SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of Machine Learning Classification Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency

Yennimar
Universitas Prima Indonesia
Medan, Indonesia
yennimar@unprimdn.ac.id

Reyhan Achmad Rizal
Universitas Prima Indonesia
Medan, Indonesia
reyhanachmadrizal@unprimdn.ac.id

***Abstract*—** The growth of SMEs in Indonesia, which has increased by 6% every year, is driven by continued growth by many parties, including the government and private institutions that often conduct business coaching and assistance. Problems that are often encountered are the lack of willingness of MSME business practitioners to apply information technology and the internet, besides that most of them live in rural areas with very limited internet access and many are not yet digital-literate, adequate digital technology utilization capabilities and the will of business people For SMEs to understand customer needs, a service that is consistent with standard service procedures will give a good impression and pay attention to customer feedback. This research was conducted by collecting data on MSME products obtained from the North Padang Lawas District Trade Industry Office followed by the development of a Paluta Market website as a marketplace for media promotion and marketing of MSME products in North Padang Lawas by applying a sentiment analysis approach using machine learning classification algorithm to produce product rating values based on public opinion of MSME products contained on the website, in addition the system is able to classify consumer comment data on MSME products from various sources from the umkm web, so that it becomes useful information for MSME businesses especially in North Padang Lawas Regency and the community at large. The results of the application of sentiment analysis of a product on the Paluta Market website can be used as a reference in improving service and product quality, so as to create a variety of new opportunities that are profitable for MSME businesses.

Keywords— Sentiment Analysis, MSME, North Padang Regency, Product Riview, K-Nearest Neighbor

I. INTRODUCTION

The main priority in social development in Indonesia is the welfare of the people. One of the efforts to increase community income is Micro, Small and Medium Enterprises (MSMEs). In 2013, the number of MSMEs in Indonesia reached 57.89 million units, while the number of large businesses in Indonesia was only 5,066 units. This shows that MSMEs have a proportion of 99.99% of the total business units in Indonesia. In the last five years, the contribution of the MSME sector in GDP increased from 57.38% to 60.34%. Absorption of domestic workers in the MSME sector also increased from 96.99% to 97.22% (depkop, 2013). This makes MSME a very strategic economic component. The 4th Industrial Revolution can improve the welfare of life of the world community. Existing technology enables the emergence of facilities that increase effectiveness and efficiency in carrying out daily life, such as booking tickets, shopping, and payments can be done online, enabling the provision of facilities at lower prices with the development of online transportation markets (GoJek, Uber, Grab), e-commerce, online platforms and more.

The position of MSME that is so strategic in the Indonesian economy, if it is incorporated with

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the presence of the Industrial Revolution 4.0, will have a great influence, to increase the accessibility and capability of MSMEs to go digital, so as to produce products that are able to compete with foreign products that have flooded e-commerce. Indonesia, however, most MSMEs live in rural areas with very limited internet access and there are still many that are not yet digital-literate, besides that it needs to be balanced with the ability to utilize adequate digital technology and the willingness of MSME businesses to understand customer needs, consistent services with standard service procedures, so as to give a good impression and pay attention to customer feedback.

An item review is useful to see how previous buyer feedback is through a positive or negative comment. User comments express their opinions about quality, price, service and delivery speed. Online shopping users often use comments from previous users when they are going to make a purchase of goods (Chen, 2012), grouping reviews of goods from consumers is influenced by emotions (sentiments) that are grouped or classified to determine their polarization, positive or negative (Indriati, 2016).

This research focuses on creating a North Sumatra UMKM marketplace website as a promotional media and product marketing by implementing sentiment analysis using a machine learning classification algorithm with the aim that each product has a rating based on the classification of public opinion on products sourced from social media and comments on the website. Our analysis framework first analyzes the NB, SVM, KNB, ANN classification model, then we analyze the results to see the most accurate model to be implemented on each MSME marketplace website product. The application of sentiment analysis to a product in this study is expected to be an evaluation and reference tool in improving service and product quality, so as to create a variety of new opportunities that are profitable for MSME business people

II. LITERATURE REVIEW

Sentiment classification algorithms such as Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNB), artificial neural network (ANN) are widely proposed by researchers because they can be applied to various problems such as television program ratings (Saifinnuha, 2015) , mobile app reviews (Firmansyah, 2016), user satisfaction levels of Indonesian cellular telecommunications service providers (Rofiqoh, 2017), television shows based on public opinion (Nurjanah, 2017), Product Opinions (Hanggara, 2017), item reviews (Haryanto, 2018) , assessment of online shopping site services (Muljono, 2018), market sales reviews (Lufti, 2018), travel agents (Ernawati, 2018), hotel reviews (Taufik, 2018) and others.

a. K-Fold Cross Validation

Cross validation is one technique to validate the accuracy of a new model that is built based on a new dataset. The data used in making the new model is called training data and the data used to validate is test data. The new classification model is created to predict a new dataset so that it can be applied in the future, so we need a validation so that the resulting model has good performance. K Fold Cross Validation is one of the cross validation methods used to calculate the prediction accuracy of a system. Data is divided into k segments that have the same or almost the same ratio. Training of data and validation k times with each experiment taking one different segment as test or validation data and k-1 other segments as training data to retrieve the average of the results of each iteration (Refaeilzadeh, Tang, & Liu, 2009).

b. Support Vector Machine

Support Vector Machine (SVM) is a learning system that uses hypothetical spaces in the form of linear functions in high-dimensional feature space, trained with learning algorithms based on optimization theory by implementing learning bias derived from statistical learning theory.

SVM is a machine learning method that works on the principle of Structural Risk Minimization (SRM) with the aim of finding the best hyperlane separating two classes in the input space. SVM can be used for classifications that are applied to handwriting detection, object recognition, voice identity and others (Ulwan, 2016).

c. TF-IDF (Term Frequency – Inverse Document Frequency)

TF-IDF (Term Frequency - Inverse Document Frequency) algorithm is an algorithm that can be used to analyze the relationship between a phrase / sentence and a collection of documents. The main core of this algorithm is to calculate the TF value and the IDF value of each keyword for each document. The TF value is calculated by the formula TF = number of selected word frequencies / number

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X

of words and IDF value is calculated by the formula IDF = log (number of documents / number of selected word frequencies). Next is to multiply the TF and IDF values to get the final answer. (Informatikalogi, 2016).

d. Sentiment Analysis

Sentiment analysis or opinion mining is a field of data mining that has the purpose of analyzing, understanding, processing and extracting textual data in the form of opinions, sentiments, evaluations, attitudes, and emotions towards an entity such as products, services, organizations, individuals, and certain topics . Sentiment classification is used to solve two class classification problems, positive and negative. The test data used is usually an online product review. Because online reviews have a rating value set by researchers. Most research papers do not use neutral classes, with the aim of facilitating classification problems, but also not infrequently researchers who use neutral classes, for example the use of three-star ratings as neutral classes (Liu, 2012).

e. Machine Learning

Machine Learning is the study of how computers can learn or improve performance based on data in order to automatically recognize data patterns to make intelligent decisions based on data. Machine learning is an area of artificial intelligence that deals with the development of techniques that can be programmed and learned from past data. Pattern recognition, data mining, and machine learning are often used to refer to something that is the same or intersects with the science of probability and statistics sometimes optimization. Machine learning as an analytical tool in data mining (Han, Kamber, & Pei, 2012)

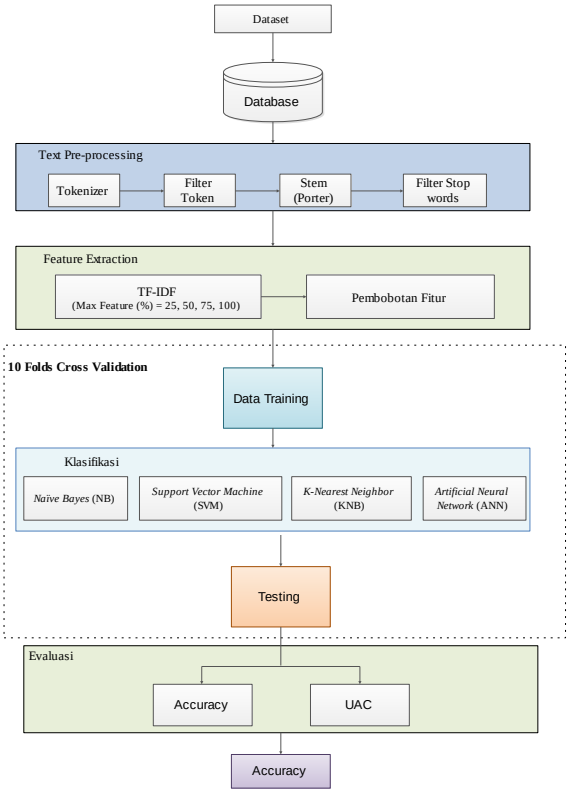
III. PROPOSED METHOD

a. Proposed model

In the study of Comparative Algorithm of Machine Learning Classification in Sentiment Analysis Product Review North Padang Lawas Regency consists of several stages carried out, namely:

1. Preparation stage.
   Research activities are focused on determining the topic, identifying and formulating the problem, determining the research objectives, determining the boundaries and research methodology.
2. Literature Study
   Literature study stages are used to find reference information about sentiment analysis, topics related to research objects, machine learning methods and literature related to the proposed method.
3. Data Collection
   At this stage the process of collecting data on the products of North Padang Lawas Regency will be carried out.
4. Stages of Design.
   At this stage the design of the palutamarket.com website system will be carried out as a media for marketing and promoting SME North Padang Lawas products.
5. Stages of Sentiment Analysis, covering several stages, viz
   a. Stages The process of retrieving SME product data from customer commentary sources on the PalutaMarket website.
   b. The pre-processing stages consist of several stages, namely:
      1. Tokenize is the process of separating words. These words are called tokens or terms
      2. Filter Token is the process of taking important words from the token results.
      3. Stem is the process of changing the form of words into basic words. This method of changing the form of words into basic words adjusts the structure of the language used in the stemming process.
      4. Stopwords filter is the process of eliminating words that often appear but do not have any effect in the extraction of sentiment of a review. The words included are like timepieces, question words.
   c. Feature Extraction is the stage of making classifiers more efficient by reducing the amount of data analyzed, while weighting to determine the polarity of the opinion target (positive or negative). In addition, giving weight to the target opinion also aims to rank product features, so that it can be known what product features are most liked and least liked by the customer
   d. Machine Learning Classification Algorithm Analysis is the method analysis stage used to analyze the accuracy of the product learning algorithm review product classification algorithms namely Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNB) and Artificial Neural Network (ANN))

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X



Picture 1. Proposed framework

Text processing aims to prepare unstructured text documents into structured data that is ready to be used for further processing. Stages of text processing include:

1. Tokenize is the process of separating words. These words are called tokens or terms.
2. Filter Token is the process of taking important words from the token results.
3. Stem is the process of changing the form of words into basic words. This method of changing the form of words into basic words adjusts the structure of the language used in the stemming process.
4. Stopwords filter is the process of eliminating words that often appear but do not have any effect in the extraction of sentiment of a review. The words included are like timepieces, question words.

a. Data obtained

The dataset used for training and testing of the proposed method is sourced from the Department of Industry and Trade of North Padang Lawas Regency, where the process of collecting data is directly to the location.

Picture 2 Dataset

IV. RESULT AND DISCUSSION

IV1. Test result

In the study of Comparative Algorithm of Machine Learning Classification in Sentiment Analysis of the Padang Lawas Utara Regency Product Review, the step taken is to classify the final comment based on the products on palutamarket.com then we make a test analysis framework. The researcher proposes to compare 4 (four) classification algorithms (SMV, NB, KNB and ANN). Tests carried out using the Intel Core i7 1.8 GHz CPU specifications, 16 GB RAM, and the Microsoft Windows 10 Home 64-bit operating system, in table 1 are the results of the classification classification of the proposed method and Figure 3 Comparison Chart.

Table 1 Comparison of Classification Results

Method	Accuracy	Precision	Recall	TP Rate	TN Rate
NB	89.00	81.00	82.00	89.00	84.00
SVM	94.00	98.00	85.00	87.00	93.00
KNB	67.00	99.00	76.00	65.00	63.00
ANN	68.00	66.00	58.00	81.00	78.00

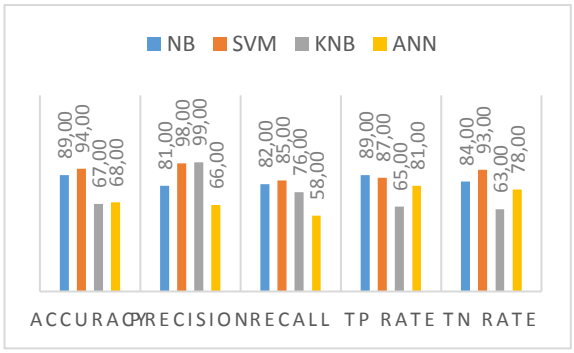


Figure 3 Comparison Chart

Table 2 Comparative Results Methods

Method	UAC
NB	0.876
SVM	0.982
KNB	0.605
ANN	0.413

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X

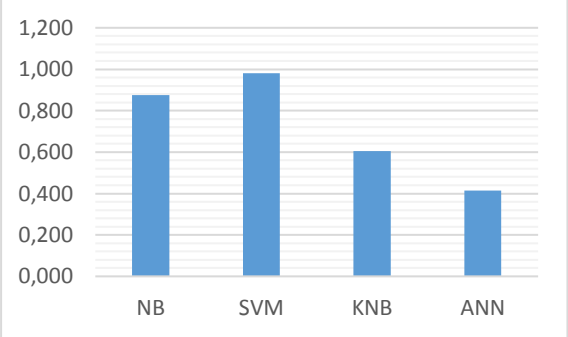


Figure 4 Graph of UAC Comparation Method

IV.2 Discussion

Based on table 1 is information on the results of the proposed method classification based on the value of accuracy, precision, recall, TP rate and TN rate, where the SVM method produces the highest value compared to other methods, while ANN produces the lowest value. the application of the method for the classification of sentiment analysis has a degree of dependence on training and testing datasets.

V. CONCLUSION AND SUGGESTION

Based on the comparative test results of Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNB), Artificial Neural Network (ANN) for the classification of sentiment analysis of UMKM products, North Padang Lawas District produces an accuracy rate of 94% and 0.9 UAC on the SVM method, however due to limited training and testing data sources it is very dependent to produce better accuracy.

1. Acknowledgment
2. Thank you to:
3. 1. Kemenristekdikti who has provided assistance in the form of financial support.
4. 2. Universitas Prima Indonesia, which has provided motivational support and facilities.
5. 3. North Sumatra Province of North Sumatra Province Trade and Industry Office as a partner.

IV. REFERENCES

https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution
http://www.kemenperin.go.id/artikel/17565/Empat-Strategi-Indonesia-Masuk-Revolusi-Industri Keempat
https://www.cnnindonesia.com/ekonomi/20161121122525-92-174080/kontribusi-umkmterhadap-pdb-tembus-lebih-dari-60-persen
http://www.depkop.go.id/pdf-viewer/?p=uploads/tx_rtgfiles/sandingan_data_umkm_20122013.pdf
https://www.weforum.org/agenda/2016/07/what-is-networked-readiness-and-why-does-it-matter
http://lkyspp2.nus.edu.sg/wpcontent/uploads/2016/10/lkysppms_case_study_technical_and_vocational_education_and_training_in_indonesia.pdf
Republik Indonesia. 2008. Undang – Undang No. 20 Tahun 2008 tentang Usaha Mikro, Kecil, Menengah. Sekretariat Negara. Jakarta.
http://www.fmeindonesia.org/optimalisasi-teknologi-digital-dalam-rangka-menerapkan-revolusi-industri-4-0-bagi-usaha-kecil-menengah-ukm-di-indonesia/ diakses pada tanggal 25 Juli 2018
Chen, H., 2012. The Impact of Comments and Recommendation System on Online Shopper Buying Behaviour. Journal of Networks.
Basari, A.S.H., Hussin, B., Ananta, I.G.P., Zeniarja, J. 2013. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Procedia Engineering 53 (2013) 453 – 462.
Saifinnuha, A.Z., 2015, Penerapan Sentimen Analisis pada Twitter Berbahasa Indonesia untuk Mendapatkan Rating Program Televisi Menggunakan Metode Support Vecotr Machine.
Indriati, Ridok, A., 2016, Sentiment analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). Journal of Environmental Engineering & Sustainable Technology.
Hanggara, S., Akhriza, M. TB., Husni, M., 2017, Aplikasi Web Untuk Analisis Sentimen Pada Opini Produk Dengan Metode Naive Bayes Classifier, Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri, ISSN 2085-4218.
Nurjanah, W. E., Perdana, R. S., Fauzi, M. A., 2017, Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada

Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10416
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X, Vol. 1, No. 12, pp 1750-1757.

Rofiqoh, U., Perdana, S.P., Fauzi, M.A., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.

Muljono, Artanti, D. P., Syukur, A., Prihandono, A., De Rosal, D., Setiadi, M., 2018, Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes, Konferensi Nasional Sistem Informasi, pp 165-170.

Haryanto, D. J., Muflikhah, L., Fauzi, M. A., 2018, Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X, Vol. 2, No. 9, pp. 2909-2916.

Lutfi, A. A., Permanasari, A. E., Fauziati, S., 2018, Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine, Journal of Information Systems Engineering and Business Intelligence, Vol. 4, No. 1, pp. 57-64.

Magdalena, H., Irawadi, S., 2018, Optimasi AHP dalam mendukung UMKM di Bangka Belitung dalam memanfaatkan E-Commerce, Jurnal Optimasi Sistem Industri, ISSN (Online) 2442-8795, DOI: 10.25077/josi.v17.n1.p16-25.2018.

Ernawati, S., Wati, S., 2018, Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel, jurnal khatulistiwa informatika, vol. VI, no. 1, e-ISSN: 2579-633X, pp. 58-63.

Taufik, A., 2018, Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel, Jurnal Teknik Komputer AMIK BSI, Volume IV No. 2, E-ISSN: 2550-0120, pp. 112-118.

Liu, B., 2012, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publisher.

Refaeilzadeh, P., Tang, L., & Liu, H. 2009. Cross Validation Editors: M. Tamer dan Ling Liu Encyclopedia of Database Systems. New York: Springer.

Ulwan, M. N. 2016. Pattern Recognition pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association. Skripsi: Program Studi Statistika Universitas Islam Indonesia.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques. Waltham: Morgan Kaufmann.

Informatikalogi. 2016. Pembobotan Kata atau Term Weighting TF-IDF. Retrieved Juli 2018, from Informatikalogi.com: https://informatikalogi.com/term-weighting-tf-idf/.