Accuracy for Sentiment Analysis of Twitter Students on E-Learning in Indonesia using Naive Bayes Algorithm Based on Particle Swarm Optimization

Mohamad Kartiko¹,²*, Sfenrianto²

¹Master of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri Jakarta, Indonesia
²Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Universitas Bina Nusantara, Jakarta, Indonesia

Abstract. Students can use social media such as Twitter for online learning (E-Learning). This study aims to analyse an accuracy for the sentiments of students about E-Learning who use Indonesian on Twitter social media both positive and negative opinions. The algorithms used are Naive Bayes (NB). Then to optimize the accuracy of the calculation results, the Particle Swarm Optimization (NB-PSO) approach is used. In order to optimize accurate results, this study uses three experimental sequences (scenario 1, scenario 2, and scenario 2) for both NB and NB-PSO algorithms. Each scenario uses different positive and negative comments. The results of the experiment show that in scenario 1 an increase in accuracy is 10.00% for NB-PSO. Scenario 2 there is an increase in accuracy of 13.33% on NB-PSO. Meanwhile, in scenario 3 an increase in accuracy is 27.22% for NB-PSO. These results prove that the accuracy of NB-PSO is better than NB for all scenarios.

1. Introduction

Social media is widely used by the public for socialization. In addition, students also use social media to support online learning (E-Learning). Understanding and improving e-learning systems can use social media strategies [1]. Social media for e-Learning also increases student academic achievement [2] [3].

To find out the opinions of students in using social media in e-Learning sentiment analysis is needed. This analysis aims to investigate whether people consistently show more positive or negative sentiments in the social media messages that they use to serve as the basis of policy [4]. Students who use social media in online learning can be used to find out how they think about e-Learning.

The approach to sentiment analysis can use various algorithms, such as Naive Bayes, the Particle Swarm Optimization (NB-PSO) and so on. The Naive Bayes (NB) algorithm has been used for sentiment analysis [5], such as opinions on social media [6], e-learning review [7], and so on.
Likewise, the Particle Swarm Optimization (NB-PSO) algorithm has used sentiment analysis to find out the opinions of social media users on Twitter [8], review learning on e-Learning [9].

The purpose of this study is to analyze sentiment on Twitter social networks with Indonesian e-Learning queries using Naive Bayes (NB). Then to obtain accuracy optimization, the application of the Particle Swarm Optimization (NB-PSO) feature selection method is also used. The study expected understanding student behavior based on their opinions on social media Twitter to help instructors improve teaching in e-learning systems.

2. Basic Theoretical

Although using sentiment analysis in social media is a recent research field, there are many works in this area, one is in the field of e-Learning. Sentiment analysis is research related to opinions or comments using various algorithms [10]. Naive Bayes is one algorithm that can be used for sentiment analysis that explains opinions in various fields [5]. One example is the opinion of students who use social media in e-Learning [6] [7].

Nowadays learning strategies using social media in e-Learning have begun to be used to improve the learning process [3]. But it is necessary to analyses student social media users' comments on e-Learning to find out their opinions. The results of the opinion should be seen how much the level of accuracy.

Sentiment analysis can be seen from the results that are more accurate. Previous research has resulted in the accuracy of the Naive Bayes and PSO methods. The use of the Naive Bayes algorithm achieves an accuracy of 79.12% [11] [12]. Meanwhile, the Particle Swarm Optimization (NB-PSO) produces an accuracy of 87.91% [11] [12].

3. Research Methodology

The data used is student comments from social networking twitter on E-learning that uses Indonesian. The data is obtained from data collection in January - March 2018 through Rapid miner software. Table 1 shows example of positive and negative comments on student twitter related to E-learning.

| Table 1. For example positive and negative comments in Indonesian |
|------------------|------------------|
| **Positive**     | **Negative**     |
| Pendidikan berkualitas butuh media E-Learning | Kampus gue minggu dpn gaada kuliah tatap muka tp e-learning eaaaaaah gaya beud pdhl cm mahasiswa biasa bukan kelas karyawan |

This study will compare the highest accuracy by dividing into three scenarios. Each scenario consists of 180 positive and negative Twitter comments. The first scenario uses 120 positive comments, and 60 negative comments. The second scenario is 90 positive comments and 90 negative comments. The third scenario is 60 positive comments and 120 negative comments.

Sentiment Analysis can refer to social media users in e-Learning. Naive Bayes Algorithm has been used for this analysis, but need to know the accuracy results. The study, Naive Bayes algorithms using Particle Swarm Optimization (NB-PSO) feature selection will be used to determine accuracy. The use of NB-PSO is expected to produce higher accuracy than without PSO. Accuracy can be calculated based on the following formula1, where: TP = True Positive; TN = True Negative, FP = False Positive, and FN = False Negative.
\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)
\]

4. Results

4.1 Naive Bayes Accuracy

Table 2, shows the results of Naïve Bayes accuracy for all scenarios (scenario1, scenario2, and scenario 3).

| Scenario   | TP  | TN  | TP  | TN  | TP  | TN |
|------------|-----|-----|-----|-----|-----|-----|
| FP         | 119 | 50  | 88  | 76  | 58  | 106 |
| FN         | 1   | 10  | 2   | 14  | 2   | 14  |
| Accuracy   | 71.67% | 56.67% | 40.00% |

The first scenario, the comment data generated by RapidMiner with the Naive Bayes algorithm shows that positive comments match the predictions of 119 data (TP). Negative comment data included in positive predictions is 50 data (TN). Positive comment data included in the negative prediction is 1 data (FP), and negative comment data is included in the negative prediction of 10 data (FN).

The second scenario, the comment data generated by RapidMiner with the Naive Bayes algorithm shows that positive comments match the predictions of 88 data (TP). Negative comment data included in positive predictions is 76 data (TN). Positive comment data included in the negative prediction is 2 data (FP), and negative comment data is included in the negative prediction of 14 data (FN).

The third scenario, the comment data generated by RapidMiner with the Naive Bayes algorithm shows that positive comments match the predictions of 58 data (TP). Negative comment data included in positive predictions is 106 data (TN). Positive comment data included in the negative prediction is 2 data (FP), and negative comment data is included in the negative prediction of 14 data (FN).

The accuracy of the first scenario is 71.67%, second = 56.67%, and third = 40.00%. Thus, the results of accuracy obtained for each scenario indicate that the more positive comments or the fewer negative comments, the greater the results of accuracy.

4.2 Naive Bayes Accuracy based on PSO

Table 3, shows the results of Naïve Bayes accuracy based on PSO for all scenarios (scenario1, scenario2, and scenario 3).

| Scenario   | TP  | TN  | TP  | TN  | TP  | TN |
|------------|-----|-----|-----|-----|-----|-----|
| FP         | 105 | 18  | 76  | 40  | 43  | 42 |
| FN         | 15  | 42  | 14  | 50  | 17  | 78 |
| Accuracy   | 81.67% | 70.00% | 67.22% |

The first scenario, the comment data generated by RapidMiner with the Naive Bayes algorithm base on PSO shows that positive comments match the predictions of 105 data (TP). Negative comment data included in positive predictions is 18 data (TN). Positive comment data included in the negative prediction is 15 data (FP), and negative comment data is included in the negative prediction of 42 data (FN).

The second scenario, the comment data generated by RapidMiner with the Naive Bayes algorithm base on PSO shows that positive comments match the predictions of 76 data (TP). Negative comment
data included in positive predictions is 40 data (TN). Positive comment data included in the negative prediction is 14 data (FP), and negative comment data is included in the negative prediction of 50 data (FN).

The third scenario the comment data generated by RapidMiner with the Naive Bayes algorithm base on PSO shows that positive comments match the predictions of 43 data (TP). Negative comment data included in positive predictions is 42 data (TN). Positive comment data included in the negative prediction is 17 data (FP), and negative comment data is included in the negative prediction of 78 data (FN).

The accuracy of the first scenario by RapidMiner with the Naive Bayes algorithm base on PSO is 81.67%, second = 56.67%, and third = 40%. Thus, the results of accuracy obtained for each scenario indicate that the more positive comments or the fewer negative comments, the greater the results of accuracy. Based on the results of testing the data for all scenarios, it is evident that the application of Naive Bayes based on Particle Swarm Optimization (PSO) to feature selection can improve accuracy on the Naive Bayes algorithm. Table 4 and Figure 1 show the comparison of the results.

### Table 4. Comparison of accuracy and increase between NB and NB-PSO

| Scenarios | Naive Bayes (NB) | Naive Bayes based on Particle Swarm Optimization (NB-PSO) | Increase Percentage |
|-----------|-----------------|--------------------------------------------------------|---------------------|
| 1         | 71.67%          | 81.67%                                                 | 10.00%              |
| 2         | 56.67%          | 70.00%                                                 | 13.33%              |
| 3         | 40.00%          | 67.22%                                                 | 27.22%              |

**Figure 1.** Comparison of accuracy between NB and NB-PSO

Based on table 4 and figure, the comparison scenario1, scenario2, and scenario3 for NB and NB-PSO. The results show that in scenario 1 an increase in accuracy is 10.00% for NB-PSO. Scenario 2 there is an increase in accuracy of 13.33% on NB-PSO. Meanwhile, in scenario 3 an increase in accuracy is 27.22% for NB-PSO. Thus, the results of accuracy obtained for each scenario indicate that the more positive comments or the fewer negative comments, the smaller the results of accuracy.

### 5. Conclusion

Based on three scenarios (scenario1, Scenario2, and scenario3) results, NB-PSO is effective to improve the accuracy of the Naive Bayes algorithm. The result recorded by NB-PSO is 10.00% for scenario1, 13.33% for scenario2, and 27.22% for scenario3. The results NB-PSO have improved the accuracy as compared to the existing study [1] [12]. In the future, we want to develop a recommendation system for sentiment analysis of social media types such as Facebook, Instagram, WhatsApp, etc. Also the use of Naive Bayes (NB) and the Particle Swarm Optimization (NB-PSO).
6. References

[1] Elkaseh, A. M., Wong, K. W., & Fung, C. C. (2016). Perceived ease of use and perceived usefulness of social media for e-learning in Libyan higher education: A structural equation modeling analysis. *International Journal of Information and Education Technology*, 6(3), 192

[2] Alwagaït, E., Shahzad, B., & Alim, S. (2015). Impact of social media usage on students academic performance in Saudi Arabia. *Computers in Human Behavior*, 51, 1092-1097.

[3] Linda, W. Sfenrianto, (2018). The Impact of Facility, Social media and Learning Material on Student Engagement Master in Technology Online Student in Indonesia, *Journal of Theoretical & Applied Information Technology*, 96(13).

[4] Bašić, S., Spijkers, B. S., Ouwerkerk, J. W., Vermeulen, I. E., van Koningsbruggen, G. M., & Fokkens, A. S. (2018, May). Assessment of an Optimistic Bias in Social Media Messages. In *WebSci’18 Main Conference Poster Session*.

[5] Chen, S., Ding, Y., Xie, Z., Liu, S., & Ding, H. (2018, April). Chinese Weibo sentiment analysis based on character embedding with dual-channel convolutional neural network. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 107-111). IEEE.

[6] Yadav, R., & Gupta, V. (2018). Self-Harm Prevention Based On Social Platforms User Data Using Naive Bayes Classifier. *Journal of Data Mining and Management*, 3(2).

[7] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

[8] Duwairi, R. M., Marji, R., Sha’ban, N., & Rushaidat, S. (2014, April). Sentiment analysis in Arabic tweets. In *Information and communication systems (icics), 2014 5th international conference on* (pp. 1-6). IEEE.

[9] El-Halees, A. (2009). Mining students data to analyze e-Learning behavior: A Case Study.

[10] Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2, 627-666.

[11] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).

[12] Dulhare, U. N. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomedical Research*, 29(12), 2646-2649.