International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Deep Learning and Statistical Process Control Approach to Improve Quality in Immigration Services

Pontiselly [a], Muhammad Ahsan [a,*], Muhammad Hisyam Lee [b]

[a] *Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[b] *Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*
*Corresponding author: *muh.ahsan@its.ac.id*

***Abstract*—Technological advancements have profoundly influenced various sectors, including tourism, by simplifying international travel and increasing the demand for passports. In this context, public feedback on the Surabaya Immigration Office's services, gathered from Google Maps reviews, presents an invaluable dataset for analysis. This study introduces a novel approach to quality management systems in immigration services by applying the Convolutional Long Short-Term Memory (Co-LSTM) for sentiment analysis of public reviews and p-attribute control charts for statistical process control. Focusing on the Surabaya Immigration Office, we analyze public feedback from Google Maps to assess service quality. Our methodology preprocesses and classifies opinion data into sentiment classes, employing Co-LSTM for its superior accuracy over traditional models. The sentiment analysis reveals a predominance of positive over negative reviews, with classification accuracy demonstrating Area Under Curve (AUC) values of 98.61% for training and 85.66% for testing data. Furthermore, the p-attribute control charts are utilized to monitor service defects, identifying areas of uncontrolled variability and suggesting the necessity for service improvement interventions. The study uncovers that the primary public grievance relates to the perceived lack of friendliness and politeness from office staff. By integrating sentiment analysis with statistical process control, this research offers a comprehensive approach for immigration service providers to enhance service quality, respond proactively to public sentiment, and ensure customer satisfaction.**

***Keywords*— Convolutional long short-term memory; sentiment analysis; service quality improvement; immigration services.**

Manuscript received 21 Mar. 2024; revised 20 Jul. 2024; accepted 3 Aug. 2024. Date of publication 31 Oct. 2024.

I. INTRODUCTION

The increasingly rapid development of technology has a very significant impact on human life. Most activities in various fields already use technology, including tourism. Technology makes human travel from one place to another more effective and efficient. In 2022, the number of national tourists every month will increase. The total number is 3,540,542 people, and this number has increased by more than 100% compared to 2021 [1].

When someone travels abroad, a passport becomes one of the mandatory documents that must be carried. Achmad Nur Saleh, the Public Relations Sub-coordinator of the Directorate General of Immigration, said that in the first quarter of 2022, the number of passport applications in Indonesia increased by 41% from the number of applicants in the fourth quarter of 2021 [2]. Then, in the first quarter of 2023, the number of passport applications reached around 13,292 applicants daily. This number shows an increase of more than two times compared to the first quarter of 2022. This rapid increase in passport applicants means that the quality of passport-making services at each immigration office must be continuously maintained. People who want to apply for a passport can visit the nearest immigration office. The public who has used the services of the immigration office can leave feedback or reviews in the Google Maps comments sections, which describe the quality of the service at the immigration office.

Sentiment analysis is one of many classifications for automatically capturing the public’s sentiments or emotions. Convolutional Long Short-Term Memory (Co-LSTM) is a hybrid classification method between CNN and LSTM that can produce better accuracy than another method [3]. Reviews with negative sentiments will be indicated as defects in the following monitoring process using one of the Statistical Process Control (SPC) methods, namely the p-control chart. Before, control charts could be divided into the variables [4]-[5] and attributes [6]-[10] based on the type of variable inspected. The p-control chart includes the attribute type and monitors the defect proportion (negative review sentiment), which uses varying samples for each subgroup.

Previous researchers have researched machine learning and Natural Language Processing (NLP) using various algorithms, such as the Naïve Bayes and SVM [7]-[13]. According to application reviews, some research related to sentiment monitoring is as follows. Research by Apsari, Ahsan, and Lee [8] which monitors the quality of rating data and user reviews of the PeduliLindungi application using p and Laney p' attribute control charts. Second, research on quality monitoring of MyPertamina application reviews using p and Laney p' attribute control charts conducted by [9, 10], also the Convolutional Neural Network (CNN) method is used for sentiment analysis of reviews, which results in an AUC value of 79.38%. The research by [28] on customer complaints monitoring with customer review data analytics proves that sentiment analysis can be combined with statistical process control analyses.

Then, the study by Kim, Park, and Kwak [11], who use review mining with lexicon-based methods from previous author research and p and EWMA control charts based on each review topic to monitor sentiment. A study by Zhang, He, and Mukherjee [12] evaluates products according to user reviews using sentiment intensity and time intervals from negative reviews. The monitoring result may be an early warning system for the institution to respond quickly when there are bad reviews from the public. The early warning system can detect signals about potential threats or risks, which means that some of the service stages are not good or not according to predetermined service standards. By listening to and understanding the public's voice, the immigration office can gain direct insight that can be used to conduct research, improvement, and innovation, as well as assess the quality of a service. After the control chart is created, it can be determined what part of the service stage or type of defect is most common or often experienced by the public over time. This research can be a reference and evaluation for the Directorate General of Immigration in maintaining the quality of services they provide at the immigration office. Although that insight may not instantly turn negative customer reviews into positive ones, the goal is to reduce the impact of the issues identified in the reviews and improve performance [19].

Thus, this research has three research questions: (1) What are the results of sentiment analysis of public reviews on Google Maps about Surabaya Immigration Office services using the Convolutional Long Short-Term Memory (Co-LSTM) method? (2) What are the results of monitoring the quality of public review sentiment on Google Maps about Surabaya Immigration Office services using the p attribute control chart? (3) What obstacles are often experienced by the community, and what recommendations can be given to the Directorate General of Immigration?

II. Materials and Method

A. Materials

1) Data Source: The data used for this study is secondary data about public reviews regarding passport-making services at the Surabaya TPI Special Class I Immigration Office, which are taken by scraping the comment column on Google Maps from December 2018 to November 2023 or accounting for five years. The data is divided into phase I data from December 2018 to November 2022 and phase II data from December 2022 to November 2023. The data used has subgroups, observations per week, with each sample size in each subset depending on the number of reviews each week.

2) Research Variable: The variable used in this research is rating-based classification. The labeling of reviews is done by using the rating value, where ratings 1-2 are categorized as negative classes, ratings 3 as neutral classes, and ratings 4-5 as positive classes. However, manual labeling was also carried out for the neutral class classification to check whether reviews with a rating of 3 were truly neutral or whether there was a tendency towards positive or negative intentions. Thus, reviews that tend to be positive are labeled as positive, and those that tend to be negative are labeled as negative. Then, the variable number of negative reviews to be observed is categorized into several groups based on the type of defect, which is done by adjusting the keywords in the review. The types of defects in passport processing services based on public reviews are shown in Table I below.

TABLE I
TYPES OF DEFECTS IN PASSPORT PROCESSING SERVICES

Category	Types of Defects
1	Obstacles during registration or filing
2	Obstacles in the photo process
3	Less clean and inadequate place or area
4	Long service and long queues
5	Obstacles in the M-Passport application
6	Officers are less friendly and polite in serving
7	Obstacles in payment
8	Obstacles in selecting dates and schedules
9	Far parking location
10	Other factors

3) Data Structure: Table II below illustrates the structure of the scraped data.

TABLE II
DATA STRUCTURE FROM SCRAPING

ReviewID	Username	Review	Rating	Label
1	$user_1$	u_1	r_1	y_1
2	$user_2$	u_2	r_2	y_2
3	$user_3$	u_3	r_3	y_3
⋮	⋮	⋮	⋮	⋮
I	$user_I$	u_I	r_I	y_I

Furthermore, the data structure for p-attribute control chart generation is shown in Table III below.

TABLE III
DATA STRUCTURE FOR THE CONTROL CHART

Week	Date	Sample Size	Negative Rating	$\hat{p}_i$
1	03/12/2018 - 09/12/2018	n_1	D_1	$\hat{p}_1$
2	10/12/2018 - 16/12/2018	n_2	D_2	$\hat{p}_2$
3	17/12/2018 - 23/12/2018	n_3	D_3	$\hat{p}_3$
⋮	⋮	⋮	⋮	⋮
m	23/11/2023 - 30/11/2023	n_m	D_m	$\hat{p}_m$

B. Data Preprocessing

The scraped data must first be pre-processed so that it becomes more accessible and manageable. The process is as follows.

1. Checking data information consists of 7 variables, namely review_id, branch, review_datetime_utc, author_title, review_rating, review_text, and label (actual).
2. Checking for missing values: Out of 3045 community review data, 1051 reviews have no comments or, in other words, only the rating. Thus, 1051 data were deleted and not used in the study.
3. Renaming variable names to make it easier to process.
4. Importing some libraries to do text mining.
5. Removing emojis that are listed in the reviews.
6. Removing characters other than letters, such as numbers, in reviews.
7. Eliminating the URL listed in the review and removing the punctuation marks listed in the review.
8. Changing the letters in the review to lowercase format.
9. Breaking down review sentences into smaller word tokens and counting the word frequency in each review.
10. Performing word normalization.
11. Importing stop word data that will be used to remove stop words in reviews (stop word removal).
12. Performing word normalization so that words with the same meaning can only be written in one kind of word.
13. Filtering review data from December 1, 2018, to November 30, 2023, amounts to 1662 data.
14. This clean data will be used to perform the analysis below.

C. Method

1) Descriptive Statistics: Descriptive statistics study how data is collected and presented in a way that is easy to understand. Pie charts and bar charts are the descriptive statistics used in this study. A pie chart is a graph that illustrates the size of the comparison of each part of the circle in percentage or proportion. A bar chart is a diagram that illustrates data using vertical and horizontal bars that have a height or length that shows the frequency of the data [13]. Furthermore, in Statistical Process Control (SPC), Pareto diagrams are often used to improve the quality of a product. A Pareto diagram is a frequency distribution (or histogram) of attribute data organized by category. It should be noted that Pareto diagrams don't automatically identify the most critical defects but are sorted only by the most frequent or most occurring to the least or rare [14].

2) Text Mining: Classification techniques, sometimes known as "categorization techniques," are used in text mining to organize an unstructured collection of text fragments into structured data by looking for patterns from the sentence structure [15]. Before doing text mining, text data needs to be processed first through text processing, which has the following stages [16].

1. Case folding: Transforming the text into lowercase format.
2. Tokenizing: Dividing a sentence into words (tokens).
3. Stemming: Taking off prefixes, suffixes, inserts, and confixes to get a basic word.
4. Filtering or Stop word Removal: The process of keeping relevant words (wordlist) or eliminating words that are not (stoplist).

3) Word Cloud: Word cloud is a visualization technique often used to visualize text data. The word cloud will describe the graph of text data or documents by plotting words with a high frequency of occurrence in a two-dimensional space. If the frequency of occurrence of a word is more frequent, the size of the word on the graph will be larger. Conversely, if the frequency of occurrence of a word is rare, then the size of the word on the graph will be smaller [17].

4) Sentiment Analysis: Sentiment analysis is a natural language processing (NLP) technique that examines, extracts, and quantifies the polarity of information about personal beliefs and feelings [18]. Artificial intelligence, computer vision, and natural language processing are developing rapidly, making it more and more possible to use sentiment analysis in artificial agents. This interdisciplinary topic of study includes social science, psychology, computer science, and other relevant fields [19]. With this analysis, the text in a sentence can be categorized according to certain classes, such as positive, neutral, and negative. Positive sentiment is a reaction or attitude that increases the value of someone or something. Some examples of public reviews that fall into this class are "Good application, can make it easier to make a passport" and "Applications that facilitate services so that applicants do not need to queue long at the immigration office and there are no additional fees. The word neutral can be interpreted as not taking sides (not participating or not helping one party). One example of a public review that falls into the neutral class is "Hopefully, the quality of the application will be improved." Negative sentiment is a reaction or attitude that lowers the value of someone or something. Some examples of public reviews that fall into this class are "The app doesn't make it easy; it makes it difficult" and "The app is often out even though there is no signal interference."

5) Holdout Validation: The principle of this holdout validation method is to split the dataset into multiple parts and use one part to train the model and the rest to validate and test it [20]. This data will be used for both model evaluation and selection. This method is good to use when the data is large and requires a fast process time. The division of datasets into training and testing data is generally done randomly, so the proportion of division may become uneven between classes or categories. This causes overrepresentation, so to prevent it, holdout stratification can be done so that each classification class can be represented in training and testing data.

6) Word2Vec: Unstructured text data can't be directly processed by computers. It should be converted first to make that data structure. In natural language processing (NLP), Word2Vec is a popular technique that enables words to be represented as vectors in a continuous vector space [21]. To capture the semantic associations between words, Word2Vec maps words to high-dimensional vectors. Word2vec comprises two architectures: the Skip-gram and Continuous Bag of Words Model (CBOW) Model. Based on the surrounding context words of a sentence, the CBOW model will forecast the target words or the word's core.

7) Convolutional Long Short-Term Memory (Co-LSTM): Co-LSTM is a hybrid method of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) algorithms. Fig 1 shows the model structure of CNN-LSTM, which, according to this method name, places the LSTM algorithm between the CNN algorithm as follows.

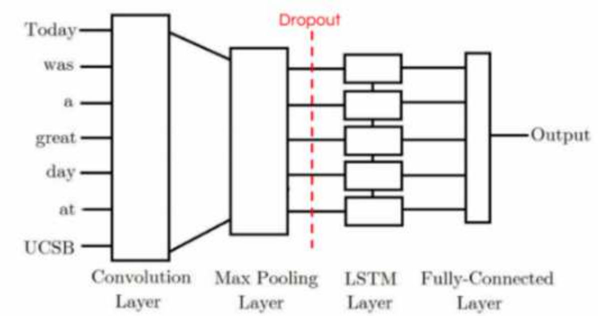

Fig. 1 Co-LSTM Illustration

The review words processed by word embedding will be used as input to the convolutional layer of the CNN network. The data will undergo a convolution process, where the layer will apply a kernel (filter) to all data, producing an activation map or 2D feature map. Next, pooling will be performed on the pooling layer with the max pooling method, which takes the maximum value of each block on the 2D feature map of a certain size. The goal is to summarize the existence of features where the result is a z vector. The output of this pooling layer will be forwarded to the layers in the LSTM network to analyze the vector of features sequentially from left to right [3]. Here, the LSTM network will use its input, output, and forget gates to travel through the gates to check the long-term dependency and discover global features. Then, to predict the actual sentiment, the LSTM layer output will be sent back through the CNN layer in the fully connected layer. This process is performed by a SoftMax activation function that will produce a probability value for each class. In the end, the class with the highest probability value will be the predicted class of the classification.

8) p Attribute Control Chart: The p attribute control chart controls the product quality whose characteristics can only be divided into two categories, such as defective or non-defective, good or bad, and so on. The percentage of defective items in each sample is the defect measure used for the p-control chart. The defective proportion of the sample itself is the ratio of the number of defective or non-conforming units in the sample to the sample size n. However, it is not uncommon for the defective portion of the process to be unknown, thus p should be estimated and calculated using the observed data. A commonly used procedure is to select m preliminary samples of size n each. If there are as many as D_i defective units in the i-th sample, then the proportion of defects in the i-th sample, which may be calculated with Equation 1 [22].

$$\hat{\boldsymbol{p}}_{\boldsymbol{i}} = \frac{\boldsymbol{D_i}}{\boldsymbol{n_i}} \tag{1}$$

The sample data taken is binomially distributed, so it can be assumed that the proportion of defects in all products is p. Thus, Equation 2 below illustrates the probability that the number of defects (D_i) in a sample size n is x units [23].

$$P(x;n;p) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, x = 0,1,\ldots,n \tag{2}$$

In this control chart, p statistics are used to estimate the proportion of defective products in some production. Equation 3 is the formula for calculating the average proportion of defective products.

$$\bar{p} = \frac{\sum_{i=1}^{m} D_i}{\sum_{i=1}^{m} n_i} \tag{3}$$

Then, there's a formula to calculate the control limits, and the center line in the p attribute control chart can be seen in Equation 4 below.

$$\begin{aligned} UCL &= \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \\ CL &= \bar{p} \\ LCL &= \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \end{aligned} \tag{4}$$

D. Research Flow

This study delved into public perception of the Surabaya Immigration Office's passport-making services by analyzing reviews and ratings from Google Maps. Leveraging a web scraping tool, researchers first collected the data. Review reviews were classified as positive, neutral, or negative to understand user sentiment. The labeling process of sentiment classes is conducted based on ratings. Ratings 1-2 being negative sentiment classes, ratings 3 being neutral sentiment classes, and ratings 4-5 being positive sentiment classes. Following text pre-processing to ensure data quality, researchers conducted exploratory analysis. The data was then strategically divided into training and testing sets to develop a model for review classification with a ratio of 80%:20%. A Convolutional Long Short-Term Memory (Co-LSTM) model was chosen, and its effectiveness in categorizing reviews was evaluated using the accuracy and AUC value of the former Co-LSTM model.

Shifting the focus to service quality monitoring, the data was segmented into two phases. A control chart technique was employed to analyze the frequency of negative reviews within each phase. If the analysis indicated a lack of statistical control, signifying potential service quality issues, researchers iteratively removed outlying negative reviews furthest from the control limits. Following each removal, the analysis was rerun to assess the impact.

Finally, a Pareto diagram was constructed to pinpoint the most common types of negative feedback. This visualization provided valuable insights into areas requiring improvement at the Surabaya Immigration Office. The study identified specific service quality issues by analyzing user reviews and subsequently formulated recommendations for improvement. The research culminated in conclusions and suggestions to guide future efforts in enhancing the passport-making experience at the Surabaya Immigration Office.

III. Result and Discussion

This chapter will explain the analysis of the data processing that has been done. The method used in this research is the p attribute control chart using public review data related to December 1, 2018 - November 30, 2023.

A. Descriptive Statistics

Analysis of data characteristics will be carried out using data visualization, namely pie charts and bar charts.

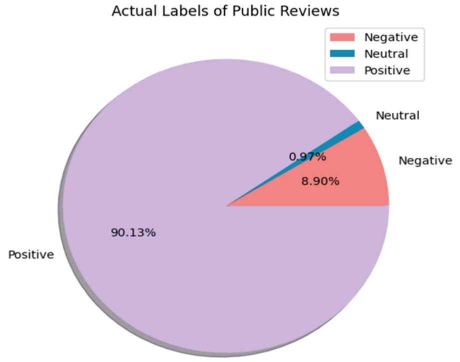

Fig. 2 Proportion of Actual Data Class Based on Rating

Based on Figure 2, 90.13% is categorized as positive, 8.90% as negative, and 0.97% as neutral. This calculation is obtained from the number of each sentiment category compared to the total data of public reviews or comments related to passport-making services from Google Maps of immigration offices in Surabaya multiplied by 100%. Thus, when viewed based on the rating, passport-making services at the Surabaya Class I Special TPI Immigration Office tend to be good, which is indicated by the number of positive ratings. Next, a bar chart will show the number of actual sentiment groups for each quarter, as shown in Figure 3.

Based on Figure 3, most reviews occurred during the second quarter of 2020. Then, after being traced, the most reviews fell during May 2020, which amounted to 270 reviews. This may be due to several factors, such as the COVID-19 pandemic, which impacts the number of people who need to deal with the immigration office to adjust their travel documents due to the imposition of travel restrictions and lockdowns. Meanwhile, the lowest number of reviews occurred in June 2021, at 0. This might happen because the services provided by the immigration office in Surabaya are satisfactory, so people do not leave many reviews. When people feel satisfied with something, they tend to comment less often than when they think unsatisfactory service. Furthermore, Figure 4 shows the number of negative reviews in each quarter.

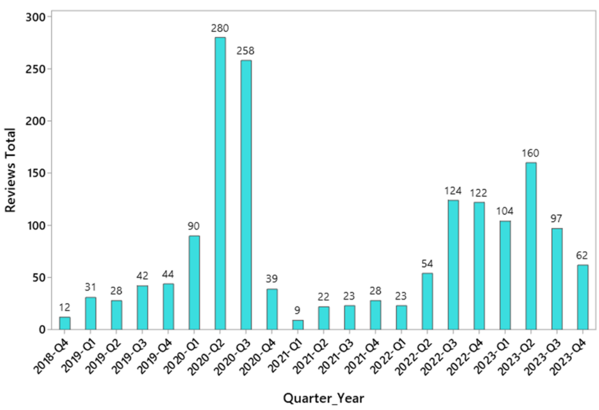

Fig. 3 Number of Reviews Each Quarter

Based on Figure 4, the highest negative reviews occurred in quarter 3 of 2023. This may be due to several policy changes related to travel between countries, which can also affect the provision of passport processing services at the immigration office in Surabaya, which apparently cannot be adequately implemented, triggering negative reviews from the public. Meanwhile, the lowest number of negative reviews occurred in several quarters, namely quarter 2 of 2020 and quarters 1-3 of 2021. Negative reviews can be used as an early warning system indicating that there are conditions in the passport-making service process that don't meet public expectations. Therefore, people convey this through the comment columns in each place where they take care of their passports on Google Maps. Table IV shows the percentage of negative reviews from December 2018 to November 2023.

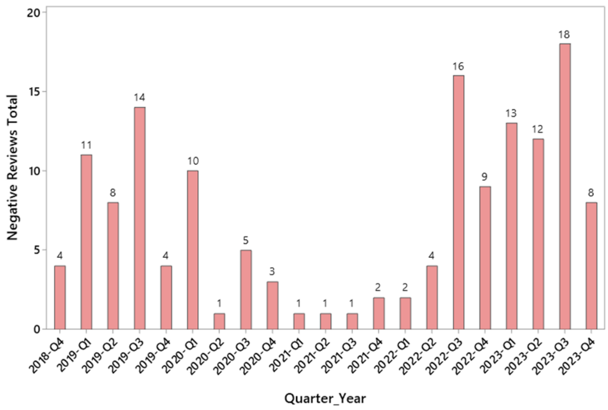

Fig. 4 Number of Negative Reviews Each Quarter

Table IV shows that the highest percentage of negative reviews was from January to February 2019 and July 2019, which amounted to 50%. This may be due to several factors that make people less satisfied.

TABLE IV
PERCENTAGE OF NEGATIVE REVIEWS EACH QUARTER

Month	Percentage of Negative Reviews Each Quarter					
	2018	2019	2020	2021	2022	2023
Jan		50.00%	14.29%	0.00%	16.67%	25.00%
Feb		50.00%	27.78%	0.00%	0.00%	4.17%
Mar		20.00%	5.17%	16.67%	12.50%	2.78%
Apr		25.00%	0.00%	0.00%	14.29%	11.76%
May		25.00%	0,37%	33.33%	9.38%	9.09%
June		33.33%	0.00%	0.00%	0.00%	6.82%
July		50.00%	1.96%	-	13.51%	9.76%
Aug		40.00%	2.48%	0.00%	18.18%	25.00%
Sept		7.69%	0.00%	4.76%	10.77%	25.00%
Oct		20.00%	4.17%	0.00%	5.56%	6.45%
Nov		7.14%	0.00%	7.69%	12.50%	19.35%
Dec	33.33%	9.09%	40.00%	33.33%	3.57%	

B. Sentiment Analysis using Co-LSTM

The public review data for the Surabaya Immigration Office services on Google Maps was used to train a classifier. The classifier could categorize the reviews into positive, neutral, and negative. The training data comprised 1321 public reviews, and the testing data comprised 331 public reviews. Then, after training several times using the Co-LSTM method, the best model prediction results were obtained, as shown in Figure 5.

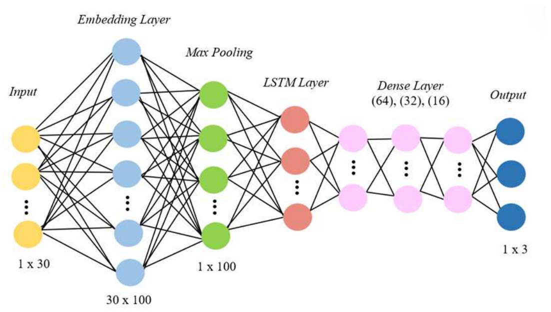

Fig. 5 The best model architecture using Co-LSTM

The performance of the Co-LSTM classifier on the training data is summarized in the confusion matrix, which is shown in Table V. The confusion matrix shows that the classifier correctly classified 1,184 positive reviews, 0 neutral reviews, and 115 negative reviews.

TABLE V
CONFUSION MATRIX TRAINING DATA

Actual	Prediction		
	Negative	Neutral	Positive
Negative	115	0	2
Neutral	13	0	0
Positive	7	0	1184

The model accuracy on the training data was 98.33%, and the AUC was 98.61%, indicating excellent classification. Meanwhile, the performance of Co-LSTM classifier on the testing data is summarized in the confusion matrix as shown in Table VI. The confusion matrix shows that the classifier correctly classified 283 positive reviews, 0 neutral reviews, and 19 negative reviews.

TABLE VI
CONFUSION MATRIX TESTING DATA

Actual	Prediction		
	Negative	Neutral	Positive
Negative	19	0	11
Neutral	1	0	2
Positive	15	0	283

The model accuracy on the training data was 91.23%, and the AUC was 85.66%, indicating good classification. Thus, the results of this sentiment analysis consist of two categories, namely negative and positive sentiment classes. Figure 6 illustrates the negative sentiment class word cloud.

Fig. 6 Word Cloud of Negative Reviews Prediction

The words that often appear in Figure 6 show that people who come to the Surabaya Immigration Office to receive passport-making or renewal services complain about things related to services provided by officers sometimes less friendly, there are long queues, and problems with online registration. Then, Figure 7 shows the word cloud of the positive sentiment class.

Fig. 7 Word Cloud of Positive Reviews Prediction

The words that often appear show that people who come to the Surabaya Immigration Office to receive passport-making or renewal services feel that the services of immigration office officers tend to be friendly and fast. The facilities provided at the immigration office can give a sense of comfort to the people who come.

C. p Attribute Control Chart Based on Phase I Review Data

The phase I p-control chart based on review data was created using public review data on passport processing services at the Surabaya Immigration Office from December 1, 2022, to November 30, 2023. This data is used to obtain the average value of the proportion of defects that will be used to monitor phase II public review data. Figure 8 shows the phase I p control chart using the 3-sigma limit to determine whether the passport-making service at the Surabaya Immigration Office has been statistically controlled.

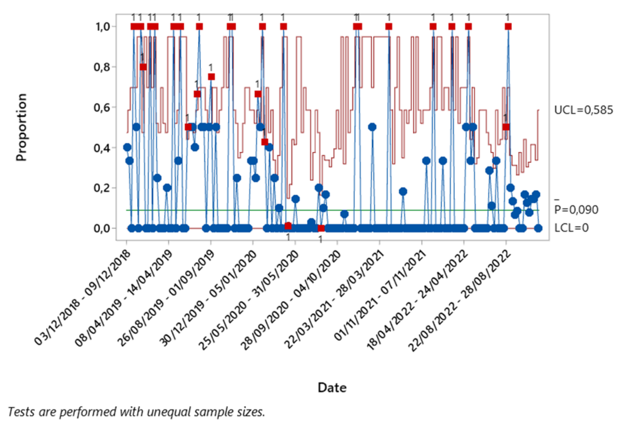

Fig. 8 Phase I p-Control Chart

Figure 8 shows that the phase I iteration 0 p control chart has an average proportion value of 0.090 with different upper and lower control limits in each subgroup. On this control chart, there are still 25 observations that are outside the upper control limit and two observations outside the lower control limit. Thus, one-by-one deletion will be carried out on observations outside the farthest control limit so that all observations in the p diagram can be statistically controlled. After 13 iterations, in-control results were obtained as shown in Figure 9 below.

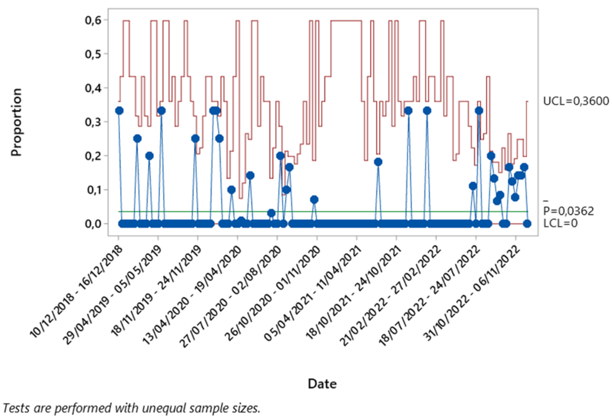

Fig. 9 Phase 1 p-Control Chart Iteration 13

The phase I iteration 13 p control chart in Figure 9 has an average defect proportion value of 0.0362 and shows that no observations leave the upper control limit, meaning that the p control chart is statistically controlled. In phase II, the proportion average of defects on the phase I iteration 13 p control chart will be used to monitor passport processing service review data at the Surabaya Immigration Office.

D. p Attribute Control Chart Based on Phase II Review Data

The phase II p control chart was created based on public review data related to passport-making services at the Surabaya Immigration Office on Google Maps from December 1, 2022, to November 29, 2023. It uses the defects proportion in the phase I p control chart, which is already statistically controlled and has a 0.0362 value. The phase II p control chart analysis is shown in Figure 10.

Figure 10 shows the phase II p control chart based on public reviews of passport-making services at the Surabaya Immigration Office on Google Maps. The analysis results in Figure 10 show that several out-of-control observation points make the phase II p-control chart not yet statistically controlled. Table VII summarizes the causes and number of out-of-control observations in each month.

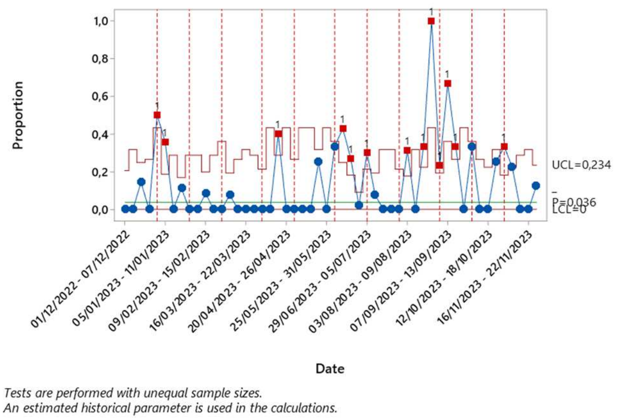

Fig. 10 Phase II p-Control Chart Based on Reviews

Table VII illustrates the causes of OOC observations and their percentages. The percentage of causes of OOC observations is calculated from the number of causes of OOC observations divided by the amount of data available in that month. In September 2023, the causes of the most OOC, 65.00%, are as follows.

1. Category 6: Officers are less friendly and polite in serving.
2. Category 3: Place or area that is less clean and less adequate.
3. Category 1: Obstacles during registration or filing.
4. Category 2: Obstacles in the photo process
5. Category 4: Long service and long queues
6. Category 7: Obstacles in payment
7. Category 8: Constraints in date and schedule selection
8. Category 10: Other factors

The percentage of the most common causes of OOC in September 2023 above is 65.00% of the total number of observations in that month.

TABLE VII
SUMMARY OF OOC OBSERVATIONS BASED ON REVIEW DATA

Month	Number of OOC Obs. (Week)	Causes of OOC Obs.	Percentage Cause of OOC Obs.	Immigration Office Location with OOC Obs.
Dec 2022	-	-	-	-
Jan 2023	2	Category 10, 4, 5, 6	27.27%	Tanjung Perak Wiyung Ciputra World BG Junction Juanda
Feb 2023	-	-	-	-
Mar 2023	-	-	-	-
Apr 2023	1	Category 1, 2, 6	17.65%	Juanda Tanjung Perak
May 2023	-	-	-	-
June 2023	2	Category 1, 6	12.12%	Wiyung Tanjung Perak Juanda BG Junction
July 2023	1	Category 6, 1, 8	21.95%	Juanda Tanjung Perak BG Junction
Aug 2023	4	Category 6, 4, 8	58.33%	Juanda Wiyung Tanjung Perak BG Junction
Sept 2023	3	Category 6	65.00%	Wiyung Tanjung Perak BG junction Juanda
Oct 2023	1	Category 1, 2, 4, 7	12.90%	Juanda BG Junction Tanjung Perak
Nov 2023	1	Category 4, 8, 6, 7	25.81%	BG Junction Juanda

E. Pareto Diagram

A Pareto diagram can be used to see the characteristics of defect categories in public reviews regarding passport-making services at the Surabaya TPI Special Class I Immigration Office. This diagram aims to show which problems have the highest frequency so that they are deemed necessary to solve as soon as possible. The Pareto diagram describes public complaints related to passport-making services. However, before making a Pareto diagram, keywords are first defined for each defect category.

a. Category 1: *'berkas', 'dokumen', 'persyaratan', 'daftar', 'materai', 'surat', 'ijazah'*
b. Category 2: *'foto'*,'dresscode'
c. Category 3: *'toilet', 'jorok', 'area', 'ruang', 'bersih'*, 'room'
d. Category 4: *'lama'*, 'over', 'time', *'antri', 'jam'*, 'crowded', *'menunggu'*

e. Category 5: *'aplikasi', 'down', 'online', 'otp',* 'login', 'error'
f. Category 6: *'petugas', 'staf',* 'staff', 'security', *'sopan', 'santun', 'ramah', 'pns', 'perempuan', 'laki', 'pegawai', 'admin', 'profesional'*
g. Category 7: *'bayar', 'uang', 'lunas'*
h. Category 8: *'tanggal', 'jadwal',* 'reschedule', *'kuota'*
i. Category 9: *'lahan', 'parkir', 'parkirnya', 'jauh', 'mobil', 'parkiran'*
j. Category 10: Words other than those listed in categories 1-9 in negative label reviews.

Figure 11 below shows the Pareto diagram of the number of public complaints related to passport-making services at the Surabaya Immigration Office.

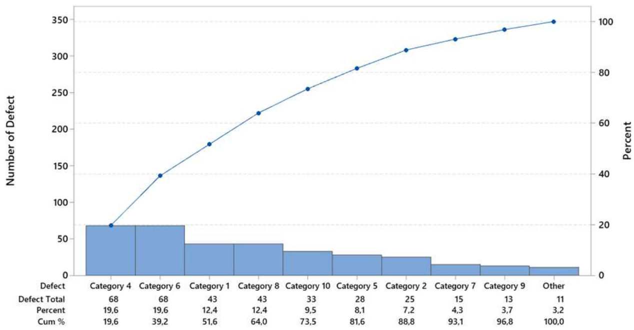

Fig. 11 Pareto Diagram

Based on Figure 11, the highest number of obstacles or defects complained about by the public regarding passport-making services at the Surabaya Immigration Office is 19.6% in category 4, namely, extended services resulting in long queues. This needs further attention because this queue makes the service inefficient, as many people come in the morning but are only finished being served in the afternoon.

F. Recommendations for Service Improvement

The number of OOC observations in the monitoring results can be used as an early warning system for better passport-making services. Here are some recommendations that can be given. Officers who are less friendly and polite are the category with the highest defects many complain about when processing passports. This is an essential indicator because officers will indirectly represent the institution regarding satisfaction with the services provided. The recommendation that can be given is that if there is no training period, officers can be given adequate training, especially regarding knowledge of hospitality (attitudes or treatment given to others) in serving people who apply for passports. On the other hand, if training has been held at the beginning, the employee training material needs to be broken down again according to the existing evaluation.

The long queue at the immigration office can be caused by the large number of people who want to apply for a passport at one time or because the number of servers to provide services is still insufficient. The recommendation that can be given is that if the queue continues to occur every day, it is necessary to add the number of servers or counters. In addition, the quota system that has been implemented in the M-Passport application can be improved better.

The public widely complains about obstacles related to registration or filing because when people have registered and uploaded files online through the M-Passport application, the complex files are still requested at the immigration office. The recommendation that can be given is that it is necessary to increase efficiency in processing and sorting the registration files of people who want to apply for a passport because it has been assisted by the M-Passport application.

The immigration office sets a quota of people who can apply for a passport daily, but to get the queue number, people must come early in the morning. In addition, sometimes people are left waiting in long queues without a queue number, and when it is time for them to be served, the quota runs out. People who have chosen the arrival schedule through the M-Passport application must wait a long time, which does not match their schedule. The recommendation that can be given is that if there is indeed a daily quota, the existing officers must be more assertive regarding giving queue numbers offline so that people who queue are those who have received a number. In addition, good integration is needed between offline queues and those who have taken a schedule or queue online so that the system with the queue number can be adequately realized.

Other factors not listed in the defect category, such as officers at the immigration office not picking up the phone from the public when needed, policies that are still unclear, and so on, also complained about by the public. Recommendations that can be given regarding this matter concern the existence of a particular officer or customer service that can answer and provide solutions to public complaints via chat or telephone.

The obstacles experienced by the public related to the photo process were such that the public had come according to their schedule to take photos, but when they got there, they were asked to wait for a long time. In addition, there was a complaint that he was expelled from the immigration office because he did not wear clothes according to the dress code, even though there was no prior notification. Recommendations that can be given are that the integration of the schedule of the online queue for taking photos must be improved, and things that must be known by the public when taking passport photos can be included in the M-Passport application or on important platforms owned by the Surabaya Immigration Office.

People complain about the M-Passport application, which is often down or in error, so people cannot access it even though, in some places, registration for passport processing is already required to be done online or through the application. The recommendation that can be given is that developers need to carry out regular monitoring of the M-Passport application, especially in parts that often experience errors, such as invalid OTP codes, inability to select the date of arrival, applications that usually suddenly exit themselves, and others.

Many complained that the car park was very far away, and they had to park in another place at an expensive cost when they did not get a place. Recommendations that can be given are the need to add facilities for an adequate car park that is large and not too far away, and if there is an application for parking fees, the fees are not too expensive.

In some places, people feel that there are areas or facilities of the immigration office that are not clean enough to make people uncomfortable. Recommendations that can be given are to conduct reminders or evaluations to cleaners at the immigration office to be able to carry out their duties regularly and periodically, considering that people who come to the immigration office to take care of passports continue to take turns from morning to evening. In addition, warnings related to "maintaining cleanliness" in certain places should be given so that the public can see if the five immigration offices in Surabaya have not been widely applied.

IV. CONCLUSION

This research demonstrates the impactful application of Convolutional Long Short-Term Memory (Co-LSTM) and p-attribute control charts in evaluating and enhancing the quality of services provided by the Surabaya Immigration Office. Our findings indicate a substantial majority of positive public sentiment towards the services offered, as evidenced by the sentiment analysis, which classified 90.1% of reviews as positive. The high classification accuracy, with an AUC value of 98.61% for training data and 85.66% for testing data, underscores the effectiveness of the Co-LSTM method in discerning public sentiment from online reviews.

Applying p-attribute control charts further allowed for a nuanced understanding of service quality over time, revealing statistically controlled processes and highlighting areas of concern requiring attention. The most significant issues identified relate to service personnel's perceived lack of friendliness and politeness, underscoring the need for targeted improvements in customer service training and evaluation.

To improve the quality of service and customer satisfaction at the Surabaya Immigration Office, our research recommends several actionable strategies. First, staff training programs should be implemented to focus on providing excellent customer service. This will address the primary concerns raised about staff conduct and professionalism. Second, the service processes should be reviewed and optimized to reduce wait times and improve efficiency. This could involve streamlining queue management and improving how services are delivered. Third, the Surabaya Immigration Office should continuously improve the usability and reliability of its digital platforms. This includes appointment scheduling and service applications. The office can better serve its customers by minimizing obstacles and improving user satisfaction. Fourth, improvements should be made to the facilities to ensure they are clean and adequate and create a more welcoming environment for those using the services. Finally, a robust feedback mechanism should be established. This would allow for real-time reporting and resolution of issues, ultimately creating a more responsive and adaptable service environment.

For future research, expanding the dataset and employing a broader array of sentiment analysis and control chart methods could offer deeper insights and more exact recommendations for service improvement. Additionally, exploring the impact of specific interventions on public sentiment and service quality metrics would provide valuable feedback on the effectiveness of the implemented strategies.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the Institute Teknologi Sepuluh Nopember for this work under the Inbound Researcher Mobility (IRM) 2024 project scheme.

REFERENCES

[1] BPS, Statistik Wisatawan Nasional 2022, Jakarta: Badan Pusat Statistik, 2023.

[2] H. Ditjenim, "Aplikasi M-Paspor Siap Digunakan di Seluruh Indonesia Mulai 27 Januari 2022," 2022. [Online]. Available: https://www.imigrasi.go.id/id/2022/01/23/aplikasi-m-paspor-siap-digunakan-di-seluruh-indonesia-mulai-27-januari-2022/.

[3] R. K. Behera, M. Jena, S. K. Rath and S. Misra, "Co-LSTM: Convolutional LSTM Model for Sentiment Analysis in Social Big Data," *Information Processing and Management,* vol. 58, no. 1, 2021.

[4] M. Mashuri, M. Ahsan, M. H. Lee and D. D. Prastyo, "PCA-based Hotelling's T2 chart with Fast Minimum Covariance Determinant (FMCD) Estimator and Kernel Density Estimation (KDE) for Network Intrusion Detection," *Computers & Industrial Engineering,* vol. 158, p. 107447, 2021.

[5] M. Imran, J. Sun, X. Hu, F. S. Zaidi and A. Tang, "Investigating zero-state and steady-state performance of MEWMA-CoDa control chart using variable sampling interval," *Journal of Applied Statistics,* pp. 1-22, 2023.

[6] A. Zaka, M. Naveed and R. Jabeen, "Performance of attribute control charts for monitoring the shape parameter of modified power function distribution in the presence of measurement error," *Quality and Reliability Engineering International,* vol. 38, no. 2, pp. 1060-1073, 2022.

[7] A. Karenina, "Sentiment Analysis on User Reviews of Mobile Banking BRIMO Using Naive Bayes Classifier and Support Vector Machine Models," Institut Teknologi Sepuluh Nopember, Surabaya, 2023.

[8] N. T. Apsari, M. Ahsan and M. H. Lee, "Monitoring the Quality of PeduliLindungi Application based on Customer Reviews on Google Play Using Hybrid Naive Bayes - Laney p' Attribute Control Chart," *International Journal on Advanced Science Engineering Information Technology,* vol. 13, 2023.

[9] D. Firmansyah, "Monitoring Kualitas pada Aplikasi MyPertamina Berdasarkan Ulasan Pengguna di Google Play Menggunakan Diagram Kendali Laney p' Berbasis Convolutional Neural Network," Institut Teknologi Sepuluh Nopember, Surabaya, 2022.

[10] P. Naila Adelia and A. Muhammad, "Monitoring Quality of KAI Access Application Based on Customer Reviews on Google Play Store Using Laney p'Control Chart Based on Convolutional Neural Network," *5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023),* pp. 175-184, 2023.

[11] S. Kim, S. Park and M. Kwak, "Review-Based Control Charts For Service Quality Monitoring: A Brief Review And Future Directions," *ICIC International,* pp. 707-714, 2020.

[12] T. Zhang, Z. He and A. Mukherjee, "Monitoring negative sentiment scores and time between customer complaints via one-sided distribution-free EWMA schemes," *Computers & Industrial Engineering,* 2023.

[13] M. Muchson, Statistika Deskriptif, Guepedia: Jakarta, 2012.

[14] D. Montgomery, Introduction to Statistical Quality Control, 7th ed, New York: Wiley and Sons Inc, 2013.

[15] E. Senave, M. J. Jans and R. P. Srivastava, "The Application of Text Mining in Accounting," *International Journal of Accounting Information System,* 2023.

[16] R. G. Katryn, "Text Preprocessing: Tahap Awal dalam Natural Language Processing (NLP)," 2020. [Online]. Available: https://medium.com/mandiri-engineering/text-preprocessing-tahap-awal-dalam-natural-language-processing-nlp-bc5fbb6606a.

[17] S. Kala, "Data Visualization: Word Clouds with Python," Juli 2020. [Online]. Available: https://medium.com/analytics-vidhya/data-visualization-word-clouds-with-python-fbb6395be18f.

[18] S. R. Goniwada, "Sentiment Analysis," in *Introduction to Datafication*, Apress, Berkeley, CA, 2023, pp. 1-14.

[19] S. Lai, X. Hu, H. Xu, Z. Ren and Z. Liu, "Multimodal Sentiment Analysis: A survey," *Displays,* vol. 80, 2023.

[20] K. Devi, "Understanding Hold-Out Methods for Training Machine Learning Models," 14 August 2023. [Online]. Available: https://www.comet.com/site/blog/understanding-hold-out-methods-for-training-machine-learning-models/. [Accessed 23 February 2024].

[21] S. Kulkarni, "Word Embedding using Word2Vec," 3 January 2024. [Online]. Available: https://www.geeksforgeeks.org/python-word-embedding-using-word2vec/. [Accessed 25 February 2024].

[22] D. Montgomery, Introduction to Statistical Quality Control, 8th ed, Hoboken: John Wiley & Sons, Inc, 2020.

[23] S. Turney, "Probability Distribution | Formula, Types, & Examples," 21 June 2023. [Online]. Available: https://www.statisticalaid.com/binomial-distribution-definition-density-function-properties-and-application/. [Accessed 25 February 2024].

[24] K. I. K. I. K. T. Surabaya, "Visi dan Misi," 2022. [Online]. Available: https://kanimsurabaya.kemenkumham.go.id/tentang-kami/visi-dan-misi/.

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

[26] R. N. Cikania, "Analisis Sentimen Review Pengguna Layanan Telemedicine Halodoc Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine," Institut Teknologi Sepuluh Nopember, Surabaya, 2021.

[27] M. Romadhona, "Analisis Klasifikasi Sentimen Review Mobile JKN Menggunakan Naïve Bayes Classifier dan Support Vector Machine," Institut Teknologi Sepuluh Nopember, Surabaya, 2021.

[28] J. Kim and C. Lim, "Customer Complaints Monitoring with Customer Review Data Analytics: An Integrated Method of Sentiment and Statistical Process Control Analyses," *Advanced Engineering Informatics,* pp. 2-13, 2021.

[29] M. Ahsan, M. Mashuri and H. Khusna, "Kernel Principal Component Analysis (PCA) Control Chart for Monitoring Mixed Non-Linear Variable and Attribute Quality Characteristics," *Heliyon,* vol. 8, no. 6, 2022.

[30] M. Ahsan, M. Mashuri, H. Kuswanto, D. D. Prastyo and H. Khusna, "Outlier Detection Using PCA Mix Based T2 Control Chart for Continous and Categorical Data," *Communications in Statistics - Simulation and Computation,* vol. 50, no. 5, pp. 1496-1523, 2021.

[31] A. A. Aly, N. A. Saleh and M. A. Mahmoud, "An adaptive EWMA control chart for monitoring zero-inflated Poisson processes," *Communication in Statistics,* vol. 51, no. 4, pp. 1564-1577, 2022.

[32] A. Johannssen, N. Chukhrova and P. Castagliola, "The performance of the hypergeometric np chart wih estimated parameter," *Stochastics and Statistics,* vol. 296, no. 3, pp. 873-899, 2022.