained in the to-be-processed text document [10]. The TF-IDF method was then used to improve accuracy. The goal of this study is to use SVM to determine the pattern of public opinion regarding television broadcasts in Indonesia, and to use TF-IDF to determine the results of increasing accuracy.

METHOD

Support Vector Machine is the model that will be used in this research.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Prior to incorporating this paradigm into research, there are multiple essential tasks that must be completed. Subsequently, the review text is processed in order to be suitable for the trained model. The discourse that is offered in this section will make use of the methodology that was utilised for this research investigation.

Preposed Model

The model suggested in this study is SVM. Prior to incorporating this paradigm into research, a number of sequential tasks must be completed.

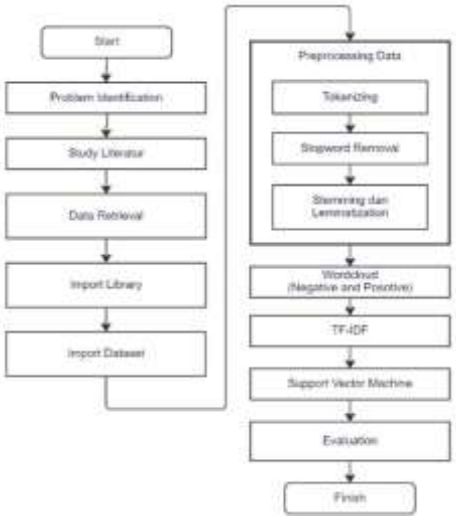

Image 1. Research Flow

Sentiment Analysis

Sentiment analysis, or opinion mining, is the study of how individuals articulate their emotions on a specific object or attribute through written language [9]. The objective of sentiment analysis is to create automated technologies that are able to extract sentiment and other subjective information from text and natural language. Contrary to what most people believe, an opinion can be defined as a cognitive perspective, judgment, or appraisal of a specific subject matter [12].

Television Shows

Television shows can reach a large audience. Experiments with television broadcast began in the late 1920s and early 1930s. Television is a visual broadcasting medium. Television is derived from the words tele and vision, which mean far (tele) and visible (vision), respectively, so television refers to viewing from a distance. Television is compared to the invention of the wheel in that it has the potential to change world civilization [13].

Datasets

The television show dataset [14] is used, serving as source material for this research. The dataset consists of 400 tweet documents from 4 television shows, namely Hitam Putih TransTV, Indonesia Lawyers Club TvOne, Kick Andy MetroTV, and Mata Najwa MetroTV. There were a total of 400 data points that were utilized, with 200 positive and 200 negative data points being labeled respectively.

Preprocessing

Many different methodologies are essential components in the field of text processing [15].

Tokenization

The term "tokenization" refers to the process of determining which words are contained inside a series of characters that are supplied. Although the separation of punctuation marks is the primary manner by which this procedure is completed, it also entails the identification of abbreviations and the utilisation of other techniques [15]. In the event that the content was obtained from a webpage, the tokenization procedure might also involve the utilization of normalization

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

strategies, such as the removal of HTML tags, the implementation of truecasing, or the conversion to lowercase [16][17].

Stop Word Removal

Words that are frequently used in language are known as stop words. Some examples of stop words include subjects, affixes, pronouns, determinants, prepositions, and conjunctions. In the course of this procedure, the text document will be cleaned of any words that are not absolutely necessary [15].

Stemming

At its most fundamental level, stemming refers to the process of removing affixes from a particular word in order to extract the root, or base form, of the word. This root is often shared by all words that are connected to the word in question [15], [16] .

Lemmatization

An alternative to stemming, lemmatization is a technique that reduces the inflectional form of a word to its root form. There are several applications for lemmatization. While stemming is used to generate authentic word forms, lemmatization is used to generate authentic word forms that correspond to the base forms of words that are listed in dictionaries [15]. The process of lemmatization, despite the fact that it has the benefit of making the output more intelligible, is a procedure that demands a greater amount of computer power than other methods [16]

Term Frequency - Inverse Document Frequency (TF-IDF)

The formula known as TF-IDF [18] combines the ideas of term frequency (TF) and inverse document frequency (IDF), which indicate the frequency of documents. "Inverse Document Frequency" is what "IDF" stands for in its full form name. In order to tackle this problem, a method called in-verse document frequency (IDF) has been suggested. This method enhances the ability to differentiate between terms in text categorization by considering their coverage frequency [19]. Inverse Document Frequency (IDF) extends beyond the concept of Document Frequency (DF) by considering the actual count of documents that include a specific phrase. According to the assumption, terms that have a lower frequency of appearance in documents are regarded to be more significant than terms that have a higher frequency of appearance [20]. To compute the IDF value of a specific phrase, use the formula 1-3:

$$IDF\ (t,d,D) = log\frac{|D|}{DF(t,D)} \quad (1)$$

$$IDF\ (t,d,D) = log\frac{|D|+1}{DF(t,D)+1} \quad (2)$$

$$TF-IDF\ (t,d,D) = TF(t,d) * IDF(t,d,D) \quad (3)$$

Support Vector Machine

The SVM algorithm, as described in reference [21] is a model used for binary classification. A straight line serves as the most suitable division between two data classes in a two-dimensional space. The distance between the data point that is closest to the decision border should be maximized when attempting to answer a classification problem, according to one of the primary concepts of SVM.

f(x) = wTx + b is the classification function that will be used to represent it. As a consequence of this,

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

the value is comparable to the distance that existed between the two dotted lines, which is denoted by the symbol r̂. This distance is the gap that exists between two dotted lines and 2̃r, in addition to the support vector that is located on the dotted line. Ir is the symbol that is used to denote the geometric distance, and the formula that is used to calculate it is 4:

$$r = yr\frac{\hat{r}}{w} \quad (4)$$

Evaluation Matrix

There is a matrix known as the Confusion Matrix that illustrates how accurate a model is at forecasting a certain scenario. Concerning the effectiveness of a classification model, the confusion matrix is one of the approaches that may be applied to evaluate its performance. To identify models that contain data as either A or B (4), the confusion matrix can be expressed in its most fundamental form as a table that is two by two. This table is used to store the confusion matrix.

Table 1. Table of Confusion Matrix

Actual Class	Predict Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Using test data that has never been seen before, a variety of evaluation strategies are utilized in order to provide an accurate assessment of the classification's effectiveness. Among the measures that are utilized most frequently in text classification are precision, recall, F-measure, and accuracy[22].

The performance of categorization can also be evaluated using accuracy, which is another statistic that is utilized. The number of samples that have been correctly identified is what determines the level of accuracy. The term "accuracy" refers to the degree to which a classification accurately predicts a single class on a consistent basis. Using a mathematical formula, such as equation (5), which can be shown, it is possible to ascertain this information.

$$\text{Accuracy} = 2 \times \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

RESULT AND DISCUSSION

For the purpose of coming up with the best possible results, the authors of this study carried out a number of different implementations employing a wide range of parameters. Next, the feature extraction step, more specifically the TF-IDF, was carried out, after which wordclouds were evaluated both before and after the preprocessing stage. These were the parameters that were examined. Wordcloud is the first indicator that is being evaluated. Examples of words that are included in the word cloud that represents negative sentiment are shown in Image 2.

Image 2. Wordcloud Negative

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 2, Maret 2024, hlm. 239 - 246 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v10i2.2935
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Additionally, the second parameter that was evaluated yielded a positive wordcloud. The words displayed in Figure 4 are derived from the Wordcloud positive word. Subsequently, the method proceeds to the feature extraction stage and assessment matrix.

Image 3. Wordcloud Positive

Evaluation Model

It can be concluded that the implementation of Support Vector Machine (SVM) using data from Television Show Review has been successful in this research. This step consists of filtering and categorizing data using SVM. Of the total 400 data points collected, 320 were used for training purposes, while the remaining 80 were used for testing purposes. Table 2 presents the results of tests conducted as part of this investigation.

Table 2. Result of confusion matrix

Actual Class	**Predict Class**	
	Positive	**Negative**
Positive	40 (TP)	0 (FN)
Negative	16 (FP)	24 (TN)

The findings of the evaluation are provided in table 2, and they reveal that the model properly identifies forty true positives (TP), does not have any false negatives (FN), mistakenly identifies sixteen false positives (FP), and correctly identifies twenty-four true negatives (TN). The research resulted in an accuracy rate of one hundred percent for the training data and an accuracy rate of eighty percent for the test data, as was illustrated in Image 4.

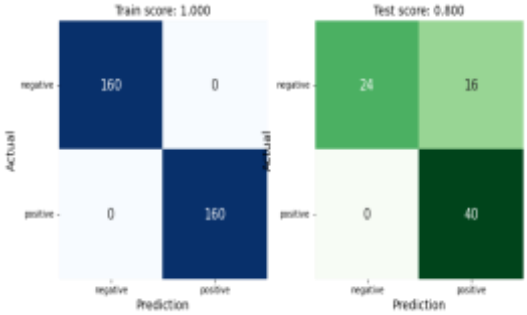

Image 4. Result

CONCLUSION

Based on the findings of the research and the debate, it appears that SVM has been successfully applied to the data of television broadcasts. A successful completion of the preprocessing step, which included filtering and classifying data using support vector machines (SVM), was required for the research undertaking. There were a total of 400 data points that were utilized, with 200 positive and 200 negative data points being labeled respectively. The training data consisted of 80% of the total data, while the remaining 20% was used for testing. Information. The data is evenly split between negative data, accounting for 0.50, and positive data, also accounting for 0.50. SVM (Support Vector Machine) achieves 80% of the highest quality test outcomes. Additional study could involve comparing this model and algorithm to others, as well as

conducting case studies using different social media platforms. Twitter and other social media platforms were included of supplementary analyses in a later study.

BIBLIOGRAPHY

[1] Raghav Mehta and Shikha Gupta, "Movie Recommendation Systems using Sentiment Analysis and Cosine Similarity," *International Journal for Modern Trends in Science and Technology*, vol. 7, no. 01, pp. 16–22, Jan. 2021, doi: 10.46501/ijmtst0701004.

[2] A. Benlahbib and E. H. Nfaoui, "MTVRep: A movie and TV show reputation system based on fine-grained sentiment and semantic analysis," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1613–1626, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1613-1626.

[3] I. Saputra and D. Rosiyadi, "Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp," *Faktor Exacta*, vol. 12, no. 2, p. 101, Jul. 2019, doi: 10.30998/faktorexacta.v12i2.4181.

[4] A. Setiyono and H. F. Pardede, "Klasifikasi Sms Spam Menggunakan Support Vector Machine," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 275–280, Sep. 2019, doi: 10.33480/pilar.v15i2.693.

[5] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *JURNAL INFORMATIKA*, vol. 6, no. 2, pp. 212–218, 2019, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[6] A. Rahmawati, A. Marjuni, J. Zeniarja, J. Informatika, U. Dian, and N. Semarang, "Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksanaan Pilkada Serentak Menggunakan Algoritma Support Vector Machine Public Sentiment Analysis On Twitter Social Media To Pilkada Serentak Event Using Support Vector Machine Algorithm," 2017.

[7] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentiment analysis untuk reviewfilm menggunakan algoritma support vector machine," *Jurnal Informatika Polinema* , vol. 4, 2018.

[8] R. Taufiq, A. E. Wardoyo, and R. Pratama, "Analisis Sentimen Pada Twitter Terhadap Kinerja Komisi Pemberantasan Korupsi (Kpk) Di Indonesia Dengan Metode Naive Bayes."

[9] S. Al Faraby, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification indonesian news," 2018.

[10] R. S. Perdana and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[11] S. Bhatia, M. Sharma, and K. K. Bhatia, "Sentiment Analysis and Mining of Opinions," in *Studies in Big Data*, vol. 30, Springer Science and Business Media Deutschland GmbH, 2018, pp. 503–523. doi: 10.1007/978-3-319-60435-0_20.

[12] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment analysis in social networks*. 2017.

[13] A. Halik, S. Sos, and M. Si, "Buku Daras Uin Alauddin Komunikasi Massa Universitas Islam Negeri (Uin) Alauddin Makassar," 2013.

[14] W. E. Nurjanah, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[15] G. Ignatow and R. Mihalcea, "An Introduction to Text Mining," 2018.

[16] D. D. Nur Cahyo *et al.*, "Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 90–95, Mar. 2023, doi: 10.25139/inform.v8i2.5700.

[17] R. Friedman, "Tokenization in the Theory of Knowledge," *Encyclopedia*, vol. 3, no. 1, pp. 380–386, Mar. 2023, doi: 10.3390/encyclopedia3010024.

[18] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, "Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/6619088.

[19] M. Lan, C. L. Tan, J. Su, and Y. Lu, "Supervised and traditional term weighting methods for automatic text categorization," *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 4, pp. 721–735, 2009, doi: 10.1109/TPAMI.2008.110.

[20] T. Sabbah *et al.*, "Modified frequency-based term weighting schemes for text classification," *Appl Soft Comput*, vol. 58, pp. 193–206, Sep. 2017, doi: 10.1016/j.asoc.2017.04.069.

[21] W. S. Noble, "What is a support vector machine?," 2006. [Online]. Available: http://www.nature.com/naturebiotechnology

[22] E. Beauxis-Aussalet, "Simplifying the Visualization of Confusion Matrix," 2014. [Online]. Available: https://www.researchgate.net/publication/302412429

[23] D. D. Nur Cahyo *et al.*, "Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 90–95, Mar. 2023, doi: 10.25139/inform.v8i2.5700.

[24] N. Ritha *et al.*, "Sentiment Analysis of Health Protocol Policy Using K-Nearest Neighbor and Cosine Similarity," European Alliance for Innovation n.o., Jan. 2023. doi: 10.4108/eai.11-10-2022.2326274.