Monitoring Quality of KAI Access Application Based on Customer Reviews on Google Play Store Using Laney p' Control Chart Based on Convolutional Neural Network

Naila Adelia Pribadi[1] and Muhammad Ahsan[1]

[1] Sepuluh Nopember Institute of Technology, Surabaya 60111, Indonesia
nailaadelia6@gmail.com

Abstract. The variety of activities that Indonesian people have has made mobilization increase. One form of transportation that can support close and long distance mobilization is rail transport. Purchasing train tickets can be made directly at the station or online. Online purchases can be made through an official application issued by PT Kereta Api Indonesia (KAI Access) or other service provider applications. The KAI Access application has a feature that allows ticket buyers to use complementary services such as railfood, connecting transportation, cancellation or rescheduling without having to go to the station, and other services. Despite its advantageous features, the app's Google Play Store rating remains relatively low at 2.5. Users also provide reviews, serving as valuable material for sentiment analysis and quality evaluation. Data spans from July 16, 2014, to February 15, 2023. Sentiment analysis, conducted through Convolutional Neural Network classification, revealed that 47.8% of reviews conveyed negative sentiment, while 52.12% were positive. Classification accuracy (AUC) for training data was 76.4%, falling under the fair category, while testing data achieved 97.3%, classified as excellent. The analysis identified common user issues, including challenges with account registration and login, application performance lag, and payment difficulties. The study ultimately demonstrates the potential for attribute control charts, specifically the Laney p', in effectively monitoring sentiments within large and varied sample sizes.

Keywords: Convolutional Neural Network, KAI Access, Laney p' Control Chart.

1 Introduction

In Indonesia, the diverse range of activities necessitates frequent travel, both by land, sea, and air. This leads to an increase in the number of passengers using transportation, including railway transport [1]. Technological advancements have enabled ticket purchases not only through direct providers but also online, including for rail transport. PT Kereta Api Indonesia facilitates online ticket sales through their official app, KAI Access, alongside other service providers [2]. While KAI Access offers various conveniences such as food orders, transport connections, and easy cancellations or rescheduling, its Google Play Store rating remains at a modest 2.5. This suggests some users may

N. Djam'an et al. (eds.), *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)*, Advances in Computer Science Research 109,
https://doi.org/10.2991/978-94-6463-332-0_20

not be entirely satisfied. A review of the past year indicates a notable number of negative feedback, emphasizing the need for quality monitoring and improvement. User reviews serving as valuable material for sentiment analysis and quality evaluation.

Sentiment analysis employs Natural Language Processing (NLP) to discern user sentiments, providing insights into their experiences [3]. To process the textual data, word weighting methods are applied for structural refinement. Meanwhile, statistical quality control, are employed to monitor and improve product quality [4]. To evaluate the quality of an application based on user reviews, it is essential to implement statistical quality control with attribute control charts. The Laney p' control chart proves suitable for this study, capable of monitoring defect proportions within large, varied sample sizes.

Several studies have been conducted on quality control of applications using sentiment analysis. Kim & Lim [5] focused on integrating sentiment analysis and process control analysis to monitor customer complaints. They suggested that this integration, along with Statistical Process Control (SPC), could enhance service quality. However, this study acknowledged certain limitations that warrant further exploration in future research. Another study by Firmansyah [6] employed Convolutional Neural Network (CNN) classification and Laney p' control chart, resulting in a more sensitive p-chart compared to the Laney p' control chart. Similarly, Apsari [7] used Naïve Bayes classification and Laney p' control chart, yielding a more sensitive p-chart compared to the Laney p' control chart. Additionally, in addition to user reviews, quality control of an application can be conducted using user-provided ratings. This approach was also employed by Hidayatillah [8] using the Laney p' control chart.

2 Literature Review

2.1 Convolutional Neural Network

The Convolutional Neural Network (CNN) is an extension of Multilayer Perceptron (MLP) designed for two-dimensional data processing. It operates similarly to MLP, but each neuron in CNN processes data in two dimensions, whereas in MLP, neurons are one-dimensional. The architecture of an MLP consists of layers with each containing neurons. It takes one-dimensional input data, propagates it through the network, and produces an output. Each connection between adjacent neurons has a one-dimensional weight parameter determining the mode's quality. CNN is applied in computer vision and Natural Language Processing (NLP), featuring convolutional and pooling layers. The convolution process applies filters to input data, generating a 2D feature map. These filters are obtained through training in a specific training set. The convolutional layer significantly influences model complexity through depth, stride, and zero padding adjustments. The pooling layer reduces output data dimensions. Pooling typically employs mean() for average pooling or max() for max pooling [9]. In this study, the output layer classifies reviews into positive (1) and negative (0) based on extracted features from preceding layers.

2.2 Laney p' Control Chart

The Laney p' attribute control chart combines principles from the Z control chart and Donald Wheeler's concepts to address sample size issues in p control charts [10]. When observation samples are very large, it leads to narrower control limits, causing many observations to fall outside these limits. The equation used to find the proportion of defective units for each subgroup is given by $p_i = \frac{D_i}{n_i}$, where p_i represents the proportion of negative reviews in subgroup i, D_i is the number of negative reviews in each sample in subgroup i, and n_i is the sample size in each subgroup i. The control limits and centerline for the Laney p' control chart are determined as follows [11].

$$\text{Upper Control Limit (UCL)} = \bar{p} + 3\sigma_{p_i}\sigma_z \quad (1)$$

$$\text{Central Line (CL)} = \bar{p} \quad (2)$$

$$\text{Lower Control Limit (LCL)} = \bar{p} - 3\sigma_{p_i}\sigma_z \quad (3)$$

with the average proportion values $\bar{p} = \frac{\sum_{i=1}^{n} D_i}{n_I} = \frac{\sum_{i=1}^{n} \hat{p}_i}{I}$. The σ_z is the sigma value for the individual chart. Therefore, standardization is carried out using $z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}}$. The σ_z is calculated as $\sigma_z = \frac{\bar{\bar{R}}_I}{1{,}128}$ and the σ_{pi} is calculated as $\sigma_{p_i} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$ with the R_i value $R_i = |z_i - z_{i-1}|$ and $\bar{R}_i = \frac{1}{I-1}\sum_{i=2}^{I} R_I$.

3 Methods

The data used in this study consists of KAI Access application ratings and user review from July 16, 2014, to February 15, 2023 on the Google Play Store. This study employs two sets of variables. For Convolutional Neural Network classification, it uses predictor variables which are user reviews of the KAI Access app on Google Play Store, and the response variable is the sentiment classification (negative or positive) based on the ratings. On the other hand, for the Laney p' chart, it uses variables derived from CNN classification results and the total negative comments in user reviews of the KAI Access app, categorized by types of defects. The types of defects are outlined in **Table 1**.

Table 1. The types of defects in the KAI Access application.

Category	Types of defects
1	Slow application
2	Payment issues
3	Application cannot be opened/frequently force-closed
4	KAI Pay issues (missing balance and activation)
5	Account suddenly logged out
6	Log-in/sign-up issues
7	Ticket booking and cancellation issues
8	Ticket not appearing
9	Server error

4 Result and Discussion

4.1 Pareto Chart

Pareto diagram aims to highlight the most frequently occurring issues that require immediate attention. It describes user complaints about the KAI Access application on Google Play from July 16, 2014, to February 15, 2023. The Pareto diagram representing the number of user complaints about the KAI Access application is shown in **Fig. 1**.

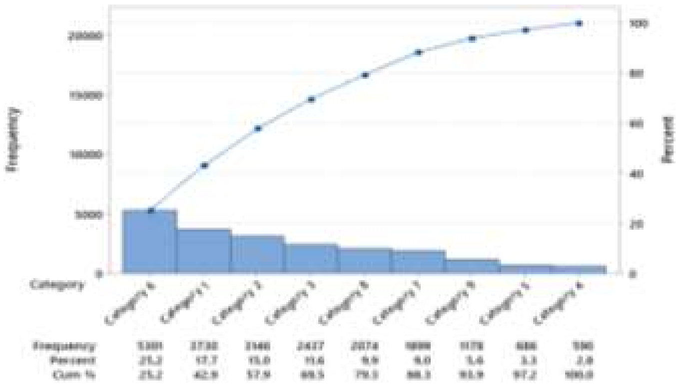

Fig. 1. Pareto chart of KAI Access application's type of defects.

From July 16, 2014, to February 15, 2023, in the KAI Access app, the predominant issue is Login/Sign-up (25.2%), followed by Slow Application (17.7%) and Payment Problems (15%). Defect categories are defined in **Table 1**. Reviews were categorized using a WordCloud from negative comments.

Fig. 2. Wordcloud of negative user reviews.

From the words appearing in the word cloud in **Fig. 2**, defect categories are identified and organized in **Table 1**. These will serve as the basis for determining defect categories in the monthly data analysis. To classify each comment into its respective defect category, programming with keywords for each defect category is required.

4.2 P Chart Based on Rating

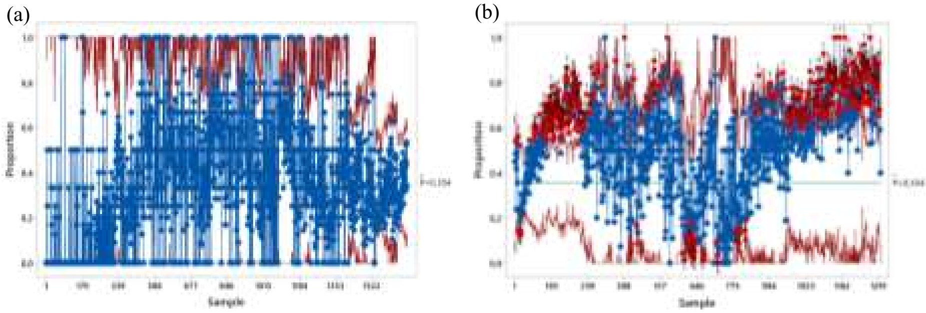

Fig. 3. P chart based on rating (a) 2nd iteration of Phase I, (b) Phase II.

Fig. 3 shows that after the second iteration of cleaning the Phase I chart from out-of-control observations, the p control chart based on ratings reveals an average proportion value of 0.354. This value will serve as a benchmark for monitoring Phase II data. However, Phase II's p-chart still exhibits some out-of-control observations.

4.3 Laney p' Chart Based on Rating

Fig. 4 shows that after the third iteration of cleaning the Phase I chart from out-of-control observations, the Laney p' control chart based on ratings reveals an average proportion value of 0.355. This value will serve as a benchmark for monitoring Phase II data. However, Phase II's Laney p' chart still exhibits out-of-control observations.

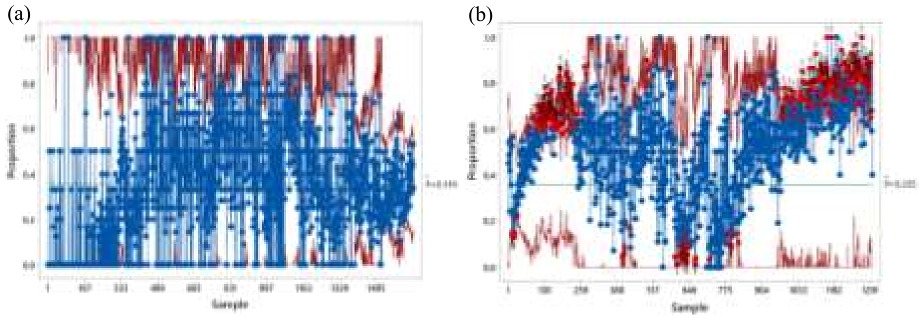

Fig. 4. Laney p' chart based on rating (a) 3rd iteration of Phase I, (b) Phase II.

4.4 Analysis of KAI Access Application User Reviews Using Convolutional Neural Network

The study employed user reviews from the Google Play Store to classify items using a technique called classification. This involves categorizing unlabeled items into specific classes. The review data was then split into training and testing sets using the stratified holdout validation method with a 90%:10% ratio. This method ensured that both positive and negative classes were evenly represented in both sets.

Before classifying the review data, it undergoes preprocessing steps. Non-text elements are removed to enhance classification accuracy. Tokenization is then applied to separate words. The subsequent Natural Language Processing (NLP) step involves converting tokens into representative numerical values through embedding. Each sentence is represented by a 1 x 40 vector. For reviews exceeding 40 words, they are truncated, and for those with fewer words, zeros fill the remaining spaces.

In the Convolutional Neural Network (CNN) deep learning method, the model learns through convolution by forming a convolutional layer and a pooling layer to extract crucial information from the feature map. The CNN model is constructed with an output layer generated by the softmax activation function. The optimal layer combination, obtained with 100 epochs in this CNN process, includes an embedding dimension of 100, Global Max Pooling 1D, a layer with 128 nodes, and an output layer. This combination will be used to build the CNN model for predicting the KAI Access application user reviews.

Table 2. Classification Accuracy of CNN Model

Dense layer combination	AUC		Accuracy	
	Training data	Testing data	Training data	Testing data
Embedding dimension = 100 Global Max Pooling 1D First Layer with 256 node Second Layer with 128 node Output Layer	75.4%	97.43%	77.5%	96.6%

Table 2 shows best model combinations built to achieve the best accuracy and AUC values for the CNN model. Upon evaluation, it was found that the classification accuracy and AUC for the training data fall under the "fair classification" category, while for the testing data fall under the "excellent classification" category. The CNN classification results, with 47.8% classified as negative and 52.12% as positive, will be used for analysis using p and Laney p' control chart based on review data.

4.5 P Chart Based on User Review

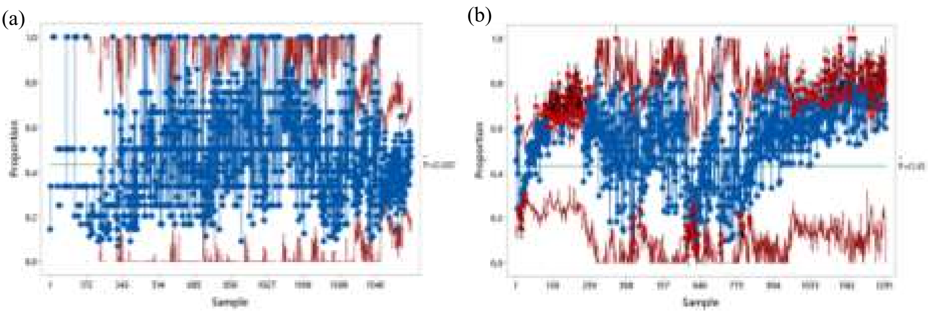

Fig. 5. P chart based on user review (a) 3rd iteration of Phase I, (b) Phase II.

Fig. 5 shows that after the third iteration of cleaning the Phase I chart from out-of-control observations, the p control chart based on ratings reveals an average proportion value of 0.430. This value will serve as a benchmark for monitoring Phase II data. However, Phase II's p-chart still exhibits some out-of-control observations.

4.6 Laney p' Chart Based on User Review

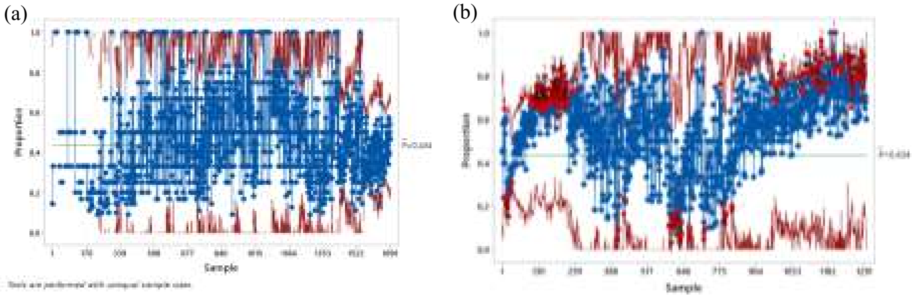

Fig. 6. Laney p' chart based on user review (a) 4th iteration of Phase I, (b) Phase II.

Fig. 6 shows that after fourth iteration of cleaning the Phase I chart from out-of-control observations, the Laney p' control chart based on ratings reveals an average proportion value of 0.430. This value will serve as a benchmark for monitoring Phase II data. However, Phase II's p-chart still exhibits some out-of-control observations.

4.7 Comparison Between p and Laney p' Chart

Table 3 shows the comparison between p and Laney p' control charts. It reveals that the p-control chart is more sensitive, detecting more out-of-control observations. The majority of issues occurred between November 2019 to March 2020 and July 2022 to February 2023, primarily related to login/registration, slow performance, and payment problems. In the first period, updates to the application might have caused login/sign-up issues. The decrease in comments in March 2020 may be linked to the Covid-19 pandemic. In the second period, increased usage led to slower performance.

Table 3. Comparison between p and Laney p' control chart

Months	Number of out-of-control observations			
	Based on rating		Based on customer review	
	P chart	Laney p' chart	P chart	Laney p' chart
August 2019	9	5	12	7
September 2019	2	1	2	2
October 2019	7	7	1	1
November 2019	21	14	13	12
December 2019	31	25	30	23
January 2020	29	26	11	19
February 2020	28	24	1	14
March 2020	24	19	21	18
April 2020	10	6	2	2

Table 3. Comparison between p and Laney p' control chart

Months	Number of out-of-control observations			
	Based on rating		Based on customer review	
	P chart	Laney p' chart	P chart	Laney p' chart
May 2020	2	2	1	5
June 2020	3	-	-	-
July 2020	1	1	1	1
August 2020	3	-	1	-
September 2020	6	2	7	2
October 2020	3	2	3	2
November 2020	7	3	1	2
December 2020	8	8	3	3
January 2021	2	-	-	-
February 2021	1	5	2	1
March 2021	6	3	8	2
April 2021	18	4	17	5
May 2021	3	3	3	3
June 2021	-	-	-	-
July 2021	5	5	1	5
August 2021	-	-	-	-
September 2021	4	1	4	2
October 2021	8	6	4	8
November 2021	1	-	8	-
December 2021	3	1	1	2
January 2022	4	5	2	3
February 2022	2	1	1	-
March 2022	7	3	6	4
April 2022	16	11	9	6
May 2022	15	15	9	9
June 2022	13	10	7	9
July 2022	24	21	16	19
August 2022	26	18	16	8
September 2022	23	23	18	18
October 2022	20	17	14	10
November 2022	21	17	9	6
December 2022	30	20	20	13
January 2022	24	20	19	12
February 2022	13	9	5	7

5 Conclusion and Future Research

Notably, the Laney p' Control Chart identified more out-of-control observations compared to the p Control Chart. This was attributed to certain observation samples with highly variable and large sizes, causing fluctuations in the control limits of the Laney p' Control Chart. However, in the Phase II data analysis conducted on a monthly basis, the p Control Chart demonstrated greater sensitivity compared to the Laney p' Control Chart. The significant number of statistically out-of-control samples indicates further improvements of the application are warranted for enhanced future performance.

Regarding user-reported issues based on ratings between July 16, 2014, and February 15, 2023, the most prevalent problems were related to account registration and login difficulties, sluggish application performance, and challenges encountered during the payment process.

For future research, considering categorizing comments with a rating of 3 as negative, a practice adopted by previous researchers, is recommended. Alternatively, implementing a multinomial control chart to distinguish ratings 1-2 as negative, 3 as neutral, and 4-5 as positive comments could provide valuable insights. Additionally, enhancing quality monitoring by using the central line value as a benchmark to assess the effectiveness of defect tolerance in the application is suggested.

References

[1] Badan Pusat Statistik, "Jumlah Penumpang Kereta Api (Ribu Orang), 2023," Badan Pusat Statistik. Accessed: Feb. 07, 2023. [Online]. Available: https://www.bps.go.id/indicator/17/72/1/jumlah-penumpang-kereta-api.html.

[2] "KAI," PT Kereta Api Indonesia (Persero). [Online]. Available: https://www.kai.id/information/full_news/4783-beragam-fitur-kai-access-permudahpelanggan-menggunakan-ka

[3] A. Khan, "Sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and Blogs," *International Journal of Computer Science & Emerging Technologies*, pp. 539–552, 2011.

[4] A. Mitra, *Fundamentals of Quality Control and Improvement*. John Wiley & Sons, Inc., 2021.

[5] J. Kim and C. Lim, "Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses," Ulsan, 2021.

[6] D. Firmansyah, "Monitoring Kualitas pada Aplikasi MyPertamina Berdasarkan Ulasan Pengguna di Google Play Menggunakan Diagram Kendali Laney p' Berbasis Convolutional Neural Network," Surabaya, 2023.

[7] N. T. Apsari, "Monitoring Kualitas Aplikasi PeduliLindungi Berdasarkan Ulasan Pelanggan di Google Play Menggunakan Diagram Kendali Atribut Laney p'," Surabaya, 2022.

[8] R. Hidayatillah, "Pengukuran Ketidakpuasan Pelanggan Menggunakan Attribute Control Chart Berdasarkan Data Customer Review Sebagai Dasar Penyusunan Rekomendasi Perbaikan Layanan," Surabaya, 2022.

[9] J. Patterson and A. Gibson, "Deep Learning A Practitioner's Approach." [Online]. Available: http://oreilly.com/safari.

[10] M. Ahsan, M. Mashuri, and H. Khusna, "Evaluation of Laney p' Chart Performance," *International Journal of Applied Engineering Research*, pp. 14208–14217, 2017.

[11] D. C. Montgomery, *Introduction to Statistical Quality Control*, Eighth. Danvers: John Wiley & Sons Inc, 2020.