Strategies for Improving the Quality of Logistics Courier Services Through Priority Problem-solving Based on Multiclass Classification

R Hendayani¹*, M C Dharmawan²**
School of Economics and Business, Telkom University, Bandung, Indonesia

Email: *ratihhendayani@telkomuniversity.ac.id,
**muhammadcahyo@student.telkomuniversity.ac.id

Abstract. PT Tiki Jalur Nugraha Ekakurir (JNE) is the most popular courier service in the country and ranked the first in the category of courier service companies in Indonesia Top Brand 2018. Despite being a top brand, JNE still faces a variety of complaints posted by customers on the internet and social media. This makes it interesting to study as a means of evaluating and finding strategies to improve the quality of logistics services using a sentiment analysis method with classification models of Naive Bayes Classifier and Support Vector Machine as well as multiclass classification. This study aims to review and determine which classification model is the most appropriate for the dataset, find out the rating of sentiments and dimensions in order to measure the quality of logistics company services and identify what problems need to be prioritized. The population in this study are all the tweets mentioned to JNE's Twitter accounts. Tweet samples in this study were collected over a 30-day period, from 1 September 2019 to 30 September 2019, totalling 11,134 tweets. This study revealed that JNE is dominated by negative sentiment of 97.82% and positive sentiment of 2.18%. The main problems (high priority) faced by JNE are the dimensions of Personnel Contact Quality and Timeliness. Then, the dimensions of Information Quality, Order Discrepancy Handling and Order Accuracy are problems with medium priority. Meanwhile, JNE's low priority issues are the dimensions of Order Conditions, Ordering Procedures, Ordering Release Quantities, and Order Quality. Therefore, JNE needs to improve the examination of personal messages and provide a quick response to complaints about customer problems and increase commitment in every delivery service to match the estimated delivery time. It is recommended in future studies to apply the Naive Bayes Classifier classification model in Twitter dataset processing because it has fast, good, and quite precise processing results. Future research can also use the application of ABC analysis which can be used as a basis for determining the number of priority dimensions of a service's problem solving.

1. Introduction
At present, there are many freight forwarding companies competing with each other to win the market. The Indonesian Logistics and Forwarders Association (ALFI) predicts that the logistics business in the country can grow to more than 30% in 2020. This significant growth is due to a drastic upward trend in e-commerce. According to Top Brand Award survey data, JNE ranked the first in the category of courier services in Indonesia's Top Brand 2018, beating some of its rivals in the second phase with a Top Brand Index level of 45%. If JNE wants to remain the best, it must compete even tighter in the future. Behind its reputation as the number 1 courier service in Indonesia, JNE is not free from complaints as posted by the customers on the internet and social media. It is very important for
companies to care about their consumers and maintain the quality of service [1]. Therefore, JNE as a Top Brand courier must maintain and monitor the level of quality of their logistics services in order to remain leading in the industry through a strategy to improve the quality of JNE logistics services based on the classification of JNE logistics service problems so that it can clearly illustrate the priority of the problems being faced by JNE.

The quality of service in this study is more specific to Logistics Service Quality (LSQ). LSQ is intended to identify the services of the courier service provider company (logistics delivery) to fit the research object. LSQ explains that customer perceptions of LSQ are based on the results of an ongoing evaluation (series) of the logistical process interactions from the courier service providers [2]. Previous studies also used LSQ to validate the scale of quality of logistics services in the Indian logistics industry [2]. There are 9 dimensions of LSQ that are [2]: Information quality is a dimension related to various information provided by courier services. Ordering procedures is a dimension related to courier service procedures for receiving orders from customers. Ordering release quantities is a dimension related to courier service policies regarding shipping/delivery quantity. Timeliness is a dimension related to the evaluation of courier services based on the time taken to get to the destination of delivery. Order accuracy is a dimension related to the accuracy of the courier service for orders taken from customers. Order quality is a dimension related to quality standards of courier services to customers. Order conditions is a dimension related to the ability of courier services in maintaining the condition of an item sent in anticipation of possible damage to the goods. Order discrepancy handling is a dimension related to handling various problems during courier service to customers. Personnel contact quality is a dimension related to benefits or clear explanations obtained by customers through personal contact with courier services and about the attitude of courier personnel towards customers. This study also uses sentiment analysis to determine the positive and negative opinions of customers, whereas previous studies also used sentiment analysis to measure the quality of e-commerce services from online customer reviews [3]. Sentiment analysis is a study of opinions, sentiments, evaluations, and emotions of people towards an entity such as products, services, organizations, individuals, issues, topics and other attributes that represent a sizable problem [4] Customer opinions contained in social media, blogs, discussion forums, the web, tweets, and other forms of media are referred to as User Generated Content (UGC) [5]. In addition, this study also uses ABC analysis to classify the priority of JNE logistical services. This ABC analysis divides the group into three categories based on the priorities of the problems. Category A contains top priority problems (high priority) which have 20% proportion of the total variables, Category B consists of medium priority problems (medium priority) with 30% proportion of the total variables, and Category C is made up of lower priority problems with 50% proportion of the total variable (low priority) [6].

Customers' opinions on the quality of logistics courier services can be used as a source of interesting and innovative ideas for the development of courier services [7]. Customer opinions can also be found on social media. Thus, the problems of online companies can be analyzed using the most popular social media, Twitter [8]. With reference to the explanation above, this study used Twitter to analyze customer responses and opinions on the quality of JNE logistics services. Data were processed by the method of sentiment analysis and multiclass classification to better understand customer opinions and JNE's position. Negative sentiment serves to identify problems and set priorities for problem solving while multiclass classification classifies the problems of the courier service using the ABC analysis. The use of ABC analysis is the contribution of this paper, that show the managerial focus based on the big data analysis result. The aim is to identify the level of service satisfaction, spot the quality of logistics services, and improve the accuracy of decision making.

2. Methods

Research framework and the stages of the study were adopted from the research of Alamsyah and Rachmadiansyah [8] regarding the mapping of the quality of transportation services and combine with the Logistic. The use of ABC analysis in this research is to determine the priority of problems that are urgent, quite urgent, and not urgent for JNE companies so that they can be used as a basis for determining strategies to improve JNE service quality. The advantages of using ABC analysis are that
it is easy to use, universal in nature to determine priorities using a variety of indicators, and can be used as a managerial tool in management [9]. The framework of this study is shown in Figure 1.

![Figure 1. Research Framework](image)

Problem identification describes problems found based on the phenomena occurring at JNE company. The purpose of this research is to determine the rating of sentiments and dimensions in order to measure the quality of JNE logistics services and find out issues that need to be prioritized based on customer opinions classified into 9 dimensions of LSQ. The research method uses a big data approach, especially sentiment analysis with multiclass classification. The focus is to choose which classification model is better in the process of classifying customer opinions from the Twitter dataset, Naive Bayes Classifier (NBC) or Support Vector Machine (SVM). The tweet data was collected using the R-Studio application by crawling the tweets that mentioned to @JNECare and @JNE_ID so the population in this study includes all such tweets found in the two Twitter accounts above. The sample of this study was 11,134 tweets obtained in a period of 30 days from 1 September 2019 to 30 September 2019. Irrelevant Tweets like those not written in Indonesian, contained advertisements or did not contain sentiments and dimensions were filtered by removing them. In the next stage, tweets are categorized into positive and negative sentiments, and into certain dimensions as in Table 1 and 2.

| Table 1. Classification of Tweets Based on Dimensions |
|---------------------------------|
| **Tweets** | **Dimension** |
| --- | --- |
| @JNECare, the following receipt number JNCL-7646014811 information on tracking is incomplete address, but the recipient received an sms with the full address, and the address is often used, please help..thxyou | Information Quality |
| @JNECare, Are you sure it was delivered directly to me? As far as I know, from the warehouse, ordinary goods are taken to the Garuda warehouse first, then delivered to the consumers, right? I am familiar with the central Jakarta JNE plot!! Sorry. I was upset reading the answer of the JNE admin who said that the weight was tracked at 2kg and the cost was 110,000 Really outrageous @JNE_ID. I chose the package yes 1 day, hoping that my item arrived at the destination quickly. But the fact that I got was my item did not arrive in 3 days ... and the worse thing was I still had to pick it up from the warehouse..ckck ... saluut..salut .. If I have not taken it myself, it might take 5 days to reach me ... so what are the different rates for???
@JNECare, How come I was told to revise, you sent the wrong one. On the receipt was clearly written Sukun Malang. Why did you send it to Dau Malang Regency?! Yes, the sender has already bubble-wraped it, min. Is it still impossible to be delivered without wood packing? @JNECare, we want to complain that my package is broken, what's your solution? How can the professionalism of a company like @JNE_ID @JNECare be this bad in responding to shipping complaints that are the customer's right? Since Thursday, I have complained, and it is now Saturday but there is no meaningful action, I don't even know the position of my item now. @JNECare, Min, please check the DM, it's important | Ordering Procedures |
| | Ordering Release Quantities |
| | Timeliness |
| | Order Accuracy |
| | Order Quality |
| | Order Conditions |
| | Order Discrepancy Handling |
| | Personel Contact Quality |
Table 2. Classification of Tweets Based on Sentiments

| Tweets                                                                 | Dimension |
|-----------------------------------------------------------------------|-----------|
| Hi all, thank you for the good cooperation, thank God the package has been well received by my customers. | Positive  |
| I lost one more item in jne ... There is no karang besuki in Dau Malang!!!!! What's wrong with your couriers? @JNECare!! | Negative  |

The next stage is the preprocessing of data in rapidminer by eliminating useless items such as emojis and improving the structure of sentences or words through stages of tokenization, filtering, and stemming so that the data is ready to be further processed in machine learning [8]. To measure how important and relevant a word is in a document, weighting is carried out using the Term Frequency Inverse Document Frequency (TF-IDF) procedure on the preprocessing operator in rapidminer. This process is important for improving classification performance [8].

After filtering out irrelevant tweets and pre-processing stages, the net data obtained were 5,036 tweets. Training data and test data are made in an optimal level of accuracy, precision and recall with a ratio of 70:30 [10]. The classification techniques used were NBC and SVM. NBC is one of the simplest Bayesian learning models in the classification method that uses teorema bayes to predict the probability of features (data) that have been labelled according to predetermined categories [11]. The second classification model is SVM, a model that is only built from the contribution of selected training data used in classification [12]. After that, the performance of the two models is compared and evaluated to determine the best classification model. Text performance evaluation is a method of measuring the performance of a classification model after being tested with a machine. One way to measure it is through the use of a confusion matrix [13]. Confusion matrix is a table that records the results of classification. The amount of data correctly classified provides us with information about the accuracy of the predicted results and the amount of data incorrectly classified allows us to know the error rate of the prediction made [12]. The results obtained from the matrix are then tested using the formulas shown in Table 3 and also with kappa calculations aiming to determine the level of accuracy or the strength of agreement / reliability in the classification model [14].

Table 3. Calculation of Classification Model Performance

| Measure       | Formula                |
|---------------|------------------------|
| Precision Positive | \( \frac{TP}{TP + FP} \) |
| Precision Negative | \( \frac{TN}{TN + FN} \) |
| Recall Positive   | \( \frac{TP}{TP + FN} \) |
| Recall Negative   | \( \frac{TN}{TN + FP} \) |
| Accuracy          | \( \frac{TP + TN}{TP + TN + FP + FN} \) |
| F-Measure     | \( \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \) |
| Kappa           | \( \frac{P_0 - P_e}{1 - P_e} \) |

Precision is the correct percentage of positive / negative estimates for reviewing the accuracy of information. Recall is the percentage of positive / negative labels to review the completeness of the data. Accuracy is a review of how precise the classification is made from the overall amount of data predicted while the f-measure is a metric that combines the values of precision and recall with the aim
to determine the balance between recall and precision that are equally important [13]. The last stage is called results and interpretation, here the research questions are answered and the final conclusions of the study along with solutions that benefit the company are given.

### 3. Results and Discussion

The following are the results of the performance measurement of the Naive Bayes Classifier (NBC) and the Support Vector Machine (SVM) using the rapidminer application. The performance of the NBC model is shown in Table 4 and SVM in Table 5. The two classification models will be compared to determine the best classification model.

#### Table 4. Performance of NBC Classification Model

| Accuracy | Recall | Precision | F-Measure | Kappa |
|----------|--------|-----------|-----------|-------|
| 96.45%   | 71.89% | 60.68%    | 65.81%    | 0.279 |

#### Table 5. Performance of SVM Classification Model

| Accuracy | Recall | Precision | F-Measure | Kappa |
|----------|--------|-----------|-----------|-------|
| 98.24%   | 52.90% | 64.19%    | 58.00%    | 0.094 |

Tables 4 and 5 show the measurement results of the two classification models. When viewed from the accuracy values, both models have very good accuracy because they are in the range of 0.90 - 1.00, namely 96.45% and 98.24% which are categorized as excellent classification [15]. When viewed from the values of recall and precision, they appear to be inversely proportional to each other. Therefore, the f-measure value is taken as the middle value between precision and recall. The result is that the Naive Bayes Classifier's f-measure value is higher than SVM as a harmonization of the precision and recall values which indicate that the classification process is running quite well. In addition, the kappa value of NBC is also higher than SVM. The NBC kappa has a value of 0.279 which falls into the fair agreement category (quite right) because the kappa value is in the range of 0.21 - 0.40 [16]. On the other hand, the SVM kappa has a value of 0.094 which is lower than that of the NBC, this makes it fall into the slight agreement category (a bit accurate) because it is in the range of 0.01 - 0.20 [16]. The explanation above proves that the NBC classification model has better performance and can provide good classification accuracy results and is more precise than the SVM classification model. Based on the results of the sentiment rating on the quality of JNE courier logistics services from the tweets collected within 30 days from 1 September 2019 to 30 September 2019, it was revealed that negative sentiment results were 97.82% and positive ones were 2.18% as presented in Table 6.

#### Table 6. Sentiment Rating Results of JNE's Courier

| Total Net Tweets | Positive | Negative |
|------------------|----------|----------|
| 5,036            | 110 (2.18%) | 4,926 (97.82%) |

Customer tweets are classified into sentiments and dimensions of the quality of logistics courier services to identify problems in their services through negative customer opinions. The dimension with the highest number of negative sentiments becomes the priority dimension for problem solving. Table 6 is illustrating the total number of negative and positive sentiments on JNE courier services successfully classified through the Naive Bayes Classifier classification method.

Table 6 shows that JNE courier logistics services currently have a variety of quality problems with the presence of dominant negative complaints. Negative complaints arise due to customer dissatisfaction with JNE services. Complaints occur because the company is providing services at a level lower than what can be realistically expected. Negative complaints are one of three customer responses due to dissatisfaction of a service [1]. In addition to the rating results regarding overall tweet sentiments, the following also presents the proportion of tweets and positive sentiment and negative sentiment for each dimension of logistics service quality from the most to the least in the JNE dataset in Table 7.
Table 7. Proportion of Tweets in Dimensions

| Quality Dimensions of Logistics Services | Proportion of Tweets | Percentage | Positive Sentiment | Negative Sentiment |
|------------------------------------------|----------------------|------------|--------------------|--------------------|
| Information Quality                      | 1,051                | 20.87%     | 16                 | 1,035              |
| Order Accuracy                           | 489                  | 9.71%      | 24                 | 465                |
| Order Conditions                         | 66                   | 1.31%      | 2                  | 64                 |
| Order Discrepancy Handling               | 581                  | 11.54%     | 6                  | 575                |
| Order Quality                            | 21                   | 0.42%      | 9                  | 12                 |
| Ordering Procedures                      | 34                   | 0.68%      | 3                  | 31                 |
| Ordering Release Quantities              | 22                   | 0.44%      | 1                  | 21                 |
| Personnel Contact Quality                | 1,625                | 32.27%     | 24                 | 1,601              |
| Timeliness                               | 1,147                | 22.78%     | 25                 | 1,122              |

Table 7 shows that the dimensions of the quality of logistics services with the highest proportion of tweets in Twitter's net dataset is Personnel Contact Quality covering a total of 1,625 tweets with a percentage of 32.27% while the dimension with the least proportion of tweets is Order Quality which only has 21 tweets or 0.42%. Table 7 also shows that the dimension that has the most negative sentiments is Personnel Contact Quality with a total of 1,601 tweets. The order of dimensions from the one with the most negative sentiments to the least is Personnel Contact Quality, Timeliness, Information Quality, Order Discrepancy Handling, Order Accuracy, Order Conditions, Ordering Procedures, Ordering Release Quantities, and Order Quality. After knowing the various proportions of negative sentiments on the dimensions of logistics service quality, ABC analysis was carried out to determine the priorities for solving JNE logistics courier service problems as presented in Table 8.

Table 8. Priority Problems Based on ABC Analysis

| Quality Dimensions of Logistics Services | Number of Negative Sentiments | Category          |
|------------------------------------------|-------------------------------|-------------------|
| Personnel Contact Quality                | 1,601                         | A (High priority) |
| Timeliness                               | 1,122                         |                   |
| Information Quality                      | 1,035                         | B (Medium priority)|
| Order Discrepancy Handling               | 575                           |                   |
| Order Accuracy                           | 465                           |                   |
| Order Conditions                         | 64                             | C (Low priority)  |
| Ordering Procedures                      | 31                             |                   |
| Ordering Release Quantities              | 21                             |                   |
| Order Quality                            | 12                             |                   |

Table 8 provides information about the results of ABC analysis categorization sorted by priority of problems. It is clear that the dimensions of Personnel Contact Quality and Timeliness are high priority issues in the quality of JNE logistics services with negative sentiments of 1,601 tweets and 1,122 tweets. The problems of high priority and medium priority and the solution are presented as follows in Table 9 and Table 10.

Table 9. High Priority Problems

| Dimensions | Problems | Mentions | Solutions |
|------------|----------|----------|-----------|
| Personnel  | Long     | “Direct Message” (DM) 1,349 times, “Check” 1,146 times, “Admin” 747 times, “Reply” 111 times, “Answer” 55 times, and “Response” 28 times. | Enhance checking of personal messages sent to JNE official accounts and provide quick and responsive replies to complaints submitted by customers through JNE’s official personal contacts. |
Low priority issues in the quality of JNE logistics services are the dimensions of Order Conditions, Ordering Procedures, Ordering Release Quantities, and Order Quality. Being in the lower priority problem category means that the 4 dimensions are not too significant and not urgent in the case of JNE customer complaints. Even so, JNE courier companies are also required to maintain and improve the quality of their logistics services in all four dimensions by scheduling repairs. The categorization of high priority, medium priority, and low priority problems using ABC analysis in this study aims to identify urgent problems needing to be addressed immediately (high priority), problems that are quite urgent (medium priority), and problems that can wait for scheduled repairs because they are in non-urgent position (low priority).

4. Conclusion
Twitter user responses to the quality of JNE logistics services from September 1, 2019 to September 30, 2019 were dominated by complaints with the negative sentiments of 97.82% while positive responses were only 2.18%. The best classification model in the process of classifying customer opinions from the JNE twitter dataset is the Naive Bayes Classifier (NBC) model because it is superior in terms of recall, f-measure and kappa value compared to the Support Vector Machine classification model. In addition, the level of accuracy in the NBC classification model is also excellent. The dimensions in the quality of JNE logistics courier services with problems to be fixed immediately are divided into 3 priority categories of problems, namely: The category of high priority issues for JNE courier service improvements includes Personnel Contact Quality and Timeliness. The category of medium priority problems consists of Information Quality, Order Discrepancy Handling and Order Accuracy. The category of low priority issues is made up of Order Conditions, Ordering Procedures, Ordering Release Quantities, and Order Quality. Based on the results of this study, it is suggested that
future studies can apply the Naive Bayes Classifier classification model in processing Twitter datasets because it has fast, good, and quite precise processing results. In addition, further research can also use the application of ABC analysis as a basis for determining the number of priority dimensions of a service's problem solving.

References

[1] Pujiah, I. A., & Fatmawati, I. 2018. Pengaruh Pelayanan yang Gagal terhadap Respon Perilaku Konsumen. Jurnal Manajemen dan Pemasaran Jasa, 11(1), pp. 1-20.

[2] Kamble, S. S., Raut, R. D., & Dhume, S. M. 2011. Validating the logistics service quality (LSQ) scale in Indian logistics industry. In International Conference on Business and Economics Research 1, pp. 81-85.

[3] Sari, P. K., Alamsyah, A., & Wibowo, S. 2018. Measuring e-Commerce service quality from online customer review using sentiment analysis. In Journal of Physics: Conference Series 971(1), p. 012053. IOP Publishing.

[4] Liu, B. 2012. Sentiment Analysis and Opinion Mining Morgan & Claypool Publishers. Language Arts & Disciplines, 167.

[5] Moens, M. F., Li, J., & Chua, T. S. (Eds.). 2014. Mining user generated content. CRC press.

[6] Heizer, J., & Render, B. 2015. Manajemen Operasi: Manajemen Keberlangsungan dan Rantai Pasokan. Jakarta: Salemba Empat.

[7] Gulec, A. 2017. Courier service quality from the clients’ perspective. Engineering Management in Production and Services, 9(1), pp. 36-45.

[8] Alamsyah, A., & Rachmadiansyah, I. 2018. Mapping online transportation service quality and multiclass classification problem solving priorities. In Journal of Physics: Conference Series 971(1), p. 012021. IOP Publishing.

[9] Pitel, N. Y., & Alioshkina, L. P. 2016. ABC analysis as a tool of optimization of marketing management of export-led enterprises. 3, pp. 87-94

[10] Arusada, M. D. N., Putri, N. A. S., & Alamsyah, A. 2017. Training data optimization strategy for multiclass text classification. In 2017 5th International Conference on Information and Communication Technology (ICoICT) (pp. 1-5). IEEE.

[11] Medhat, W., Hassan, A., & Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), pp. 1093-1113.

[12] Prasetyo, E., & MINING, D. 2014. Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta, CV. Andi Offset.

[13] Syarif, I. 2014. Dimensionality Reduction Algorithms on High Dimensional Datasets. EMITTER International Journal of Engineering Technology, 2(2), pp. 28-38.

[14] Han, J. K. M., and Pei, J. 2012 Data Mining Concepts and Technique 3rd Edition (Waltham:Elsevier)

[15] Mayadewi, P., & Rosely, E. 2015. Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. SESINDO 2015, 2015 pp. 329–334.

[16] Fong, S., Lan, K., & Wong, R. 2013. Classifying human voices by using hybrid SFX time-series preprocessing and ensemble feature selection. BioMed research international, 2013. pp. 1-27.