Utilizing Machine Learning Algorithms to Detect Emotions from Tweets

Christopher Reynard Julian
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
christoper.julian001@binus.ac.id

Gian Reinfred Athevan
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
gian.athevan@binus.ac.id

Nathanael Geoffrey Hanjaya
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
nathanael.hanjaya@binus.ac.id

Henry Lucky
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
henry.lucky@binus.ac.id

Derwin Suhartono
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
dsuhartono@binus.edu

***Abstract* —Emotion detection has been a popular topic for a long time. The need to develop and improve the technologies has grown throughout the year. This is mostly because of the vast implementation in marketing, education, security, healthcare, and other fields. Emotion can be detected variously, from gestures, text, speech, facial expressions, and others. Today, the emotion detection technology is still not perfect, even with the huge access to textual and visual data. This is mostly because of the complexity of human emotions that can't be easily captured by a machine. Therefore, a reliable method is required in order to create an accurate emotion detection system. A few of the popular machine learning methods that are implemented for emotion detection are Support Vector Machine, Naïve Bayes, and Logistic Regression. From our experiment, SVM is considered to be the best methods that have demonstrated high accuracy, especially on a quality data set.**

Keywords—Support vector machine, Naïve bayes, Logistic regression, Emotion detection system

I. Introduction

Human emotion is an important aspect of people's daily life. The meaning of a text or speech can be defined from emotion and therefore, acknowledging emotions is popular nowadays. Emotion detection (ED) is the process of identifying discrete emotions expressed in text and is a part of Affective Computing which is setting computers to have human-like abilities such as observation and interpretation by enhancing their intelligence and research on this topic is still growing according to [1].

There are many implementations and uses for ED, especially in social media. One of the most prevailing usages of ED is in digital marketing and retail [2]. Sellers can use ED to give recommended ads based on someone's current moods to give more accurate advertising and compel the buyer to buy their product. ED can also be used in customer service, by being able to detect emotion in email, the seller can understand what needs to be changed about their product [3]. Another use of ED is in medicine where it can detect mental diseases like dementia or depression [4], particularly on social media such as Twitter posts, so suicide acts can be prevented.

Currently, artificial intelligence (AI) has developed rapidly and has made a significant contribution to providing effective solutions that might be useful in helping the computer to understand human languages. AI can solve difficult issues and create solutions to improve technologies and society [5]. In developing this emotion detection system, the techniques used can be categorized into 2 parts, which are lexical-based approaches and machine learning approaches. Lexical-based approaches are approaches that use keyword recognition (KR) and lexical affinity to build major grammatical and logical rules. This KR technique makes use of lexical resources in the form of emotion dictionaries. Another approach used in developing this ED system is machine learning (ML), which is also one of the technological derivatives of AI. Machine learning itself is a technology that can imitate human intelligence by learning from its surroundings or inputting data related to various desired aspects according to the purpose of its manufacture or construction [6].

Both lexical based and machine learning approaches are also divided into several methods. However, since most papers used machine learning methods, then those machine learning

978-1-6654-5668-5/22/$31.00 ©2022 IEEE

subdivisions are the most often used method in this topic research. The subdivisions of ML approaches are *supervised learning* and *unsupervised learning.* According to [7], supervised machine learning is doing predictions using labeled data or trained datasets where the true values are acknowledged. Meanwhile, unsupervised learning approaches aim to find hidden structures in unlabeled data to build models for emotion classification. Some algorithms usually perform slightly better than other algorithms. For example, in research conducted by [8]) they compared four types of machine learning classifiers, J48 for Decision Tree, Naïve Bayes, K-Nearest Neighbor, and SMO (Sequential Minimal Optimization) for the implementation of SVM. The result was SMO has the highest accuracy rate (85.5%). There is also research by [9] which used Naïve Bayes as their method to classify the dataset and they got an accuracy of 78.6%.

Our main objective is to make a review or summary of all the work that has been done in this area of Emotion Detection by providing major-related topics, approaches used, datasets, and many more. Furthermore, emotion is difficult to understand since it is complicated and it is not easy to distinguish them, especially the complexity of emotional language in human emotions so many factors affect the accuracy [10] Our second objective is to analyze a machine learning algorithm that is efficient and more accurate at classifying emotion into various parts. Therefore, through this paper, we will discuss how to construct the machine learning algorithm and the possibilities in this Emotion Detection system that may occur in the future.

II. Literature Review

Emotion detection is growing more popular in this generation, and it can be useful in many fields, such as detecting insulting words [11]. [12] also researched detecting emotion to recognize suicide ideas and thoughts to prevent them from happening. Another usage of emotion identification is proposed by [13] which uses this technique in brand management to identify potential customers. It can also be implemented to find an application for mental health which is done by [14] using Natural Language Processing (NLP) and Natural Language Generation (NLG).

According to [15], many approaches can be used to do emotion detection, such as the word-based approach, learning-based approach, and hybrid approach. Machine learning classifiers can also be utilized, including Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree, CNN (convolutional neural networks), and many others. [16] stated that most Machine Learning methods are frequently used because they provide superior performance and do not require a direct mention of emotion.

Emotion detection has been evaluated by a lot of researchers, one of the types of research is by [17], which investigated both speech-based emotion recognition and text-based emotion detection. The research showed that even with a lot of different methods of emotion detection, the accuracy of these methods is still not particularly good, especially when detecting sarcasm. Another research that focused on reviewing the current methodology is [18] who also said that emotion detection still has a lot of aspects to improve. This research also suggests that a multimodal system is the best way to get the highest level of precision, for example combining text and facial to detect emotion.

Text-based emotion detection has become more popular these few years. For example, the authors in [19] use several machine learning methods such as Naïve Bayes, Generalized Linear Model, Fast-Large Margin, Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine algorithms to classify the emotion of various English statements from social media. [20] also analyzed the comparison between Naïve Bayes, SVM, KNN, and Decision Tree by using International Survey on Emotion Antecedents and Reactions (ISEAR) dataset. Another research that used the ISEAR dataset is [21]. They used two kinds of datasets, a regular dataset from ISEAR and an irregular dataset from the Open American National Corpus (OANC) dataset. They predict the type of emotion using the proposed ensemble approach which consists of K-Nearest Neighbor, MLP, and Decision Tree algorithms, and compare it with other machine learning methods such as Random Forest, CNN, and a few more. They also use Tree-structured Parzen Estimator to tune the parameters of the basic classifiers.

Moreover, there is also research studied by [22] that analyzes the emotion detection system by searching emotional words from a pre-defined emotional keyword database (phrasal verbs and negative words) and got an overall accuracy of 65%. [23] proposed an approach to identify emotions in text messages using supervised machine learning classifiers. They developed a system called Emotex to create models for the classification of emotion and developed a framework called EmotexStream to classify live streams of tweets for real-time emotion tracking. Another example of emotion detection research is [24] which discussed how emotion embedding models are used to detect emotions in a story text. They collected more than 144 hundred tweets and constructed the emotion embedding model by employing the CNN learning algorithm.

III. METHODOLOGY

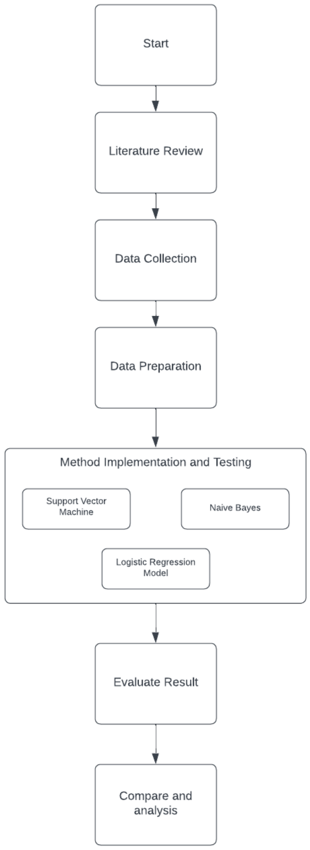

Fig. 1. Method Workflow

The figure above illustrates the flow of our experiments, starting from data collection and ending in comparison and analysis. For further details, it is explained in 3.1.

A. *Overview*

This section will explain further the workflow of our study. Our study starts with a literature review, sorting out the 3 best algorithms for ED, data collection, data preprocessing, modeling, fine-tuning, and evaluation for each algorithm. This literature review was conducted to research the top 3 algorithms. The second step is to collect for training and testing the model. Then, before the data is used to train and test, the data needs to be preprocessed first to allow the data to be understood by the model, prevent errors and make sure no bad data enter the model that can severely affect the accuracy of the model. After that, we moved on to modeling the machine learning algorithms with Support Vector Machines (SVM), Logistic Regression Model (LR), and Naïve Bayes (NB). This model will be trained and fine-tuned. After the model is finished then we evaluate the models with F1 metrics, compare results, and draw conclusions.

B. *Objective*

Our main goal of this study is to classify English sentences into their proper emotions such as joy, love, anger, and fear using several popular machine learning algorithms that we got from doing a literature review. For our dataset, we collected 12.000 English tweets from Twitter with various kinds of emotions. Each data has been labeled with numbers that stand for what emotion they are expressing, 0 is for sadness, 1 is for joy, 2 is for love, 3 is for anger, 4 is for fear, and 5 is for surprise. Furthermore, the data is classified by analyzing the terms or words in each sentence. For instance, if the sentence consists of the word frustrated, annoyed, or mad, we put it in the Anger dataset. An example of the data is shown below.

TABLE I. EXAMPLE OF DATASET CONTENT

Text	Label
"I can have for a treat or if I am feeling festive"	1
"I feel like a faithful servant"	2
"I am just feeling cranky and blue"	3
"I just know to begin with I am going to feel shy about it"	4
"I feel a funny mix of emotions"	5

The table above shows an example of how the content of the dataset is organized. The dataset shows the text and label. The label represents the emotion. For example, "I can have for a treat or if I am feeling festive" with the label of 1 meaning joy.

C. *Dataset*

Below is the figure of the dataset we use in the form of a bar graph, with the number of the emotions available on the dataset according to their emotions:

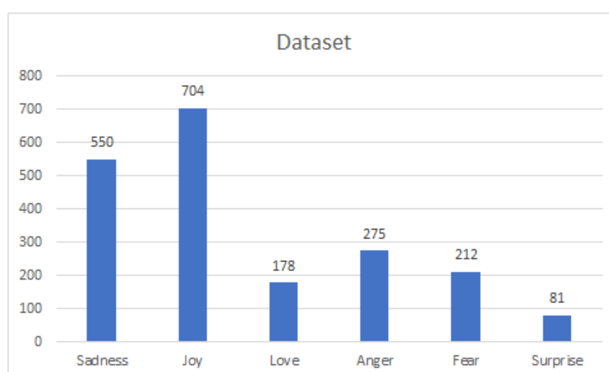

Fig. 2. Number of dataset according to their respective emotion

Based on the figure above, it can be seen that Joy emotion has the most number among all the emotions at about 704, whereas Surprise is the least with only 81 number. Meanwhile, Sadness emotion has the second most number approximately at 550. In addition, the number of Love, Anger, and Fear emotions are 178, 275, and 212 respectively.

D. Data Preparation

The data was then divided into 75% training data and 25% testing data. Before we processed it using the classifiers, we preprocessed the data to increase its quality. The preprocessing techniques include removing symbols, tokenization, lemmatization, removing stop words, and making all the sentences into lower case form.

Symbols, punctuation, and emoticons are removed in the first step because they are irrelevant. Next, we transformed the data into small blocks of words, also removing often used words such as I, the, or not a. Then, the lemmatization process was done by extracting the root word from each term. After the data was preprocessed, we implemented a feature extraction technique which is TF-IDF to improve the classification process. This technique was used to calculate the weight of each word in the data to know how relevant the word is.

E. Method

Three algorithms were used in this study as classifiers, which are SVM, NB, and LR. The first algorithm that we will test is SVM in the Python programming language

SVM is a supervised classification method that originally is a binary classifier but can be used as a multi-class classifier. SVM usually gives high performance in other emotion detection studies, and it does not require any parameter tuning. Initially, we must convert the textual data into the form of a token so that we will be able to identify the emotion words in which it plays a role as the keyword. After that, we need to arrange the text to be taken as the input while the text standing for types of emotion sent will be represented as the output of the text. Furthermore, the data will need to be labeled based on the emotions and we need to create a function to store it. Lastly, it is recommended to train and test several machine learning algorithms to find the best model with the highest accuracy.

The other algorithm that we will test is Naïve Bayes, NB is also a part of the supervised method in machine learning. It relies on the Bayesian Theorem to make a predictive classification model. It is one of the most powerful algorithms since it generates an exact result when implemented for text-based data, making it to be relatively fast and reliable compared to any other algorithms. It predicts membership probabilities for each class such as the probability that the given record or data point belongs to a particular class. The most likely class is defined as the one having the highest probability. Furthermore, the Naive Bayes algorithm has various advantages, consisting of being simple to apply, using less training data, being able to produce real-time predictions, and much more.

The last algorithm that we implemented is LR which is a common supervised machine learning algorithm in pattern recognition. Even though this is not a new method, the LR model is rarely used in detecting text-based emotion, therefore we are trying to implement this method in our study. In the implementation, the logistic regression classifier is applied to classify the emotion from the input text. To create a logistic regression model, we must create a frequency dictionary. The frequency dictionary is the number of occurrences of every word in the data set. This frequency dictionary will be used to extract features that will be used for training the logistic regression model.

IV. Result And Discussion

We have taken 12,000 tweets data from the dataset, and it remains the same after we do several data preprocessing methods. After we filter the data, we split them into 9000 (75%) training data and 3000 (25%) test data and run them using several machine learning algorithms which are SVM, Naïve Bayes, and Logistic Regression Model. Then, we measure the accuracy and the precision, recall, and F1 score of each classifier. Table 2 shows the metrics result of SVM, while Tables 3 and 4 describe the results of using the Naïve Bayes and Logistic Regression Model, respectively.

The best accuracy was achieved by using SVM as the classifier (0.84). Also in this method, Anger received the highest precision (0.87), and Sadness achieved the highest recall (0.91) and F1 Score (0.88). In Naïve Bayes, Fear achieved the best precision (0.83) while Joy received the best recall (0.91) and F1 Score (0.81). The accuracy of this classifier is the lowest, which is only 0.71. With Logistic Regression Model, Fear achieved the highest precision (0.89), Joy received the best recall (0.93), and Sadness got the best F1 Score (0.87). The accuracy for this method is 0.81. Therefore, based on the above findings, it is discovered that SVM outperforms the other algorithms in terms of accuracy and F1-Score.

TABLE II. METRICS RESULT FOR SVM

Emotion	Precision	Recall	F1 – Score
Sadness (0)	0.85	0.91	0.88
Joy (1)	0.83	0.90	0.86
Love (2)	0.80	0.57	0.67
Anger (3)	0.87	0.81	0.84
Fear (4)	0.86	0.81	0.83
Surprise (5)	0.74	0.57	0.64
Average F1-Score			0.78

The table above displays the performance metrics using the Support Vector Machine. The table shows the precision, recall, and F1-score of each emotion. From the results, it can be seen that Surprise is the emotion that the lowest number in all of precision (0.74), recall (0.57), and F1-score (0.64). In addition, Love is at the same level in recall. Meanwhile, the overall accuracy of this method is 0.84.

TABLE III. METRIC RESULTS FOR NAÏVE BAYES

Emotion	Precision	Recall	F1 – Score
Sadness (0)	0.66	0.90	0.76
Joy (1)	0.73	0.91	0.81
Love (2)	0.63	0.19	0.30
Anger (3)	0.78	0.57	0.66
Fear (4)	0.83	0.42	0.56
Surprise (5)	0.67	0.02	0.04
Average F1-Score			0.52

From the table above, it is evident that Love has the lowest precision result at around 0.63, while the lowest result of recall and F1-score is marked by Surprise at 0.02 and 0.04 respectively. The overall accuracy of this method is 0.71.

TABLE IV. METRICS RESULT FOR LOGISTIC REGRESSION

Emotion	Precision	Recall	F1 – Score
Sadness (0)	0.82	0.92	0.87
Joy (1)	0.76	0.93	0.84
Love (2)	0.81	0.48	0.60
Anger (3)	0.88	0.73	0.80
Fear (4)	0.89	0.67	0.76
Surprise (5)	0.85	0.43	0.57
Average F1-Score			0.74

Similar to 2 previous table, the table above also display the result of Precision, Recall, and F1-score. In this method, Joy is the emotion that has the lowest precision at 0.76, whereas Surprise is the emotion that generated the lowest result in recall (0.43) and F1-score (0.57). In addition, the overall accuracy of 0.81

V. CONCLUSION

Human emotions have become one of the most widely used topics for research in the Computer Science area since it uses Machine Learning algorithms in its development. These human emotions can be expressed in many ways, which are facial expressions, speech recognition, or text recognition. In this paper, we are experimenting with text-based emotion detection in which we perform data testing using textual data which is tweets from Twitter as our dataset.

Furthermore, we are managed to conclude that some emotions still lack accuracy, a few factors that affect this are the lack of data in that emotion or the machine learning algorithm's lack of the ability to predict that emotion because of its complexity. An example of this is the surprise emotion, across all machine learning algorithms, they score the lowest F1 – score, especially in Naïve Bayes with only 0.04. mostly because of the lack of data but even with the same data Naïve Bayes scored significantly lower than the other algorithm. While some emotions are still lacking in accuracy, overall accuracy for SVM and Logistic Regression is considerably good, while Naïve Bayes has a slightly lower overall accuracy. The highest overall accuracy is done using SVM with 0.84 accuracies. Thus, according to the result, we found that the model that has the highest accuracy is SVM.

REFERENCES

[1] C.-H. Wu, Y.-M. Huang, and J.-P. Hwang, "Review of affective computing in education/learning: Trends and challenges," *British Journal of Educational Technology*, vol. 47, no. 6, pp. 1304–1323, Nov. 2016, doi: 10.1111/bjet.12324.

[2] D. Al-Hajjar and A. Z. Syed, "Applying sentiment and emotion analysis on brand tweets for digital marketing," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov. 2015, pp. 1–6. doi: 10.1109/AEECT.2015.7360592.

[3] S. L. Happy, A. Dasgupta, P. Patnaik, and A. Routray, "Automated Alertness and Emotion Detection for Empathic Feedback during e-Learning," in *2013 IEEE Fifth International Conference on Technology for Education (t4e 2013)*, Dec. 2013, pp. 47–50. doi: 10.1109/T4E.2013.19.

[4] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, Dec. 2017, pp. 858–862. doi: 10.1109/ISS1.2017.8389299.

[5] M. A. Goralski and T. K. Tan, "Artificial intelligence and sustainable development," *The International Journal of Management Education*, vol. 18, no. 1, p. 100330, Mar. 2020, doi: 10.1016/j.ijme.2019.100330.

[6] I. el Naqa and M. J. Murphy, "What Is Machine Learning?," in *Machine Learning in Radiation Oncology*, Cham: Springer International Publishing, 2015, pp. 3–11. doi: 10.1007/978-3-319-18305-3_1.

[7] D. R. Schrider and A. D. Kern, "Supervised Machine Learning for Population Genetics: A New Paradigm," *Trends in Genetics*, vol. 34, no. 4, pp. 301–312, Apr. 2018, doi: 10.1016/j.tig.2017.12.005.

[8] Muljono, N. A. S. Winarsih, and C. Supriyanto, "Evaluation of classification methods for Indonesian text emotion detection," in *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*, Aug. 2016, pp. 130–133. doi: 10.1109/ISEMANTIC.2016.7873824.

[9] S. Azmin and K. Dhar, "Emotion Detection from Bangla Text Corpus Using Naïve Bayes Classifier," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, Dec. 2019, pp. 1–5. doi: 10.1109/EICT48899.2019.9068797.

[10] M. A. Riordan and L. A. Trichtinger, "Overconfidence at the Keyboard: Confidence and Accuracy in Interpreting Affect in E-Mail Exchanges," *Human Communication Research*, vol. 43, no. 1, pp. 1–24, Jan. 2017, doi: 10.1111/hcre.12093.

[11] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities,"

Engineering Reports, vol. 2, no. 7, Jul. 2020, doi: 10.1002/eng2.12189.

[12] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content," *Complexity*, vol. 2018, pp. 1–10, Sep. 2018, doi: 10.1155/2018/6157249.

[13] F. Greco and A. Polli, "Emotional Text Mining: Customer profiling in brand management," *International Journal of Information Management*, vol. 51, p. 101934, Apr. 2020, doi: 10.1016/j.ijinfomgt.2019.04.007.

[14] R. A. CALVO, D. N. MILNE, M. S. HUSSAIN, and H. CHRISTENSEN, "Natural language processing in mental health applications using non-clinical texts," *Natural Language Engineering*, vol. 23, no. 5, pp. 649–685, Sep. 2017, doi: 10.1017/S1351324916000383.

[15] V. v Ramalingam, A. Pandian, A. Jaiswal, and N. Bhatia, "Emotion detection from text," *Journal of Physics: Conference Series*, vol. 1000, p. 012027, Apr. 2018, doi: 10.1088/1742-6596/1000/1/012027.

[16] S. Al-Saqqa, H. Abdel-Nabi, and A. Awajan, "A Survey of Textual Emotion Detection," in *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, Jul. 2018, pp. 136–142. doi: 10.1109/CSIT.2018.8486405.

[17] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhajj, "Emotion detection from text and speech: a survey," *Social Network Analysis and Mining*, vol. 8, no. 1, p. 28, Dec. 2018, doi: 10.1007/s13278-018-0505-2.

[18] J. M. Garcia-Garcia, V. M. R. Penichet, and M. D. Lozano, "Emotion detection," in *Proceedings of the XVIII International Conference on Human Computer Interaction*, Sep. 2017, pp. 1–8. doi: 10.1145/3123818.3123852.

[19] A. Chowanda, R. Sutoyo, Meiliana, and S. Tanachutiwat, "Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation," *Procedia Computer Science*, vol. 179, pp. 821–828, 2021, doi: 10.1016/j.procs.2021.01.099.

[20] A. F. Ab. Nasir *et al.*, "Text-based emotion prediction system using machine learning approach," *IOP Conference Series: Materials Science and Engineering*, vol. 769, no. 1, p. 012022, Feb. 2020, doi: 10.1088/1757-899X/769/1/012022.

[21] F. Ghanbari-Adivi and M. Mosleh, "Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE)," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8971–8983, Dec. 2019, doi: 10.1007/s00521-019-04230-9.

[22] D. Seal, U. K. Roy, and R. Basak, "Sentence-Level Emotion Detection from Text Based on Semantic Rules," 2020, pp. 423–430. doi: 10.1007/978-981-13-7166-0_42.

[23] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," *International Journal of Data Science and Analytics*, vol. 7, no. 1, pp. 35–51, Feb. 2019, doi: 10.1007/s41060-018-0096-z.

[24] S.-H. Park, B.-C. Bae, and Y.-G. Cheong, "Emotion Recognition from Text Stories Using an Emotion Embedding Model," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Feb. 2020, pp. 579–583. doi: 10.1109/BigComp48618.2020.00014.