Journal of Social Research

P-ISSN: 2827-9832
E-ISSN: 2828-335x
http://ijsr.internationaljournallabs.com/index.php/ijsr

SENTIMENT ANALYSIS OF RELATED PUBLIC OPINION SMART CITY PASURUAN REGENCY ON MEDIA SOCIAL USING SVM ALGORITHM

Sony Nuhmana
Institut Sains dan Teknologi Terpadu Surabaya, Surabaya
nuhmana.sony@gmail.com

ABSTRACT

Development in Pasuruan Regency is developed sustainably in a process, thus impacting daily community service activities. With interaction, the community feels what is done in the service process. With Smart City, these services are transformed into a more modern and computer-based form. Along with the development of social media. People often provide comments or opinions through social media, including social media, namely Facebook. Facebook is a social networking and microblogging service that allows its users to send and read text-based messages. Community opinion is important information about their reactions to district government programs, so it is important to record and analyze. The research conducted was an analysis of public sentiment using the Support Vector Machine method. The most effective method for classifying text is the Support Vector Machine (SVM). The accuracy of SVM classification is 92.6%. In this study, the value obtained for the positive sentiment was below 1.0., and negative at 1.0. As well as Recall (sensitivity) which is the ratio of positive true predictions compared to overall positive true data. Then the positive value has a value of 1.0 and the negative is at 0.5. So it is determined that the selection of opinions is to use high precision. This research shows that threshold variations applied to the system have a major influence on classification results. This statement is supported by the results of research on the 10th test which showed that with a threshold value of 0.120 – 0.170 produces precision above 80% for the Administration, Facilities, Staffing, and Service classes.

Keywords: *Sentiment Analyst, Classification, Support Vector Machine, Pasuruan District*

INTRODUCTION

The development of information technology today is very fast. The rate of ICT usage is multiplying exponentially. For example, according to the Ministry of Communication and Informatics, internet users in Indonesia increased by 11 percent from the previous year, from 175.4 million to 202.6 million users (Kominfo, 2021). The condition means that social media users are also growing. Information on social media users in Indonesia reached 170 million users in early January 2021.

With the existence of Public Information Disclosure, the role of social media becomes interesting to be used as an information resource (Latifah & Elfiandri, 2021). The process of obtaining this information can be managed into a conclusion or analysis (Yunita & Devitra, 2017).

Social Network Analysis (SNA) is conducted to see opinions or opinion tendencies toward a problem, whether it tends to have a negative or positive view (Stieglitz et al., 2014). One example of the use of sentiment analysis in the real world is the identification of public opinion toward the work programs of local governments (Dale et al., 2020). The public services performed by the current government apparatus are meeting people's expectations (Mulgan &

Albury, 2003). This can be known from various public sentiments conveyed through mass media and social networks, so that quite objective images can be obtained (Bernays, 1947).

Based on reviews that have been expressed by Huberman, Romero, and Wu (Bernardo A. Huberman, n.d.), in their research on predictions of films to be favored In this article, it has been aimed at how social media can be used to predict future results. Specifically, it uses chat rates from nearly 3 million tweets from polluter sites, Twitter, and created a linear regression model to predict a movie's box-office revenue before it's published. it was then shown that the results outperformed the accuracy of those from the Hollywood Stock Exchange and that there was a strong correlation between the amount of attention to a particular topic (in this case an upcoming film) and its future ranking.

The demands of complexity in urban areas such as population growth problems (Broere, 2016), climate change, congestion, poverty, crime, natural disasters, and so on must be able to be solved in the concept of smart cities/districts. This is expected to be the answer to these various problems (Hooke & Jeeves, 1961). The Smart City concept must be able to provide basic service support for the wider community that is adapted to the conditions of regional characteristics and the needs of the people in the region (Zavratnik et al., 2020).

The vision of smart city/regency development is the creation of a unified green city that is competitive and technology-based supported by the synergy of a smart economy, smart people, smart government, smart mobility, and smart living.

Pasuruan Regency as one of the areas selected in SMART City by the Ministry of Communication and Informatics in 2019 has a target.

The targets or objectives of the development of Smart Cities / Regencies (Smart City), among others:

1. A city/district performs well with a view to the economy, population, government, mobility, environment

2. A city/district that controls and integrates all infrastructure including roads, bridges, tunnels, tracks, subways, airports, ports, communications, water, electricity, and building management.

3. Smart cities/districts can connect physical infrastructure, IT infrastructure, social infrastructure, and business infrastructure to improve city/district development.

4. Smart cities/districts make cities more efficient and livable

5. The use of smart computing to make smart cities and their facilities include education, health, public safety, smarter, interconnected, and efficient transportation.

Sentiment Analysis or known as opinion mining is computer science research on a person's opinions, judgments, and emotions towards an institution or entity, and events (Ravi & Ravi, 2015). The sentiment analysis system for a comment aims to automatically separate first between opinions or non-opinions. After successfully recognizing comments that contain opinions, the next step is to separate them into positive or negative opinions (Wilson et al., 2005).

By grouping comments into positive or negative opinions, training can be carried out on existing data by searching for important keywords for positive or negative opinion comments (Bilal et al., 2016).

Sentiment analysis is a subfield of Natural Language Processing (NLP) relating to the determination of opinion and subjectivity in text, which has many applications. In this paper,

we learned about the classification process for sentiment analysis from user opinions on the implementation of the smart city program using the Support Vector Machine (SVM) (MS Gauragangi Patil, 2014). The goal is to build a classifier that can perform sentiment analysis, by marking comments from users as positive comments or negative comments.

Sentiment analysis, which is usually referred to as opinion mining, is a Natural Language Processing and Information Extraction task that aims to obtain the author's feelings expressed in a positive or negative comment, a question, or as a request obtained by analyzing a large number of sentiment documents. Sentiment analysis is a computational technique for extracting, classifying, understanding, and determining opinions presented in various forms of content. In sentiment classification, there are four different levels of sentiment analysis, namely sentence level, document level, phrase level, and word level. Subjectivity and sentiment are relevant properties of a language. Subjectivity refers to the linguistic expression of one's opinions, beliefs, or speculations. The main task of subjectivity is to classify content into objective or subjective.

The following is an overview of the steps and techniques commonly used in sentiment classification approaches order:

1. Text Preprocessing

Data pre-processing is a process to prepare and clean the data in the dataset for the purposes of the classification process. The following is a hypothesis to pre-process the data correctly, by curating noise in the text will help improve classifier performance and speed up the classification process, so it can help the sentiment analysis process in real-time.

The following are the steps to do text preprocessing:

a. Tokenization, turning content into pieces of words.
b. Stop Word Removal, delete all common words.
c. Stemming, which is the truncation of words to find the base word in order to group formations derived from the same base word
d. Transformation

The weight for each word inside is calculated using the help of the TF-IDF. Using the help of TF-IDF we can easily determine what words are inside a set of documents, and determine which words are more profitable to use in the subsequent process. The value of the TF-IDF will be calculated for each word in the document.

2. Feature Selection

Feature selection is used to make classifying more efficient by reducing the amount of data to be analyzed as well as identifying relevant features to consider in the classification process. Ideally, the feature selection stage will refine the feature, which is an input into the classification or learning process. The following are the stages of feature selection:

a. Identify parts of the corpus to contribute to both positive and negative sentiments.
b. Combining parts of the corpus in such a way that the document belongs to the polar category.

3. Classification

The purpose of text classification is to classify data into predetermined classes, where the data will be classified into positive and negative classes. Text classification is a matter of supervised learning. The first step in the text classification process is to convert the document from a string form to a form suitable for the learning algorithm and classification process.

Support Vector Machine (SVM) is a universal learning method. The advantage of SVM is that its ability to learn does not depend on the dimensions of the feature space. SVM uses g(x) as a discriminate function

$$g(x) = w^T f(x) + b$$

4. SVM learning Algorithms for Text Categorization

SVM has a special format for its input and output. The input is vector space and the output is 0 or 1 (positive or negative). Text documents in their original form are not suitable for the learning process. This text document must be converted to the appropriate format for the machine learning algorithm input. For this reason, the preprocessing process is needed, after doing the preprocessing process, we will carry out a transformation process, where each word represents one dimension and the same words will lie on the same dimension as well. At this stage, machine learning algorithms are used for the process of learning how to classify documents, such as creating input and output mapping models. SVM has proven itself to be one of the learning algorithms powerful enough to perform text categorization. The advantages of SVM are:

a. High Dimension Input Space, when text classification has to deal with many features (probably more than 1000 features). Since SVM does not depend on the number of fit, it has the ability to handle a large number of features.
b. Document Vector Space, although it has a high dimensional representation, each vector document has only a few non-zero elements. Most text categorization problems are linear problems.

Characteristics of SVM:

a. It is a machine learning algorithm that uses vector space.
b. Use binary categories.
c. The number of words increases if we increase the number of documents.
d. Representation using words as in text.

From this study, we can conclude that SVM recognizes some text properties such as:

High Dimensional feature space
Some irrelevant features.
Sparse instance vector.

In addition, it can also be concluded that SVM provides good performance in text categorization. With the ability to generalize high-dimensional feature space, SVM removes the need for feature selection.

METHOD

In this study, there are three main processes that will be carried out by the system, namely the input process, preprocessing, and classification process with the Supervised Vector Machine method. Broadly speaking, the workflow of the text classification system is represented in Figure 4.

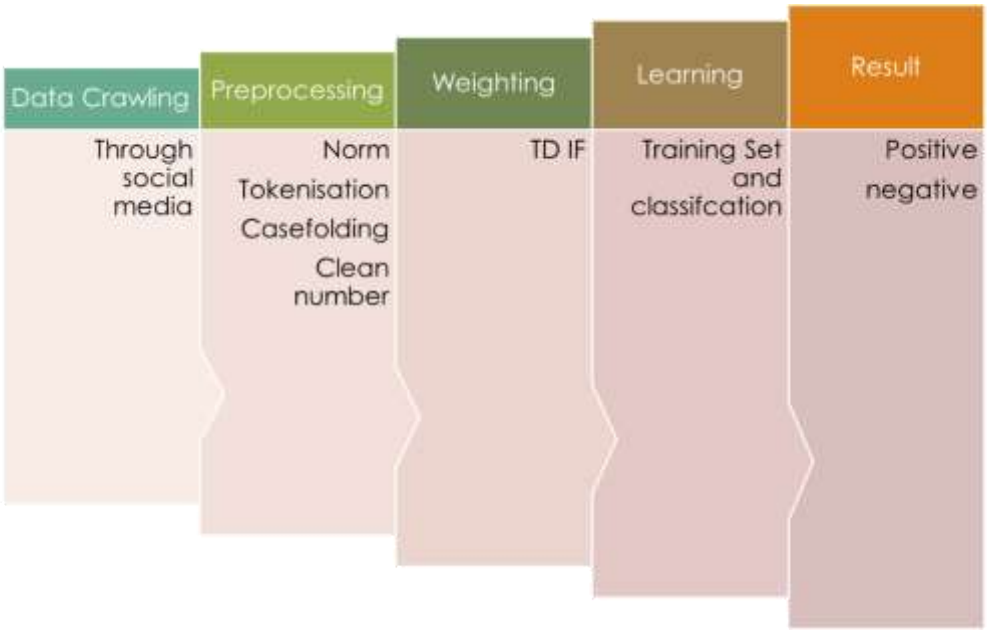

Figure 4 Flow of the1 Sentiment Analysis system in general

In this study, the data used was comment text data from pasuruan district social media. The dataset is obtained directly from the agency that is the object of research. Furthermore, preprocessing steps will be carried out, namely letter equalization or called *case folding, tokenization, stemming, stopword removal,* and weighting using TF-IDF.

Making a complaint classification model requires data that has been labeled as a source of learning data. The dataset is loaded into two variables, namely training data and testing data.

RESULTS AND DISCUSSION

Based on the compilation of the program created to perform sentiment analysis operations, then for the evaluation of this classification managed to get the precision & recall value as in Table 2.

Table 2 Precision & Recall Results1

Items	Positive	Negative
Precision	0.879654255	1.0
F1 Score	0.935974531	0.1083743842
Recall	1.0	0.572916666

From Table 2, Precision is a representation of the level of accuracy between the information requested by the user and the answers given by the system. In this study, the value obtained for the positive sentiment was below 1.0 and negative at 1.0. As well as Recall (sensitivity) which is a positive true prediction ratio compared to the overall positive true data. Then the positive value has a value of 1.0 and the negative is at 0.5. So it is determined that the selection of opinions is to use high precision.

CONCLUSION

In this study, it was shown that the threshold variation applied to the system had a great influence on the classification results. This statement is supported by the results of research in the 10th testing which shows that with a threshold value of 0.120 – 0.170 produces precision above 80% for the Administration, Facilities, Staffing, and Service classes. Meanwhile, the average f-measure reaches 87.20% compared to other threshold values. Based on the test results in this study, it can be concluded that the K-Competitive Autoencoder method with a combination of thresholds is adaptively able to classify complaint texts very well.

REFERENCES

Bernardo A. Huberman, S. A. (n.d.). *Predicting the Future With Social Media*. Researchgate.

Bernays, E. L. (1947). The engineering of consent. *The Annals of the American Academy of Political and Social Science*, *250*(1), 113–120.

Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University-Computer and Information Sciences*, *28*(3), 330–344.

Broere, W. (2016). Urban underground space: Solving the problems of today's cities. *Tunnelling and Underground Space Technology*, *55*, 245–248.

Dale, A., Robinson, J., King, L., Burch, S., Newell, R., Shaw, A., & Jost, F. (2020). Meeting the climate change challenge: local government climate action in British Columbia, Canada. *Climate Policy*, *20*(7), 866–880.

Hooke, R., & Jeeves, T. A. (1961). ``Direct Search''Solution of Numerical and Statistical Problems. *Journal of the ACM (JACM)*, *8*(2), 212–229.

Kominfo, A. (2021). Warganet Meningkat, Indonesia Perlu Tingkatkan Nilai Budaya di Internet. *Diakses Pada*, *10*.

Latifah, L., & Elfiandri, E. (2021). MANAJEMEN PENGELOLA WEBSITE WWW. PEKANBARU. GO. ID SEBAGAI MEDIA INFORMASI PUBLIK DINAS KOMUNIKASI INFORMATIKA STATISTIK DAN PERSANDIAN. *Jurnal Riset Mahasiswa Dakwah Dan Komunikasi*, *3*(3), 155–169.

MS Gauragangi Patil, M. V. (2014). Sentiment Analysis Using Support Vector Machine. *International Journal of Innovative Research in Computer and Communication Engineering*.

Mulgan, G., & Albury, D. (2003). Innovation in the public sector. *Strategy Unit, Cabinet Office*, *1*(1), 40.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, *89*, 14–46.

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics: An interdisciplinary approach and its implications for information systems. *Business & Information Systems Engineering*, *6*, 89–96.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 347–354.

Yunita, I., & Devitra, J. (2017). Analisis dan perancangan sistem informasi manajemen aset pada SMK Negeri 4 Kota Jambi. *Jurnal Manajemen Sistem Informasi*, *2*(1), 278–294.

Zavratnik, V., Podjed, D., Trilar, J., Hlebec, N., Kos, A., & Stojmenova Duh, E. (2020). Sustainable and community-centred development of smart cities and villages. *Sustainability*, *12*(10), 3961.