JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALYSIS OF PUBLIC OPINION SENTIMENT REGARDING POLICE INSTITUTIONS BASED ON TWITTER USING THE SUPPORT VECTOR MACHINE (SVM) METHOD

Said Ahmad Sirojudin[1*], Try Susanti[1], Mhd. Theo Ari Bangsa[1]
[1]Sistem Informasi, Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi
email: *ahmadsirojudin03082019@gmail.com

Abstract: Twitter occupies the top position of the most popular social media platform in Indonesia. Police and other related issues were the subject of much discussion. The aim of this research is to analyze public sentiment towards the National Police Agency using Twitter with the support vector machine method. The research started by crawling Twitter data. The data contains a total of 6,925 entries for three keywords. Next, we move on to the preprocessing stage consisting of (cleaning, case folding, tokenization, and filtering). Next is the tf-idf feature extraction stage, finally the classification and evaluation stage. The results of manual data inspection (73:27) showed accuracy of 70.66%, precision of 70.68%, and recall of 99.76%. Testing the second data (82:18), found accuracy 86%, precision 86.21%, recall 99.71%. The results of manual data checking (82:18) showed accuracy of 70.66%, precision of 70.68%, recall of 99.76%. Testing the second data (82:18), found accuracy 86%, precision 86.21%, recall 99.71%. From the data system testing results (80:20), accuracy was 87.55%, positive precision 87.53%, negative precision 88.24%, positive recall 99.48%, and negative recall. the rate is 99.48.% – The result is 21.43%. Data testing results (60:40) showed accuracy of 86.89%, positive precision of 86.84%, negative precision of 88.46%, positive recall of 99.61%, and negative recall of 16.43%. Single test data validation system (80:20), accuracy 87.55, overall test cross validation system (k fold 5 accuracy) 86.673%.

Keywords: data mining;police agencies;support vector machines

Abstrak: Twitter menduduki posisi teratas platform media sosial terpopuler di Indonesia. Polisi dan masalah terkait lainnya menjadi pokok bahasan banyak pembicaraan. Tujuan penelitian ini untuk menganalisis sentimen masyarakat terhadap Badan Kepolisian Nasional menggunakan Twitter dengan metode support vector machine. Penelitian dimulai dengan crawling data Twitter. Data memuat total 6.925 entri dari tiga kata kunci. Selanjutnya beralih ke tahap preprocessing terdiri dari (pembersihan, pelipatan kasus, tokenisasi, dan pemfilteran). Selanjutnya tahap ekstraksi fitur tf-idf, terakhir tahap klasifikasi dan evaluasi. Hasil pemeriksaan data manual (73:27) menunjukkan akurasi 70,66%, presisi 70,68%, dan recall 99,76%. Menguji data kedua (82:18), menemukan akurasi 86%, presisi 86,21%, recall 99,71%. Hasil pemeriksaan data secara manual (82:18) menunjukkan akurasi 70,66%, presisi 70,68%, recall 99,76%. Menguji data kedua (82:18), menemukan akurasi 86%, presisi 86,21%, recall 99,71%. Dari hasil pengujian sistem data (80:20), akurasi 87,55%, presisi positif 87,53%, presisi negatif 88,24%, recall positif 99,48%, dan recall negatif. tarifnya adalah 99,48.% – Hasilnya 21,43%. Hasil pengujian data (60:40) menunjukkan akurasi 86,89%, presisi positif 86,84%, presisi negatif 88,46%, recall positif 99,61%, dan recall negatif 16,43%. Uji tunggal sistem validasi data (80:20), akurasi 87,55, uji keseluruhan sistem validasi silang (akurasi k fold 5) 86,673%.

Kata Kunci: data mining;instansi kepolisian;mesin vektor pendukung

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The Republic of Indonesia National Police (Polri) is a state apparatus that maintains public order and security. Its mission is to protect, provide services and enforce the law for the Indonesian people[1] The Indonesian state is based on the rule of law, not just power. Indonesia embodies the rule of law as an ideology for security, order, justice and welfare for its citizens.

The police are two elements that are interconnected with the community. It will not run smoothly and productively without the police, without the commu nity there is no police. This fact places the Police in a dual role by fulfilling its duties, as a social worker and law enforcer in the community (service and devotion) and social fields[2]

Public opinion on police performa nce can be either positive or negative. Public trust in law enforcement in this country is decreasing day by day. Public dissatisfaction with law enforcement is increasing in this country, as clearly seen from the results of a survey by institu tions in Indonesia, where public distrust of law enforcement currently reaches 56%, and only 29.8% reported it. They are satisfied, compared to the previous government, only 22.6% stated that the police are better under the current government than under the previous government[3]

Social media is a community platf orm that focuses on user presence and encourages user activity and collabora tion. Therefore, it can be said that social media is an online media that has the ability to strengthen social relationships.

Twitter is one of the social media that is always trending in Indonesia, now there is a lot of information that can be answered quickly and accurately from various points of view. Therefore, Twit ter not only has a positive impact but also a negative impact on Twitter users and non-users, one of which is public opinion towards the national police institution[4]

To see large data information on social media Twitter still uses manual methods, the difficulty of this manual method is that it cannot collect so much data automatically. Therefore, sentiment analysis is needed so that it is easier to classify public opinion on tweets from Twitter users quickly and automatically.

Sentiment analysis or opinion research includes data-based data whose purpose is to analyze, understand, mana ge and extract textual data as opinion data. Some algorithms include Support Vector Machine (SVM), Naïve Bayes Classifier, K-Nearst Neighbor (KNN), Artificial Network (ANN) and Decision Tree. The algorithm used Support Vector Machine has been proven to have more stable classification accuracy than other algorithms and is stronger and more generalizable[5]

Conducted research by implement ing the Naïve Bayes Classifier method to carry out sentiment analysis of online transportation in Indonesia on Twitter media. Accuracy test results from testing 109 data resulted in an accuracy value of 84%[6]

Conducted research by implemen ting the Naïve Bayes Classifier method to carry out sentiment analysis on Twitter media belonging to the National Police Public Relations Division where the tweets to be analyzed were classified into three topics: police activities, community services and community comments. The accuracy results of clustering testing on this system are 86%[7]

The test results obtained using the K-Fold Cross Validation and Confusion Matrix methods on a model created using

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

the Support Vector Machine algorithm gave accuracy values of 79.6%, precision 76.5%, recall 72.8%, and F1-score 74 .6% for Telkom, as well as 83.2% accuracy, 78.8% precision, 71.6% recall, and 75% F1-score for Biznet[8]

The author uses the Support Vector Machine (SVM) method in sentiment analysis regarding the opinions of Indi home service users on Twitter, with the aim of obtaining a sentiment classifica tion model using SVM, and to find out how much accuracy is produced by the SVM method applied to sentiment anal ysis, as well as to find out how satisfied are Indihome service users based on Twitter. After testing using the SVM method, the results were accuracy 87%, precision 86%, recall 95%, error rate 13%, and f1-score 90%[9]

This research uses the CRISP-DM (Cross Industry Standard Process for Data Mining) model and the algorithm used in this research is K-Nearest Neighbors. Based on the results obtained from the modeling stage using the knearest neighbors algorithm and a ratio of 60:40 for training data and testing data, the precision and recall accuracy values resulting from each application are 85.14%, 91.91% and 76% for seeds. .44% while for bareksa it is 81.70%, 87.15%, 75.73%[10]

The purpose of this study is to classify sentiment towards public opinion based on tweets from Twitter users using the Support Vector Machine (SVM) method.

METHOD

Method of collecting data

The data collection phase uses RapidMiner's crawling capabilities. The data studied was obtained from a collection of tweets on the Twitter platform. Data collection from websites, databases, documents, and other sources is automated through data crawling. Data collection was carried out using the RapidMiner application to crawl tweet data and identify the keywords most talked about by the public. #institution polri, #humas polri, and #polda.

Research Stages

Research stages as a collection of facts or data obtained from a research object. The research object in this case is the tweet data on Twitter social media, then the data will be used as a data source. In conducting this research, a research process is needed, so that the research process continues to run according to plan. Figure 1 shows the flow of this research.

Figure 1. Research Stages

There are three keywords used in this crawling stage: (Police Institution), (Polri Public Relations), and (Regional Police). A total of 756 data keywords (Polri Institution) and 1,255 data on (Polri Public Relations). Found from 4,914 data items on the keyword (Regional Police). If your data has an ID, you need an attribute selection operator. This operator only displays text. Next, you can create a CSV write file.

Table 1 below is the result of crawling data on the Twitter application.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 1. Crawling Data

No	Text
1	The Indonesian National Police Public Relations Team is working with a number of mass media to provide security and social security assistance ahead of the election.
2	The Lembang Police visited the Baroka Playgroup (KB) in Panusupan Village and the Al-Kwar Playgroup in Smanpir Village and provided material on introducing the Polri profession.
3	Recruitment of police members is free of charge, as conveyed by Ahmad Ramadan. RT @antaranews:

Data Preprocessing

By excluding all characters except letters, this technique eliminates the number of Unwanted and inappropriate characters in sentiment analysis. These characters include numbers, #, @, emojis, and even links to websites within the article. The collected data is unstructured and requires pre-processing. Therefore, it is important to replace RT with URLs, hashtags, mentions, and icons to tidy up the data from looking useless.

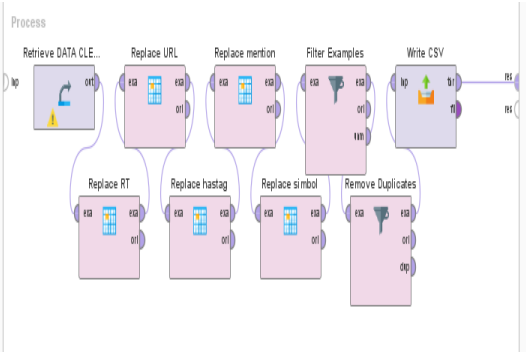


Figure 2. Data Cleaning Process

Cleansing Replace RT

In the Replace RT operator, it is used as its function is only to remove retweets such as RT @antaranews:

Cleansing Replace Url

In this process, all non-alphabetic characters in the message are removed to reduce unwanted and senseless characters in sentiment analysis. The characters are like (link) in a post.

Cleansing Replace Hashtag

In this process, all non-alphabetic characters in the message are removed to reduce unwanted and unreasonable characters in sentiment analysis. These characters are the same as the symbol (#) in a message.

Cleansing Replace Mention

In this process, all non-alphabetic characters in the message are removed to reduce unwanted and unreasonable characters in sentiment analysis. Characters such as the (@) symbol that is present in a post.

Cleansing Replace Symbol

In this process, all non-alphabetic characters in the message are removed to reduce unwanted and unreasonable characters in sentiment analysis. These characters are like symbols (!@#$%^&*_) that are in a post.

Cleansing Filter Examples

When using the examples filter, all empty data has no meaning, such as? will be permanently deleted, so there is no empty text data.

Cleansing Delete Duplicates

When using remove duplicate all duplicate data will be permanently deleted, so no data will be accidentally deleted.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

TF-IDF Feature Extraction

Tokenize

After undergoing cleaning and letter folding, sentences are separated into words based on user tweets or comments in this phase.

Case Folding

The process of changing all characters in a post to lowercase or all characters to lowercase.

Stopword

Less meaningful words in a document, such as "in," "to," and "what," are less important than stop words.

Filter Token

The token filter removes words that have no significance by entering a minimum of 3 characters and a maximum of 25 characters. Therefore, words with less than 3 letters and words with more than 25 characters will be removed because they are not needed.

Evaluation

Cross Validation

To verify the classification results of this method, we conducted a cross-validation performance evaluation. Here are some percentages of positive and negative opinions on the classification results. Evaluation and testing were carried out by reviewers using the k-fold cross-validation test classification model. We use K = 5 data divided into 5 test data with the same number of dates, containing one test data and four training data.

Table 2. Overview of Fold Validation Scheme 5

First process	Second process	The third process	The fourth process	The fifth process
Partition 1	Partition 2	Partition 3	Partition 4	Partition 5
Partition 1	Partition 2	Partition 3	Partition 4	Partition 5
Partition 1	Partition 2	Partition 3	Partition 4	Partition 5
Partition 1	Partition 2	Partition 3	Partition 4	Partition 5
Partition 1	Partition 2	Partition 3	Partition 4	Partition 5

RESULTS AND DISCUSSION

TF-IDF Feature Extraction

The result of tf-idf processing on documents from a total of 2389 data is that each text word is decomposed into 5931 words per word (regular attributes). The number 0 is the amount of data from the TF-IDF calculation of the words that appear in the document, which is as much as seen below in Figure 3.

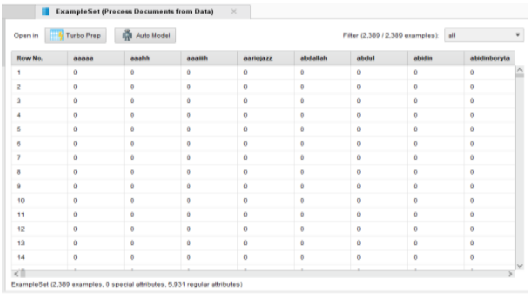


Figure 3. TF-IDF results

Support Vector Machine Method Modeling Results

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 3. Test Results of 600 Testing Data

Prediction positive and negative	True Negative	False Positive
Prediction Negative	2 True Negative	174 False Positive
Prediction Positive	1 False Negative	423 True Positive
Amount	600	

Accuracy results $= \frac{423+2}{423+2+174+1} *$ 100% = 71

Precision results $= \frac{423}{174+423} *$ 100% = 70.85

Recall results $= \frac{423}{1+423} * 100\%$ = 99.76

Table 4. Test Results of Test Data 400

Prediction positive and negative	True Negative	False Positive
Prediction Negative	0 True Negative	55 False Positive
Prediction Positive	1 False Negative	344 True Positive
Amount	400	

Accuracy results $= \frac{\mathbf{344+0}}{\mathbf{344+0+55+1}} *$ 100% = 86

Precision results $= \frac{344}{55+344} *$ 100% = 86.21

Recall results $= \frac{344}{1+344} * 100\%$ = 99.71

Here are the results of the manual test date. The results of the manual test date on Data Test 600 (73:27) showed a precision of 71%, a precision of 70.85%, a recall of 99.76%. The next test on 400 data tests (82:18) showed a precision of 86%, a precision level of 86.21%, a recall level of 99.71%. See the table above for numbers 3 and 4.

Automatic Split Validation Results

In Figure 4, the confusion matrix is created using 915 (60:40) test data. Classifying using the support vector machine (SVM) method, we find that s[split ratio = 0.6 True Positive = 772. False Positive = 117, False Negative = 3, True Positive = 23, the calculation result is 86.84% precision, and 99.61% recall. So the confusion matrix of 86.89% Accuracy is achieved by Support Vector Machine (SVM)

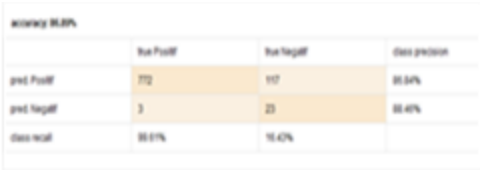
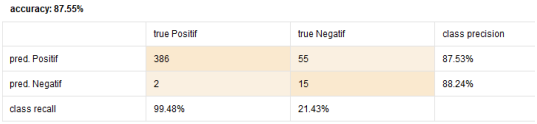


Figure 4. Results of 40% accuracy split validation

In Figure 5, the confusion matrix is created using 458 (80:20) test data. When classifying using the support vector machine (SVM) method, the split ratio = 0.6 is assessed as True Positive = 386, False Positive = 55, False Negative = 2, True Positive = 15. Then the results of this accuracy are calculated, namely 87.53%, 99.48% recall rate. The accuracy of 87.55% is determined by the confusion matrix using the support vector machine (SVM)

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 53 – 60 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3234
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi



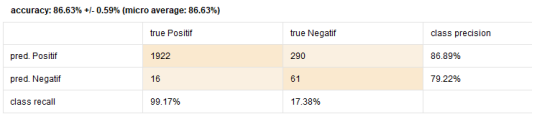
accuracy: 87.55%

	true Positif	true Negatif	class precision
pred. Positif	386	55	87.53%
pred. Negatif	2	15	88.24%
class recall	99.48%	21.43%	

Figure 5. 20% Validation Split Accuracy Results

Automatic Crossvalidation Results

In Figure 6, the confusion matrix created for classification using the support vector machine method is evaluated as k fold = 5, namely True Positive = 1922, False Positive = 290, False Negative = 16, True Positive = 61. The calculated precision and results are 86.89%. the ongoing recall has 99.17% recall. Based on the results of the support vector matrix confusion machine obtained 86.63% Accuracy.



accuracy: 86.63% +/- 0.59% (micro average: 86.63%)

	true Positif	true Negatif	class precision
pred. Positif	1922	290	86.89%
pred. Negatif	16	61	79.22%
class recall	99.17%	17.38%	

Figure 6. Cross validation accuracy results

In Figure 7, the circle diagram with 85% positive data and 15% negative data. The number of positive sentiments is 1938 data and the number of negative sentiments is 351 data.

Figure 7. Number of positive negatives

Evaluation

The test results above prove that the support vector machine algorithm with test results has a good accuracy score if accuracy is measured using a confusion matrix. Test Data 600 has an accuracy of 71% if tested manually, Test Data 400 has an accuracy of 86% if tested manually, Split System Verification Data Test (60:40) has an accuracy of 86.89%, Split System Verification Data Test (80:20). Expand the total data cross-validation system to obtain an accuracy of 87.55% and 86.673% for k-fold accuracy 5 From Figure 6 shows the best test comparison results.

Table 5. Predicted Results

Comparison of Test Results	Accuracy	Precision	Recall
Test Results of 600 Testing Data	71%	70.85%	99.76%
Test Results of Test Data 400	86%	86.21%	99.71%
Split validation data (60:40)	86.89%	86.84%	99.61%
Split validation data (80:20)	87.55%	87.53%	99.48%
Overall cross validation data	86.63%	86.89%	99.17%

CONCLUSION

By using the support vector machine technique, positive and negative emotions can be classified. Analysis of public sentiment towards the national police agency based on Twitter using the support vector machine (SVM) method is based on crawling raw data of 6,925 items from three keywords: police agency, public information Polri, and local police. This conclusion is supported by manual testing on labeling 600 test data (73:27) with an accuracy of 71% and on labeling 400 test data (82:18) with an accuracy of 86%. Testing on the data system (80:20) achieved an accuracy of 87.55% and labeling data (60:40) achieved an

accuracy of 86.89%. The more training data, the higher the accuracy achieved.

BIBLIOGRAPHY

[1] H. Ritonga, M. Marlina, and M. Mustamam, "PENINDAKAN PROPAM (POLRI TERHADAP ANGGOTA POLISI YANG MELAKUKAN PENGANIAYAAN (Studi Di Bidang Propam Kepolisian Resor Nias Selatan)," *J. Ilm. METADATA*, vol. 4, no. 3, pp. 215–227, 2022.

[2] Edi Saputra Hasibuan, "Persepsi Masyarakat Terhadap Penerimaan Anggota Polri," *J. Huk. Sasana*, vol. 7, no. 1, pp. 33–50, 2021, doi: 10.31599/sasana.v7i1.526.

[3] E. Alfian, "Tugas dan Fungsi Kepolisian Untuk Meningkatkan Kepercayaan Publik terhadap Penegak Hukum," *Leg. J. Huk.*, vol. 12, no. 1, p. 27, 2020, doi: 10.33087/legalitas.v12i1.192.

[4] M. F. Rizki, K. Auliasari, and R. Primaswara Prasetya, "Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 548–556, 2021, doi: 10.36040/jati.v5i2.3808.

[5] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.

[6] T. Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, "Tweet mengenai transporta-si online," *J. GEEJ*, vol. 7, no. 2, pp. 5–16, 2020.

[7] D. Kurniawati, E. Prayitno, D. F. Sari, and S. N. Putra, "Sentiment Analysis of Twitter Use on Policy Institution Services using Naive Bayes Classifier Method," *J. Int. Conf. Proc.*, vol. 2, no. 1, p. 33, 2019.

[8] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.

[9] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.

[10] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.