

PAPER • OPEN ACCESS

Twitter Users Opinion Classification of Smart Farming in Indonesia

To cite this article: James Nata Salim *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 012165

View the [article online](#) for updates and enhancements.

Twitter Users Opinion Classification of Smart Farming in Indonesia

James Nata Salim^{1*}, Dedi Trisnawarman², Muhammad Choirul Imam³

¹²³⁴Information System, Faculty of Information Technology Universitas Tarumanagara, Jakarta 11440, Indonesia

Abstract. In the past year, Indonesia has digitalized agriculture called "Smart Farming" or "Agriculture 4.0" to follow the era of the industrial revolution 4.0. Public opinion on social media is a strong factor in determining whether or not smart farming is implemented in Indonesia. So the purpose of this research is to find data from Twitter and then analyzed using the sentiment analysis method, which will be classified using Naïve Bayes method. This process starts with searching data on Twitter, preprocessing text, classification, and finally testing. The accuracy testing results gives a value of 0.97, and Recall produces a value of 1.0, F1 produces a value of 0.98 and an AUC value of 0.5.

1. Introduction

Indonesia is a country that is rich in endless natural resources so that it has been dubbed as an agricultural country that depends on the economy of the agricultural sector. In 2011 Indonesian people's rubber plantations covered an area of 2.9 million hectares or around 85% of the total national rubber area with a production of around 80% of the total national natural rubber production [1], palm oil plantations nationally in 2008 had an area covering 7 million hectares, with a production of 19.2 million tons. [2], and so far, Indonesia's Gross Domestic Product (GDP) agriculture sector's GDP has reached Rp 1,005.4 trillion [3]. Some of Indonesia's achievements show that Indonesia has great potential to become the largest producer of agricultural products in the world. But with a population of 266.91 million people [4], Indonesia has not been able to utilize natural resources optimally. The Central Statistics Agency said that the area of paddy fields continues to decline. Their notes in 2018 stated that the area of land was 7.1 million hectares, a decrease compared to 2017, which was still 7.75 million hectares [5].

The development of technology has reached the era of the industrial revolution 4.0, where it began to develop a variety of sophisticated tools. In the agricultural sector also has begun to digitize the so-called Smart Farming or agriculture 4.0 with the aim to simplify and increase productivity throughout the agricultural process. For farmers, the 4.0 industrial revolution era that prioritizes technology should also benefit them, especially to get information and innovations to increase the added value of their agricultural products even to help solve various existing problems [6]. The introduction of industry 4.0 to agriculture began to be introduced in 2018, which is expected to help the Indonesian government to deal with this era [7]. But there are still some challenges that need to be faced, such as the problem of cost, wide area distribution, changing the mindset of traditional farmers, and teaching technology to farmers [6]. From the explanation above it can be seen that the concept of Smart Farming or agriculture



4.0 in Indonesia has pros and cons, so it is needed a reference whether smart farming is suitable to be implemented in Indonesia or not. Of the many social media sites available, users prefer microblogging services like Twitter to learn about product services, social events, and political trends. Twitter is considered as an important source of information in sentiment analysis applications [8]. In this era, social media also participated in expanding the business movement and as a means of marketing agricultural products [9]. Therefore this study uses Twitter data to explore opinions related to smart farming that is applied in Indonesia.

2. Related Work

There had been some research that focuses on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, either positive or negative, or neutral in some cases. A research use various machine learning algorithms like Naïve Bayes, Max Entropy, and Support Vector Machine, they provide research on twitter data streams. Research results show that machine learning methods, such as SVM and naïve Bayes, have the highest accuracy and can be regarded as the baseline learning methods [10]. The second research combines textual and non-textual features to improve the performance of sentiment prediction.

In this research, they apply Naïve Bayes for textual classification and Fisher Score to determine non-textual (like and retweet) features. By combining two kinds of features, they evaluate the performance using F1-measure gives 0.838 of accuracy with α and β are 0.6 and 0.4, respectively [11]. The last one studies on twitter sentiment analysis using machine algorithms. The tweet format is very small, which generates a whole new dimension of problems like the use of slang, abbreviations etc. In this paper, they aim to review some papers regarding research in sentiment analysis on Twitter, describing the methodologies adopted and models applied, along with describing a generalized Python-based approach. This research topic has evolved during the last decade, with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data [12].

3. Method

The method used begins with collect data, preprocessing text, classification, then the last is testing. The details are described as follows.

3.1. Collect Data

The data is the tweets of Twitter users who argue about smart farming in Indonesia. Crawling data on Twitter is done using Python through consumer key access, consumer secret, access key, and access secret obtained after verifying developer accounts or APIs on Twitter. In this process, there are several keywords used to search, such as "smart farming 4.0", "modern agriculture", and "farmer 4.0". For the search process to be carried out, one Python library called Tweepy is needed. After the data is found, it will be saved in CSV file format, then labelling positive or negative sentiments manually.

3.2. Preprocessing Text

- Tokenization. A process of separating text into small pieces called tokens. A token is a meaningful unit of text, such as a word, that we are interested in using for analysis [13]. Tokenization aims to separate or differentiate each term in the text you want to process, so that at the end of this process we get the arrangement of terms.
- Case Folding. This step ensures all uppercase is changed to lowercase and also only letters a through z are contained in the document [14]. This process is important because not all text documents are consistent in writing capital or even writing errors occur.

- Stemming. A process of reducing the words in to root form effectively. Stemming process involves affix removal algorithm which removes prefixes and suffixes of the word in the document [15].
- Removing Stop Words. Stop words are common in various sentiment analysis studies to prepare texts that are easy to process [16]. This process involves removing 126 conjunctions such as “adalah”, “di”, “akan”, etc.

3.3. Classification

Bayesian Classifiers are a popular supervised classification paradigm. Naive Bayes can be seen as in equation (1) and equation (2). An advantage of Naïve Bayes' is that it only requires a small amount of training data to estimate the parameters necessary for classification. In Naïve Bayes technique, the basic idea to find the probabilities of categories given a text document by using the joint probabilities of words and categories [17], [18].

$$P(C/x) = P(x/C)/P(x) \quad (1)$$

Next, an assumption is made that the data point $x = \{x_1, x_2, \dots, x_j\}$. The probability of each attributes occurring is independent, we can estimate the probability of x as follows.

$$P(C/x) = P(C) \cdot \prod P(x_i/C) \quad (2)$$

3.4. Testing

After getting the results, testing is needed to prove the performance and accuracy of these results. There are 5 tests in this study Confusion Matrix, Accuracy, Recall, F1, and AUC.

4. Experiment & Result

4.1. Collect Data

Index	tweet	Sentiment
0	b'Teknologi Smart Farming 4.0 bantu optimalkan hasil pertanian\nhttps://t.co/X3uQRwJqNA https://t.co/n89R6uHREC'	p
1	b'RT @TelkomIndonesia: Selanjutnya Mobility eXperience Area. Sobat akan mendapatkan pengalaman eksklusif masa depan dari mobile technology. S\	P
2	b'Salah satu implementasi konsep Internet of Things (IoT) dengan pemasangan alat RiTX guna mendukung Teknologi Pertan\	P
3	b'Saksikan bingkisan menarik tentang IOT SMART FARMING CILI, JUTAWAN LEMBU & BUAH AJAIB dalam #Agrotek pukul 6.00 pet\	P
4	b'Saksikan bingkisan menarik tentang IOT SMART FARMING CILI, JUTAWAN LEMBU & BUAH AJAIB dalam #Agrotek, pukul 6.00 pe\	P
5	b'Selanjutnya Mobility eXperience Area. Sobat akan mendapatkan pengalaman eksklusif masa depan dari mobile technology\	P
6	b'Saksikan bingkisan menarik tentang IOT SMART FARMING CILI, JUTAWAN LEMBU & BUAH AJAIB dalam #Agrotek pukul 6.00 pet\	P
7	b'@azrawlldn @azlanR Smart Farming. Kalau ada tanah nak tanam cili kita boleh guide https://t.co/NhmrmBAQzC'	p
8	b'@auahhmager @collegemenfess Orang pertanian kalo gabung sama orang IT/elektro keren tuh. Bikin smart agriculture.'	p
9	b'Mengenal start-up digital pertanian karya anak bangsa yang akan unjuk kebolehan di ajang internasional.\	P
10	b'Kami serius melindungi hutan & inisiatif terbaru kami dengan @MUR merupakan langkah ke arah yg benar: mengembangkan\	P

Figure 1. Tweets Before Preprocessing

After crawling data using Python, 174 tweets were found from various users who wrote some opinions on current agricultural developments, can be seen in Figure 1. The "sentiment" column on the right is the result of manual labelling.

4.2. Preprocessing Text

Index	sentiment	tweet
0	1	b teknologi smart farming 4 0 bantu optimal hasil tani nhttps t co x3uqrwjqna https t co n89r6uhrec
1	1	b rt telkomindonesia lanjut mobility experience area sobat akan dapat alam eksklusif masa depan dari mobile technology s xe2 x80 xa6
2	1	b salah satu implementasi konsep internet of things iot dengan pasang alat ritx guna dukung teknologi tan xe2 x80 xa6 https t co dqokg8nnh
3	1	b saksi bingkis tarik tentang iot smart farming cili jutawan lembu amp buah ajaib dalam agrotek pukul 6 00 pet xe2 x80 xa6 https t co lcx1mad...
4	1	b saksi bingkis tarik tentang iot smart farming cili jutawan lembu amp buah ajaib dalam agrotek pukul 6 00 pe xe2 x80 xa6 https t co 6h9yulas...
5	1	b lanjut mobility experience area sobat akan dapat alam eksklusif masa depan dari mobile technology xe2 x80 xa6 https t co npoypuuhpj
6	1	b saksi bingkis tarik tentang iot smart farming cili jutawan lembu amp buah ajaib dalam agrotek pukul 6 00 pet xe2 x80 xa6 https t co 8bwryed...
7	1	b azrawlldn azlanr smart farming kalau ada tanah nak tanam cili kita boleh guide https t co nhmrmBAQzC
8	1	b auahhmager collegemenfess orang tani kalo gabung sama orang it elektro keren tuh bikin smart agriculture
9	1	b kenal start-up digital tani karya anak bangsa yang akan unjuk boleh di ajang internasional xe2 x80 xa6 https t co dy71hhalkq
10	1	b kami serius lindung hutan amp inisiatif baru kami dengan wur rupa langkah ke arah yg benar kembang xe2 x80 xa6 https t co ayuxlhxgbl

Figure 2. Tweets After Preprocessing

It can be seen in Figure 2 that the data tweets after preprocessing text are cleaner because there are not many symbols, conjunctions, uppercase letters, and so on.

4.3. Classification

Figure 3 is the result of the Naïve Bayes Classification. 1 for positive and 0 for negative

Index	0
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	0
0	1

Figure 3. Classification Results

4.4. Testing

Here are the results of the five tests.

Confusion Matrix: `array([[0, 1], [0, 33]])`

Accuracy Score: 0.9705882352941176

Recall Score: 1.0

F1 Score: 0.9850746268656716

AUC Score: 0.5

5. Conclusion

All results in this study showed good results. Even at the level of accuracy obtained can reach 0.97, Recall score reaches 1.0, and F1 also gets a value of 0.985. The response from Twitter users to smart farming in Indonesia also turned out to be positive.

References

- [1] Syarifa, L.F., Agustina, D.S. and Nancy, C., 2013. Evaluasi pengolahan dan mutu bahan olah karet rakyat (bokar) di tingkat petani karet di Sumatera Selatan. *Jurnal Penelitian Karet*, 31(2), pp.139-148.
- [2] <https://kemenperin.go.id/artikel/1075/Indonesia-Produsen-Kelapa-Sawit-Terbesar>
- [3] <https://katadata.co.id/berita/2019/05/11/pertumbuhan-pdb-pertanian-ri-2018-lampau-target>
- [4] <https://databoks.katadata.co.id/datapublish/2019/01/04/jumlah-penduduk-indonesia-2019-mencapai-267-juta-jiwa>
- [5] <https://www.cnnindonesia.com/ekonomi/20181025153705-92-341433/bps-sebut-luas-lahan-pertanian-kian-menurun>
- [6] <https://www.kompasiana.com/diarronayu/5ce515426b07c50853506e9f/menatap-masa-depan-konsep-pertanian-modern-di-indonesia?page=all>
- [7] Dar, M.H., Sarkum, S., Nasution, A.P. and Mustamu, N.E., 2018, November. Strategies to Overcome the Selling Price Fluctuations of Palm Oil Fresh Fruit Bunches (FFB): The Farmers Efforts in Welcoming Industry 4.0. In Proceeding 1st ICUTK International Conference
- [8] Asghar, M.Z., Kundi, F.M., Ahmad, S., Khan, A. and Khan, F., 2018. T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme. *Expert Systems*, 35(1), p.e12233
- [9] White, D., Meyers, C., Doerfert, D. and Irlbeck, E., 2014. Exploring agriculturalists' use of social media for agricultural marketing. *Journal of Applied Communications*, 98(4), pp.72-86.
- [10] Vishal A. Kharde & S.S. Sonawane, 2016. Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications* (0975 – 8887) Volume 139 – No.11, April 2016
- [11] Perdana S.R., Pinandito A., 2018. Combining Likes-Retweet Analysis and Naïve Bayes Classifier within Twitter for Sentiment Analysis. *Journal of Telecommunication, Electronic and Computer Engineering* e-ISSN: 2289-8131 Vol. 10 No. 1-8
- [12] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G. and Badhani, P., 2017. Study of Twitter sentiment analysis using machine learning algorithms on Python. *International Journal of Computer Applications*, 165(9), pp.0975-8887
- [13] Silge, J. and Robinson, D., 2017. Text mining with R: A tidy approach. "O'Reilly Media, Inc."
- [14] Surjandari, I., Naffisah, M.S. and Prawiradinata, M.I., 2015. Text mining of twitter data for public sentiment analysis of staple foods price changes. *Journal of Industrial and Intelligent Information* Vol, 3(3)
- [15] Rani, S.R., Ramesh, B., Anusha, M. and Sathiaseelan, J.G.R., 2015. Evaluation of stemming techniques for text classification. *International Journal of Computer Science and Mobile Computing*, 4(3), pp.165-171
- [16] Jianqiang, Z. and Xiaolin, G., 2017. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access*, 5, pp.2870-2879
- [17] Dey, L., Chakraborty, S., Biswas, A., Bose, B. and Tiwari, S., 2016. Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier. *arXiv preprint arXiv:1610.09982*
- [18] Lina L. Dhande, Girish K. Patnaik, 2014. "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier", *IJETTCS*, Volume 3, Issue 4 July-August, ISSN 2278-6856.