

Analysis and Notification Alert of Complaints of Health Service Recipients

Preeyanoot Moontee¹ and Trasapong Thaiupathump^{2,*}

¹ Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand

² Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

Preeyanoot_moontee@cmu.ac.th

Abstract. Complaints are expressions of dissatisfaction or grievances about a service that can be used to improve and develop the quality of service. However, there are several levels of complaints according to the level of severity, such as dividing the severity into two levels: the level that did not require warning and the level to be notified to corresponding personnel and require immediate actions to solve the problems. The purpose of this independent study was to study the analysis and notification of complaints in health care services by using complaints from health care recipients and the severity of complaints assessed by experts. The researcher obtained the information from the study to develop a system for analyzing complaints and notifying when there are complaints that are categorized as having to be notified through the LINE application. The study found that Multinomial Naïve Bayes had the highest efficiency in complaint classification compared to the Accuracy value of 71%.

Keywords: Natural language processing, Risk management, Complaint, Line notify

1 Introduction

Risk management is a operations with the goal of reducing the impact that will affect the loss of the organization [1]. A medical facility has processes for handling complaints arising from the provision of care services or the performance of personnel as a guideline for practice and systematically improve the quality of service to reduce the likelihood of litigation cases. Then, the gathered complaints will be analyzed and the notification through the Line application. Therefore, the involved associated parties can resolve complaints that are likely to affect the organization the most as the first priority to avoid negative consequences that will affect the reputation and trust of service users.

* Corresponding author.

2 Literature Review

2.1 Guidelines for managing complaints

Complaint refers to a complaint/suggestion/comment arising after a service user has received the service or has used the service. Complaint management refers to the process of handling complaints at a level acceptable to the organization and set reasonable rotection measures and achieve the specified objectives [1].

2.2 Natural language processing (NLP)

Natural language processing is a machine learning technology which enables computers to interpret, manipulate, and understand human language [2]. PyThaiNLP is a text processing and analysis of the Thai language library which has a wide range of functions that can be used to develop models to solve complex Thai natural language processing problems [3].

3 Data and Methodology

3.1 Data

The data used for the analysis were obtained from the collection of complaints arising from the dissatisfaction of service users in many hospitals in Chiang Mai. The collected data consisted of 1,783 complaints.

3.2 Methodology

This study used the information from web scraping of publicly available data on the internet to analyze and notify the complaints of the health service users. The collected data were labeled with the severity levels: 0 represents the complaints that do not require notification and 1 represents the complaints that require notification as shown in Table 1. The ratio of training set and test set is 85:15.

Severity Level	Number of Complaints	Percentage
0	1,129	63.32
1	654	36.68

Table 1 The number of complaints classified by level of severity

It is clearly seen that the data is imbalanced data. The imbalanced data problems for training set can be solved by using oversampling method to increase the amount of data in the minority group to be close to amount of data in the majority group. Table 2 shows the proportion of data before and after mitigating the imbalanced data.

Severity Level	Before Adjusting	After Adjusting
0	968 (63.9%)	1,936 (46.9%)
1	547 (36.1%)	2,188 (53.1%)

Table 2 The number of the training set data before and after adjusting the data balance

In the process of feature extraction Term-Frequency Inverse-Document Frequency, or TF-IDF was used to separate words based on their importance in a message.

This study used machine learning technique to classify the data to obtain the best answer. The researchers therefore created four models: Logistic Regression, Multinomial Naïve Bayes (Multinomial NB), Linear Support Vector Classification (LinearSVC), and Random Forest. Notifications were connected to Line notify, provided that “any message after analyzing the data and interpreting the result as 1, the complaint will send a notification to the LINE group of people involved in complaint management by showing details such as the date of the complaint and the details of the complaint”.

4 Result

4.1 Comparison of model performance

From modeling complaint classification by feature extraction method with TF-IDF, the performance of each model was shown in Table 3:

Model	%Accuracy	%Recall	%Precision	%F1 score
Logistic Regression	70	78.57	68.32	73.09
Multinomial NB	71	87.39	60.25	71.32
LinearSVC	65	73.29	66.46	69.71
Random Forest	51	86.05	22.98	36.27

Table 3 Results from model test

The results of the model test showed that, when compared to other models, the Multinomial NB had the highest efficiency in classification of complaints with the accuracy of 71%. The F1 score indicates the performance of the multinomial NB model calculated from the Precision and Recall values, with the higher the F1 score indicating the better model performance. When considering the F1 score value, type 0 complaints was 71%, while type 1 complaints were 70%. It indicates that Multinomial NB can classify each type of complaint in a good way from all results.

4.2 Notification of Complaints

The notifications were sent to Line Notify to notify complaints to the Line group of those involved by specifying details, including the date of complaint and complaint details, can be displayed as in Figure 1.

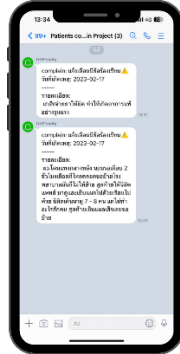


Figure 1 Complaint notification display screen

5 Discussion and Conclusion

5.1 Conclusion

The Multinomial Naïve Bayes (Multinomial NB) model using Term Frequency Inverse Document Frequency (TF-IDF) feature extraction and oversampling asymmetry was the most efficient model in the analysis to classify health care complaints in this study with the accuracy of 71%. However, the model still has mispredictions, which are often found in complaints containing metaphorical phrases or idioms. It was possibly because these complaints were rarely found.

5.2 Recommendations

Each medical facility might have different specificities. Therefore, the selected model must be modified to suit the data set of each hospital for the most accurate classification performance. For system developers, it can be used as a supplementary system (Plugin) to be used to add capabilities to the main program.

References

- [1] Faculty of Medicine Siriraj Hospital, Mahidol University. (2021). *Guidelines for managing complaints of the Faculty of Medicine Siriraj Hospital Mahidol University*. Retrieved May 19, 2023, from: <https://cmu.to/IOpF2>
- [2] AWS. *What is NLP?*. Retrieved February 7, 2023, from: <https://aws.amazon.com/th/what-is/nlp/>
- [3] Phillip Smith. (2018). *Sentiment analysis of patient feedback*. Retrieved May 19, 2022, from: https://theses.bham.ac.uk/id/eprint/7406/5/Smith_Phillip17