JIL Journal of Information and Learning, (2024), 35(3), 52-63
Journal homepage: https://so04.tci-thaijo.org/index.php/jil

Research Article

Implementing Sentiment Analysis to Enhance Service Quality in Luxury Hospitals: A Case Study in Thailand

Supattalak Phoha and Sarawut Ramjan*

College of Innovation, Thammasat University

*Corresponding author: sarawut@citu.tu.ac.th

Received June 19, 2024 ■ Revised August 23, 2024 ■ Accepted August 28, 2024 ■ Published December 24, 2024

Abstract

This research intends to improve the service quality of a luxury hospital by accurately classifying a customer's feelings. The data set was collected from customers' online reviews of a hospital certified by JCI standards. This experiment relies on five data mining techniques, including Naive Bayes, Decision Tree, k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Deep Learning. SVM has the highest accuracy at 89.71% and confirms the accuracy of the mining model in the real world by 30 experts. Therefore, the luxury hospital could use the SVM model to classify customer feedback and prepare better services for its customers.

Keywords: sentiment analysis, luxury hospital, service quality, Joint Commission International (JCI)

Introduction

Hospital can use lessons learned from luxury hotels, where they use data analytics to identify touch points that need improvement (Brutus, 2023). Joint Commission International (JCI) is a service standard that is used in luxury hospitals to improve service quality at each touchpoint to meet customer expectations (Hashemi et al., 2014; Cece & Köse, 2022; He et al., 2024). This standard covers many service dimensions, such as safety and facilities, and influences tourists who consider traveling to different countries (Asl et al., 2014; Asl et al., 2015). Thus, JCI standards could be used to attract premium customers to use the services of luxury hospitals.

By embracing JCI standards, luxury hospitals stay at the forefront of healthcare, continuously improving to meet patient expectations, such as reducing operating room times (Irvan Masoudi et al., 2015; Kitsirichawanan & Mongkolsawat, 2019). This focus on quality not only elevates patient satisfaction, but also boosts staff productivity and motivation (Tarieh et al., 2022), contributing to a healthier and more efficient healthcare environment. To evaluate their performance against the key performance indicators (KPIs) of the standard, hospitals often rely on the voice of the customer, gathering feedback through various channels. Users frequently interact with hospitals through apps like Facebook pages and hospital chatbots, providing a steady stream of feedback and comments (Afifi & Amini, 2018; Cooke et al., 2021). This approach allows hospitals to capture real-time insights into patient experiences and gauge service quality, contributing to a hospital's short-term learning and ongoing adjustments (Liao et al., 2023). However, amid the overwhelming amount of information on social media and hospital apps, some statements are nonsensical or irrelevant while others contain valuable feedback that could be used to improve service quality (Liang et al., 2013). Therefore, big data analysis is a challenge for hospitals.

Sentiment analysis could be used for automatically classifying content into binary data, including positive and negative feelings. Hence, hospitals can use this technique to reduce staff effort (Paruchuri et al., 2021). Because Thailand has a unique culture and language, the experiment is focused on the Thai language to analyze Thai language patterns from customer online reviews to improve model accuracy. This research presents a sentiment analysis model that classifies customer feelings and has the following research objectives:

1. To develop a sentiment analysis model that can classify customer feelings from a large scale of customer online reviews.
2. To use a sentiment analysis model to classify customer feelings from online customer reviews to improve the service of luxury hospitals.

The research results can help hospitals to improve service quality according to customer feedback and enhance the customer experience with luxury-level service quality. Therefore, this model can be used as a tool to support hospitals that intend to be certified by the JCI standard, as well as those aiming to transform into luxury hospitals.

Literature Review

1. Healthcare Quality Standards, service quality management is needed for hospitals (Mogakwe, Ally, et al., 2020; Mogakwe, Magobe, et al., 2020). On the other hand, doctors and nurses are used to operating treatments with a good process (Gow et al., 2011). Hospital staff prefer to operate in an environment that encompasses positive perceptions from colleagues.

From the patient's point of view, the service performance of a hospital is reflected in its congestion management among many patients (Munyewende et al., 2014). For patients, hospitals that effectively manage overcrowding in emergency rooms signal superior treatment capabilities (Hwang et al., 2008). The coveted JCI standard serves as a benchmark, augmenting the job satisfaction of medical staff (He et al., 2024), and its implementation helps in attracting health tourists to countries with certified hospitals (Campra et al., 2021). Within the realm of JCI standards, quality enhancement emerges as a critical evaluation criterion, aiming to instill patient confidence and to enhance the hospital's reputation (HongFan et al., 2023). Notably, Thailand boasts upscale hospitals certified according to JCI standards, underscoring its commitment to quality healthcare ("Bumrungrad International Hospital Receives Its 7th Joint Commission International (JCI) Re-Accreditation," 2023).

2. Sentiment Analysis in Healthcare: Healthcare professionals are actively improving their services to address escalating public health worries. They are achieving this by closely monitoring public sentiment on social media platforms and by utilizing sentiment analysis techniques to assess whether patient opinions express positivity or negativity (Usman et al., 2024). Additionally, they are examining service ratings on hospital websites to gain further insights (Asghar et al., 2016). The sentiment analysis of patients' opinions regarding the medication used for their treatment revealed a distinction in the choice of words used to express their feelings about the medication compared to general topics. Even though the wording is similar, the connotations vary when discussing medication, expressing different sentiments than when discussing general matters (Goeuriot et al., 2012). Once staff monitors customer feedback from an online review system, they can determine whether the content is positive or negative. This helps the hospital improve service quality based on the suggestions (Khaleghparast et al., 2023). To define a data mining technique suitable for sentiment analysis of Thai luxury hospitals, this research will experiment with the following 5 algorithms:

2.1 Naive Bayes is an algorithm that follows Bayes' theorem, and it has efficiency for data classification (Alwateer et al., 2021). Naïve Bayes is used for sentiment analysis about service quality evaluation of hospitals by classifying the feelings of patients (Gill et al., 2023). Although user experience involves complex emotions, Naïve Bayes can be used to capture wording that reflects the feelings of patients about hospital service (Mir Habeebullah Shah & Dr, 2020; Rahim et al., 2021). Thus, the Naïve Bayes algorithm is as follows:

- *Training phase:* (1)
  - Prior Probability:

$$P(Y = c) = \frac{\text{Number of instances with class } c}{\text{Total number of instances}}$$

  - Likelihood Calculation:

$$P(X_i = x_i \mid Y = c) = \frac{\text{Number of instances with } X_i = x_i \text{ and } Y = c}{\text{Nubmer of instances with } Y = c}$$

- *Prediction Phase:*
  - Class Prediction:

$$\hat{Y} = \frac{\arg\max}{c} (P(Y = c) \prod \frac{n}{i = 1} P((X_i = x_i \mid Y = c))$$

To evaluate the Naive Bayes performance, this research uses an evaluation matrix including accuracy, precision, recall, and F1 score. For the experiment, this research used a multinomial Naive Bayes model and set the Laplace smoothing parameter to 0.001.

2.2 Decision Tree is a technique that is suitable for analyzing customer opinion about hospital services (Chakrapani et al., 2023). Feature selection is used to identify a service touch point that influences customer satisfaction such as staff personality (Nurfaizah et al., 2019). Therefore, Decision Tree will help a hospital to identify a service touch point that should be improved to enhance the customer experience (Dangi et al., 2022). There is an algorithm of Decision Tree as follows:

- *Splitting Criteria:* (2)
  - Entropy (H):
    $H(S) = -\Sigma_{i=1}^{n} p_i \log_2(p_i)$
  - Information Gain (IG):
    $GI(S,A) = H(S) - \Sigma_{vEValues(A)} \frac{|S_V|}{|S|} H(S_V)$
- *Tree Construction:*
  - Recursive Splitting

The Decision Tree algorithm starts with a base case: if all samples in a node are of the same class, it makes a leaf node with that class label. Next, it chooses the feature with the highest information gain to split the data into subgroups based on the value of the chosen feature. This process is done again for each subgroup until a stop condition happens, such as reaching the maximum tree depth or the minimum number of samples per leaf. This repetitive method allows the Decision Tree to learn and classify data well. Every step is important for building a model that can predict results correctly.

  - Prediction

For predicting a new sample, the process starts with traversing the decision tree. It uses the feature value of the sample to guide along the tree branch until it reaches a leaf node. At the leaf node, classification is done by looking at the majority class of samples in that leaf. This majority class then becomes the predicted class for the new sample. This method helps predict results for new data effectively by using the structure and learning from the decision tree.

For the decision tree experiment, this research set the parameters including the criterion as 'gini', the splitter as 'random', the maximum depth was 1000, the maximum features as 5000, and the random state set to 100.

2.3 K-Nearest Neighbor (k-NN) is suitable for analyzing customer opinions of those who use hospital services (Chakrapani et al., 2023). For analyzing an online customer review, k-NN is a data mining technique that can be used for identifying critical service touch points for a hospital (Navele et al., 2022). The k-NN algorithm is as follows:

- *Distance Metric:* (3)
  - Euclidean Distance:
    $d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2}$
- *Find K Nearest Neighbors:*
  - *Query Point* x_q
  - Training Data X:
    $X = \{x_1, x_2, ..., x_N\}$
  - Distance Calculation:
    $D = \{d(x_q, x_1), d(x_q, x_2), ..., d(x_q, x_N)\}$
  - Sort Distance: *Sort(D)*
  - K Nearest Neighbors:
    $NN = \{x_{1NN}, x_{12NN}, ..., x_{K_{NN}}\}$
- *Prediction:*
  - Classification (Majority Vote):
    $\hat{y}_q = \arg\max_{c} \sum_{i=1}^{k} I(y_i = c)$
  - Regression (Mean):
    $\hat{y}_q = \frac{1}{K} \sum_{i=1}^{k} y_1$

For the experiment with K-Nearest Neighbors, this research set the value of k to 3 and used distance-based weighting.

2.4 Support Vector Machine: SVM is a supervised learning technique that can be used for data classification as well as regression. SVM is considered one of the most trusted algorithms (Anam et al., 2023). In natural language processing (NLP), SVM can be adopted to analyze text (Hayurian & Hendrastuty, 2024). Thus, the SVM algorithm is as follows:

$$f(x) = w \cdot x + b \quad (4)$$

Where:

x is the input feature vector.
w is the weight vector.
b is the bias term.

In the experiment with the dataset that using a Support Vector Machine, this research used the radial basis function (RBF) as the kernel and set the regularization parameter C to 1.0.

2.5 Deep learning is constructed following brain structures. There are many layers of processing, which support complex problem solving (Prakash et al., 2024). For sentiment analysis, LSTMs and RNNs are used to analyze a wide spectrum of human aspects (Botta et al., 2024). The overall process can be summarized as:

Input: sequence of word embeddings x (5)

- LSTM Processing: $h_t = \mathrm{LSTM}(x_t, h_{t-1})$ for $t = 1$ to T
- Final Hidden State: $h_T = \mathrm{LSTM}(x_T, h(_{T-1}))$
- Fully Connected Layer: $z = W \cdot h_T + b$
- Output: $\hat{y} = \sigma(z)$

For the experiment with the dataset that using a deep learning technique, this research set the parameters as follows: a batch size of 32, 10 epochs, and an LSTM network with 2 layers that the first layer with 128 units and the second layer with 68 units.

Table 1

Classification Model Evaluation Matrix

Items	Meaning	Citations
Precision	Precision measures the accuracy of positive predictions. It is defined as the ratio of true positive predictions to the total number of positive predictions made.	(Streiner & Norman, 2006)
Recall	Recall, also known as sensitivity, measures the ability of a model to identify all relevant instances within a dataset. It is calculated as the ratio of true positives to the sum of true positives and false negatives.	(Yacouby & Axman, 2020)

Table 1
(continued)

Items	Meaning	Citations
F1-score	The F-Score is the harmonic means of precision and recall. It is used when you need to balance precision and recall and is particularly useful in scenarios where an imbalance in the dataset might make one measure more important than the other. The F1 Score reaches its best value at 1 (perfect precision and recall) and worst at 0.	(Fujino et al., 2008)
Support	Support is the number of actual occurrences of the class in the specified dataset. It indicates the size of the dataset that was used to calculate the precision, recall, and F1-Score for each class.	(Yacouby & Axman, 2020)

3. Related Research on Identifying Gaps in Existing Knowledge: This research highlights the value of using social network data, like Twitter and Facebook, for Sentiment Analysis to understand consumer sentiments and create economic value, as noted by other studies in 2021 (Fahd et al., 2021). Real-time data analysis is essential for businesses to stay informed about consumer behavior, especially in sectors like healthcare, where customers frequently seek information online about services. Automatic sentiment analysis of customer feedback can help healthcare providers improve services and maintain a competitive edge, while also aiding customers in making informed decisions.

One referenced study developed a sentiment analysis model for healthcare reviews using LSTM networks, achieving 91% accuracy, outperforming the Naïve Bayes model (Pradhan & Panda, 2023). Another study found that younger and higher-income service recipients perceive private hospitals as offering better service quality than public hospitals (Yarimoglu & Ataman, 2022). Additionally, a study applied Sentiment Analysis to assess feedback on a hospital's m-health app, revealing that system quality, which accounted for 65.07% of reviews, had the highest rate of negative sentiment, primarily due to data security concerns (Pandesenda et al., 2020).

These findings emphasize the importance of sentiment analysis in luxury hospitals in Thailand, addressing challenges like data security and real-time processing. There is a gap in research focused on luxury hospitals, which this study aims to fill by using advanced sentiment analysis to enhance service quality and offer actionable insights for high-end clientele.

Despite the recognized value of sentiment analysis, there is a gap in research focusing on healthcare settings, particularly in non-English contexts like Thailand (Denecke & Deng, 2015; Menaouer et al., 2022). Existing studies tend to concentrate on English-language feedback, leaving a gap in understanding when it comes to unique linguistic and cultural nuances (Xu et al., 2022). Additionally, few studies compare different sentiment analysis algorithms to determine which is best suited for healthcare-related data (Khanam & Sharma, 2021; Sharma et al., 2023).

4. In hypothesis development, this research identifies a gap in the research from the literature review. The experiment will evaluate the efficiency of 5 algorithms, including Naive Bayes, decision tree, k-nearest neighbors, support vector machine, and deep learning, to identify the best algorithm that fits the characteristics of the dataset in the context of Thai luxury hospitals. Moreover, this research will apply the best algorithm to a test with a group of domain experts to prove that the model could be used in the real world. Based on the literature review and the identified gaps, the following hypotheses are formulated:

Hypothesis 1: Sentiment analysis can classify customer feedback in hospitals with high accuracy, providing useful information about hospital service quality.

Hypothesis 2: From Naive Bayes, decision tree, k-nearest neighbors, support vector machine, and deep learning, one algorithm will show the best accuracy for hospital service quality data.

Hypothesis 3: The sentiment analysis model will help hospitals make feedback monitoring easier, so they can respond better and follow JCI standards more effectively.

Research Methodology and Results

This research is a data analytics experiment that covers data exploration, data pre-processing, data mining, and data interpretation (Chu, 2014; Ramjan & Sunkpho, 2023). This research will compare the accuracy of each data mining technique to identify the best algorithm that fits the dataset. Then, the researchers will apply the best algorithm in a field experiment to validate the efficiency of the algorithm when adopted in the real world (Enrico et al., 2021; Shimada et al., 2017).

1. Data Collection and Exploration: This research intends to improve the efficiency of machine learning using big data. It investigates a potential data source concerning hospitals accredited by the Joint Commission International (JCI) in 2022. This research gathered data from Google about online customer reviews for 57 hospitals and from Honest Docs about online customer reviews for 37 hospitals. From data exploration, the Google dataset has many features, such as user account, content, and review score, with a total of 10,280 instances. On the other hand, the Honest Docs dataset has features including hospital name, phone number, review score, publishing date, and content, with a total of 2,548 instances. Then, those datasets were combined into a new dataset that included 12,828 instances. During the data exploration process, this research classified the dataset according to the review score, which is the label feature. This class consists of 5 levels: 5 stars for extreme satisfaction, 4 stars for high satisfaction, 3 stars for neutral, 2 stars for low satisfaction, and 1 star for very low satisfaction.

2. Data preparation, this research related to natural language processing (NLP) which is used to identify that content is positive, negative, or neutral.

2.1 Text Pre-Processing, Thai language has a unique characteristic, including spaces between words, no full stops, and frequent use of emoticons. Therefore, text pre-processing needs to be designed with a specific process for the Thai language. These include removing emoticons, eliminating special characters, content tokenization, and performing word tokenization. Removing emoticons and special characters from text helps to focus on the alphanumeric content and simplifies it. Content tokenization breaks down text into manageable pieces, such as words or sentences, which is essential for detailed analysis. Specifically, word tokenization splits text into individual words, facilitating deeper linguistic and statistical analysis. The text preprocessing is carried out using the PyThaiNLP Library from Python. Examples of these processes are shown in Table 2.

Table 2

Example Results of Data Preprocessing with the PyThaiNLP Library

Text Pre-Processing Processes	Original Data	Thai to English Translation of Original Data	Pre-Processed Data	Thai to English Translation of Pre-Processed Data
Emoticon Removing	โรงพยาบาลสะอาด ❤ มีมาตรฐาน คุณหมอ พยาบาลใจดี	A clean hospital that meets standards (❤), staffed by kind doctors and nurses.	โรงพยาบาลสะอาด มีมาตรฐาน คุณหมอ พยาบาลใจดี	A clean hospital that meets standards, staffed by kind doctors and nurses.

Table 2
(continued)

Text Pre-Processing Processes	Original Data	Thai to English Translation of Original Data	Pre-Processed Data	Thai to English Translation of Pre-Processed Data
Special Characters Removing	ไม่ใช้บริการอีกแน่ๆ ๆๆๆๆ	I will definitely not use the service again!แน่ๆๆๆๆๆ = definitely, certainly, for sure (the repetition of the letter "ๆ" adds emphasis)	ไม่ใช้บริการอีกแน่	I definitely won't use the service again.
Content Tokenization	รอนานมากๆ มาคิวแรกๆ แต่คนมาหลังได้ไปตรวจสุขภาพ	I had to wait a very long time. Although I was first in line, people who arrived after me were called for their health check-ups before me.	รอ คิว คน ตรวจ สุขภาพ	รอ (ror) = wait คิว (khio) = queue คน (khon) = person/ people ตรวจ (truat) = check/ examine สุขภาพ (suk-kha-phap) = health
Stop Word Removing	โรงพยาบาลมีการบริการที่น่าประทับใจมาก	The hospital offers excellent services.	โรงพยาบาล บริการ น่าประทับใจ	โรงพยาบาล = Hospital บริการ = Service น่าประทับใจ = Impressive

Furthermore, this research aims to comprehensively ensure data quality by addressing spelling corrections, abbreviations, and slang. This is achieved by using dictionary lookup to compare words with a dictionary, correcting any misspelled words found. For handling abbreviations, an abbreviation dictionary is created to convert all abbreviations to their full forms. Subsequently, a slang dictionary is used to create a repository of slang terms and convert them to formal language. Through these methods, this research can improve the consistency and accuracy of the data for processing.

2.2 Feature Transformation: To enhance machine learning for classifying comments as positive or negative, this research conducts feature transformation by creating a new class from two existing classes: assigning five stars for extreme satisfaction and four stars for high satisfaction, indicating comments that are important and urgent. Three stars signify important but not urgent, two stars indicate not important but urgent, and one star represents unimportant and not urgent. Once the research categorized words into new classes, it counted the frequency of each word in each category to prepare the data for conversion into numerical form in the next step as illustrated in Table 3.

Table 3
Example of Words and Their Frequencies by Data Category

IU		INU		NIU		NINU	
โรงพยาบาล (Hospital)	2,049	รอ (Wait)	1,650	ดี (Good)	1,568	ดี (Good)	1,931
หมอ (Doctor)	1,828	หมอ (Doctor)	1,615	โรงพยาบาล (Hospital)	1,277	บริการ (Service)	1,713
รอ (Wait)	1,712	โรงพยาบาล (Hospital)	1,425	บริการ (Service)	1,173	โรงพยาบาล (Hospital)	1,278

2.3 Vocabulary Construction: After preprocessing the text, which contained 12,947 words, this research created a vocabulary list by assigning each unique word an ID, disregarding grammar, word frequencies, and order to extract features for text classification. As the dataset size increases, so does the number of features, resulting in structured data for analysis. The researcher employed the Term Frequency-Inverse Document Frequency (TF-IDF) method to identify distinctive words by combining the normalized frequency of words in a document (TF) with their importance across all documents (IDF) (Manning et al., 2008), as shown in Equations 6 - 8.

Term Frequency (TF) is the ratio of the number of times a particular word appears in the document to the total number of words in the document. It can be calculated as follows:

$$IDF(t, D) = \frac{\text{Frequency of term } t \text{ in document } d}{\text{Total number of tersm in document } d} \quad (6)$$

Inverse Document Frequency (IDF): This measures the importance of the term across a set of documents. The IDF of a particular word is the logarithm of the ratio of the total number of documents to the number of documents containing the term. It helps in diminishing the weight of terms that occur very frequently in the document set and in increasing the weight of terms that occur rarely. The IDF is calculated as follows:

$$IDF(t, D) = \log\left(\frac{\text{Total number of documents } |D|}{\text{Number of documents containing term } t}\right) \quad (7)$$

The TF-IDF score is then calculated by multiplying these two statistics:

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (8)$$

Here, t represents the term, d represents a document, and D is the corpus of documents. This formula gives a high weight to any term that is not very common across the document set but appears frequently in a particular document, thus helping to identify and weight the most characteristic words in each document.

The resultant TF-IDF value helps to isolate the words with unique features for each document as shown in the example in Table 4:

Table 4

Example of TF-IDF Values for Words in the Dataset

Document	Impression	Disgrace	Suffering	Hospitality
D00001	0.363	0.344	0.344	0.347
D00002	0	0	0	0
D00003	0.125	0	0.089	0.0429
D00004	0	0.229	0	0
D00005	0.320	0	0.344	0

3. Data Mining and Results: This research utilized five data mining techniques to determine the most effective method for the sentiment analysis of customer reviews for luxury hospital services. The Naive Bayes technique applied an algorithm from Equation 1, the Decision Tree used an algorithm from Equation 2, k-Nearest Neighbor employed an algorithm from Equation 3, support vector machine uses an algorithm from Equation 4, and deep learning employs algorithm from Equation 5. The results are displayed in Table 5.

Table 5 presents the results of applying five different data mining algorithms, Naive Bayes, Decision Tree, k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Deep Learning to the sentiment analysis of customer feedback in luxury hospital services. The evaluation metrics used are Precision, Recall, F1-score, and Accuracy. These metrics are critical in determining the performance of each algorithm in classifying the sentiment accurately.

Table 5
Results of Data Mining Algorithms

	Label	Precision	Recall	F1-score	Support
Naive Bayes	IU	0.72	0.75	0.74	659
	INU	0.82	0.88	0.85	602
	NIU	0.79	0.75	0.77	626
	NINU	0.79	0.74	0.76	679
	Accuracy	0.75175			
Decision Tree	IU	0.84	0.62	0.72	659
	INU	0.84	0.99	0.91	602
	NIU	0.79	0.94	0.86	626
	NINU	0.81	0.74	0.77	679
	Accuracy	0.81800			
K-Nearest Neighbor	IU	0.82	0.40	0.54	659
	INU	0.81	0.99	0.89	602
	NIU	0.67	0.95	0.78	626
	NINU	0.77	0.71	0.74	679
	Accuracy	0.75175			
Support Vector Machine	IU	0.84	0.88	0.86	659
	INU	0.97	0.96	0.97	602
	NIU	0.89	0.91	0.90	626
	NINU	0.88	0.82	0.85	679
	Accuracy	0.89166			
Deep Learning	IU	0.87	0.70	0.78	659
	INU	0.88	0.99	0.93	602
	NIU	0.80	0.96	0.87	626
	NINU	0.88	0.78	0.82	679
	Accuracy	0.85151			

Although Naive Bayes displays an overall accuracy lower than other data mining models, this model still shows quite good efficiency with Precision at 0.72–0.82, Recall at 0.74–0.88, and F1-score at 0.74–0.85. Thus, the overall accuracy of Naive Bayes is 75.18%. Naive Bayes shows consistent efficiency across all classes. Notice that the INU class has better precision and recall than other classes. This indicates robustness in Naive Bayes for text classification. The Decision Tree has good performance with Precision at 0.79–0.84, Recall at 0.62–0.99, and F1-score at 0.72–0.91. Overall performance is 81.80%. The INU and NIU classes have a high recall value, which reflects the performance of text classification. On the other hand, the IU class has a low recall value, which means it lacks efficiency in identifying the label class in the IU segmentation.

k-Nearest Neighbor (k-NN) reflects a wide range of performance values, with Precision at 0.67–0.82, Recall at 0.40–0.99, and F1-score at 0.54–0.89. Overall performance is 75.18%. The highest performance is displayed in the INU and NIU classes, with a high recall value. On the other hand, the UI class has a low recall value, which means it lacks efficiency in text classification. Therefore, the wide range of performance values displays moderate accuracy for k-NN. Support Vector Machine (SVM) shows the best performance with Precision at 0.84–0.97, Recall at 0.82–0.96, and F1-score at 0.85–0.97. Overall performance is 89.71%. Besides, SVM shows good performance in every class with high precision, recall, and F1-score values, especially in the INU class. This reflects that SVM is a robust and appropriate data mining model for Thai language, with a special focus on a luxury hospital dataset. Deep Learning models show strong performance, especially in the INU and NIU classes with high recall rates, indicating their capability to capture relevant instances effectively. The performance is relatively lower for the IU class, where recall is moderate.

Observed Discrepancies: the IU class shows significant variance in recall across models, with k-NN performing the worst (0.40) and SVM performing well (0.88). This indicates that some models struggle more with identifying all relevant instances in this class. All models perform exceptionally well in the INU class, particularly Decision Tree, k-NN, and SVM, which achieve near-perfect recall. This consistency suggests that the characteristics of the INU class are well-captured by the models. The NIU class also shows high recall across models, with k-NN and Deep Learning achieving particularly high scores. This suggests that these models are effective in identifying urgent but not important instances. NINU has a low level of efficiency across all data mining models. SVM shows the highest stability value. Recall and precision are low in some models, which reflects the inefficiency of each model in classifying data with unique characteristics, such as the Thai language. In contrast, decision tree and deep learning models perform better when classifying feelings from online customer reviews in each class. Therefore, SVM has the best overall efficiency compared to other data mining techniques, especially in the context of the Thai language in the environment of Thai luxury hospitals. Naive Bayes and k-NN show more variability, indicating their suitability might depend on the specific characteristics of the data and the importance of balanced performance across all classes.

The analysis results from Table 5 support Hypothesis 1, as the five algorithms show different accuracy levels in online customer review classification. The accuracy ranges from 75.18% for Naïve Bayes and k-NN to 89.17% for SVM, with high precision and recall values from all data mining models. This range indicates that hospitals can trust the performance of sentiment analysis for classifying online customer reviews. Thus, this result also supports Hypothesis 2. Among the five algorithms tested, the Support Vector Machine (SVM) demonstrated superior accuracy (89.17%), followed by Deep Learning (85.15%) and Decision Tree (81.80%).

4. Field experiment, this research employs a field experiment technique to evaluate the accuracy of SVM in the real world (Bandiera et al., 2011; Battisti, 2017). This step relies on the opinions of 30 domain experts from a Thai luxury hospital. These experts came from various domains, including digital marketing, public relations, and facility management in hospitals, to collect various aspects about online customer reviews. During the experiment, this research uses SVM to classify online customer reviews into four class labels, including important and urgent, important but not urgent, urgent but not important, and neither important nor urgent. On the other hand, domain experts reviewed the same online customer reviews to classify them into the four class labels as well.

During the experiment, each domain expert analyzed 10 different contents, as did the SVM, which analyzed the same dataset. Once comparing the results of data classification from SVM and the domain experts, this research found that the accuracy of SVM was 90%, while the result from the domain experts was 58.67%. Therefore, this difference in results shows the superior performance of SVM for data classification, as it can classify data into four class labels from different customer online reviews. Besides, during the experiment, the domain experts still struggled with complex content, such as content number 7 from the dataset, which most domain experts could not identify a label class for. However, one of the domain experts classified it correctly. This discovery confirms that SVM has high performance, and that machine learning still needs improvement in automatic data classification from domain experts.

Therefore, the experiment result supports Hypothesis 3, that SVM has higher accuracy than domain experts. This result comes from the different experiences of each domain expert, which influence data classification. On the other hand, SVM has higher stability in data classification. This evidence confirms that Thai luxury hospitals could use an SVM model to classify online customer reviews through an automatic process.

Conclusions and Recommendations

This research deals with data classification from online customer reviews. The experiment covers five data mining techniques, including Naive Bayes, Decision Tree, k-Nearest Neighbor, Support Vector Machine (SVM), and Deep Learning. The result shows that SVM has superior performance at 89.17%, indicating that SVM is the most suitable sentiment analysis model for the Thai language, especially in the environment of Thai luxury hospitals. The field experiment relies on 30 domain experts from various positions in Thai luxury hospitals. The experiment shows that SVM has higher accuracy than domain experts. This result confirms the efficiency of SVM in classifying online customer reviews. Therefore, Thai luxury hospitals could develop a real-time application by embedding the ability of SVM to classify real-time online customer reviews and respond to customer demands. Ultimately, they can improve the service quality of the hospital under JCI standards

To improve the efficiency of the sentiment analysis model, Thai luxury hospitals could use an updated dataset from online customer reviews to train machine learning models to support changes in customer behavior. Moreover, the data classification result should be approved by domain experts to prevent frustration from content misinterpretation. This process will ensure that the results from the data mining model align with human considerations. For further research, researchers could expand the experiment to other languages that might affect luxury hospitals, such as Chinese or Japanese. These languages are used by foreigners who travel to Thailand, supporting various language contexts and improving the efficiency of data classification. These solutions will ensure that Thai luxury hospitals can meet JCI standards and improve overall customer satisfaction.

Practical Implementation Challenges

Sentiment analysis faces an uncontrolled challenge from human error, such as misspelling or slang, which affects accuracy. To overcome this problem, the PyThaiNLP library was created to deal with the Thai language. Besides, big data is another issue. The scalability of cloud storage is a point that Thai luxury hospitals need to consider. In addition to high-performance processing, Thai luxury hospitals need to prepare a budget for implementing that technology. Although the integration of the sentiment analysis model into real-time data processing is a complex project, Thai luxury hospitals could design an API and middleware software to connect them. On the other hand, data privacy is a concerning issue, and hospitals should follow data protection protocols such as the PDPA law in Thailand. Lastly, a training program is needed for staff to create trust among key stakeholders in using this system. This planning will support Thai luxury hospitals in reaching higher service quality and responding to customer expectations.

References

Afifi, I., & Amini, A. (2018). Factors affecting relationship marketing in creating customer loyalty in the hospital services business. *ASEAN Marketing Journal, 10*(2), 91-108. https://scholarhub.ui.ac.id/cgi/viewcontent.cgi?article=1095&context=amj

Alwateer, M., Almars, A. M., Elhosseini, M. A., Areed, K. N., Haikal, A. Y., & Badawy, M. (2021). Ambient healthcare approach with hybrid whale optimization algorithm and Naïve Bayes classifier. *Sensors, 21*(13), 4579:1-4579:27. https://doi.org/10.3390/s21134579

Anam, M. K., Fitri, T. A., Agustin, A., Lusiana, L., Firdaus, M. B., & Nurhuda, A. T. (2023). Sentiment analysis for online learning using the lexicon-based method and the support vector machine algorithm. *ILKOM Jurnal Ilmiah, 15*(2), 290-302. https://doi.org/10.33096/ilkom.v15i2.1590.290-302

Asghar, M., Ahmad, S., Qasim, M., Zahra, S., & Kundi, F. (2016). SentiHealth: creating health-related sentiment lexicon using hybrid approach. *SpringerPlus, 5*, Article 1139. https://doi.org/10.1186/s40064-016-2809-x

Asl, I. M., Nasiri, T., Shams, L., & Hashemidehaghi, Z. (2014). Relationship between health care organization management standards of the joint commission international and health tourism in selected hospitals in Tehran. *International Journal of Travel Medicine & Global Health, 2*(1), 19-22. https://www.ijtmgh.com/article_33268_5b818dc2417025e2f1e73ccd1f9820a8.pdf

Asl, I. M., Nasiri, T., Dopeykar, N., Shams, L. (2015). Observance of the patient-oriented standards of joint commission international and its relation to attracting of health tourists in selected hospitals in Tehran, Iran. *Journal of Payavard Salamat, 9*(2), 192-202. https://payavard.tums.ac.ir/article-1-5719-en.html

Bandiera, O., Barankay, I., & Rasul, I. (2011). Field experiments with firms. *Journal of Economic Perspectives, 25*(3), 63-82. https://doi.org/10.1257/jep.25.3.63

Battisti, J. E. Y. (2017). Field experiments: Design, analysis and interpretation. *RAE-Revista de Administração de Empresas, 57*(4), 410-411. https://periodicos.fgv.br/rae/article/view/71365

Botta, A., Mohini, P. V., Nooka Raju, A., & Balaji, C. (2024, April 9-10). *Deep Learning for sentiment analysis in financial markets* [Conference Session]. 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Jamshedpur, India. https://doi.org/10.1109/ICRTCST61793.2024.10578359

Brutus, J. P. (2023, January 16). What can hospitals learn from luxury hotels? KevinMD.com. https://www.kevinmd.com/2023/01/what-can-hospitals-learn-from-luxury-hotels.html

Bumrungrad international hospital receives Its 7th joint commission international (JCI) re-accreditation. (2023, January 15). Contify Life Science News. https://www.bumrungrad.com/en/blog-news/7th-joint-commission-international-jci-re-accreditation

Campra, M., Riva, P., Oricchio, G., & Brescia, V. (2021). Association between patient outcomes and joint commission international (JCI) accreditation in Italy: an observational study. *Quality Access to Success, 22*(181), 93-100. https://iris.uniupo.it/retrieve/e163b890-224a-c12c-e053-d805fe0a7f21/QAS_Vol.22_No.181_April.2021_p93-100%20%281%29.pdf

Cece, S., & Köse, İ. (2022). Comparative analysis of joint commission international and healthcare information and management systems-electronic medical record adoption model measurement models using text mining. *Journal of Health & Nursing Management, 9*(3), 371-382. https://doi.org/10.54304/SHYD.2022.37801

Chakrapani, K., Kempanna, M., Safa, M. I., Kavitha, T., Ramachandran, M., Bhaskar, V., & Kumar, A. (2023). An enhanced exploration of sentimental analysis in health care. *Wireless Personal Communications, 128*(2), 901-922. https://doi.org/10.1007/s11277-022-09981-8

Chu, W. W. (2014). Data mining and knowledge discovery for big data: Methodologies, challenge and opportunities. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40837-3

Cooke, M., Richards, J., Tjondronegoro, D., Raj Chakraborty, P., Jauncey-Cooke, J., Andresen, E., Theodoros, J., Paterson, R., Schults, J., Raithatha, B., Wilson, S., & Alcock, M. (2021). myPainPal: Co-creation of a mHealth app for the management of chronic pain in young people. *Informatics for Health & Social Care, 46*(3), 291-305. https://doi.org/10.1080/17538157.2021.1892697

Dangi, D., Dixit, D. K., & Bhagat, A. (2022). Sentiment analysis of COVID-19 social media data through machine learning. *Multimedia Tools & Applications, 81*(29), 42261-42283. https://doi.org/10.1007/s11042-022-13492-w

Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence In Medicine, 64*(1), 17-27. https://doi.org/10.1016/j.artmed.2015.03.006

Enrico, C., Wendy, R., Piyush, M., Sandeep, R., Ville-Petteri, M., Pieta, B., Markus, W., Eva, W., Saba, A., Aaron, C., & Blair, K. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health & Care Informatics, 28*(1), Article e100444. https://doi.org/10.1136/bmjhci-2021-100444

Fahd, K., Parvin, S., & Souza-Daw D. A. (2021, November 24-26). *A framework for Real-time sentiment analysis of big data generated by social media platform* [Conference session]. The 31st International Telecommunication Networks and Applications Conference (ITNAC), Sydney, Australia. https://doi.org/10.1109/ITNAC53136.2021.9652148

Fujino, A., Isozaki, H., & Suzuki, J. (2008, January 7-12). *Multi-label text categorization with model combination based on F1-score maximization* [Conference session]. Third International Joint Conference on Natural Language Processing: Volume-II, Hyderabad, India. https://aclanthology.org/I08-2116.pdf

Gill, K. S., Anand, V., & Gupta, R. (2023, April 27-29). *Foetal health classification using input dataset and fine tuning it using k-nearest neighbour, naïve bayes and decision tree classifier* [Conference session]. 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI), Tiruchengode, India, https://doi.org/10.1109/ICAEECI58247.2023.10370853

Goeuriot, L., Na, J. C., Kyaing, W. Y. M., Khoo, C., Chang, Y. K., Theng, Y. L., & Kim, J. J. (2012, January 28-30). *Sentiment lexicons for health-related opinion mining* [Conference session]. 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12), NY, United States. https://doi.org/10.1145/2110363.2110390

Gow, J., George, G., Mutinta, G., Mwamba, S., & Ingombe, L. (2011). Health worker shortages in Zambia: An assessment of government responses. *Journal of Public Health Policy, 32*(4), 476-488. https://doi.org/10.1057/jphp.2011.41

Hashemi, B., Motamedi, M., Etemad, M., Rahmati, F., Forouzanfar, M. M., & Kaghazchi, F. (2014). An audit of emergency department accreditation based on joint commission international standards (JCI). *Emergency, 2*(3), 130-133. https://doi.org/10.22037/aaem.v2i3.120

Hayurian, L. A., & Hendrastuty, N. (2024). Comparison of Naïve Bayes algorithm and support vector machine in sentiment analysis of Boycott Israeli products on Twitter. *Jurnal Teknik Informatika (Jutif), 5*(3), 731-738. https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1813

He, J., Cang, Y., Wei, C., Wang, H., Xu, Y., & Han, Y. (2024). Application of joint commission international standard combined with 6S management mode in the management of emergency medical equipment. *Hong Kong Journal of Emergency Medicine, 31*(1), 22-27. https://doi.org/10.1002/hkj2.12009

HongFan, Z., Siou-Tang, H., Mark, J. B., Lilly, E., & Herng-Chia, C. (2023). Perceptions of Chinese hospital leaders on joint commission international accreditation: a qualitative study. *Frontiers in Public Health, 11*, Article 1258600. https://doi.org/10.3389/fpubh.2023.1258600

Hwang, U., Richardson, L., Livote, E., Harris, B., Spencer, N., & Sean Morrison, R. (2008). Emergency department crowding and decreased quality of pain care. *Academic Emergency Medicine, 15*(12), 1248-1255. https://doi.org/10.1111/j.1553-2712.2008.00267.x

Khaleghparast, S., Maleki, M., Hajianfar, G., Soumari, E., Oveisi, M., Golandouz, H. M., Noohi, F., dehaki, M. G., Golpira, R., Mazloomzadeh, S., Arabian, M., & Kalayinia, S. (2023). Development of a patients' satisfaction analysis system using machine learning and lexicon-based methods. *BMC Health Services Research, 23*, Article 280. https://doi.org/10.1186/s12913-023-09260-7

Khanam, R., & Sharma, A. (2021, July 2-4). *Sentiment analysis using different machine learning techniques for product review* [Conference session]. International Conference on Computational Performance Evaluation (ComPE), Shillong, India. https://doi.org/10.1109/ComPE53109.2021.9752004

Kitsirichawanan, T., & Mongkolsawat, D. (2019). Operating room renovation project management: a case study of a joint commission international (JCI) accredited private hospital. *Architectural technology and innovation, 34*(2), C1-C19. https://so04.tci-thaijo.org/index.php/NAJUA-Arch/article/view/228498

Liang, P. W., & Dai, B. R. (2013, June 1-6). *Opinion mining on social media data* [Conference session]. IEEE 14th International Conference on Mobile Data Management, Milan, Italy. https://doi.org/10.1109/MDM.2013.73

Liao, Y.-Y., Soltani, E., & Wilkinson, A. (2023). From product to service quality: The role of managerial mindsets. *Production Planning & Control, 34*(8), 705-726. https://doi.org/10.1080/09537287.2021.1957173

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Menaouer, B., Zahra, A. F., & Mohammed, S. (2022). Multi-class sentiment classification for healthcare tweets using supervised learning techniques. *International Journal of Service Science, Management, Engineering, and Technology, 13*(1), 1-17. https://doi.org/10.4018/IJSSMET.298669

Mir Habeebullah Shah, Q., & Dr, R. K. S. (2020). Performance of naïve bayes in sentiment analysis of user reviews online. *International Journal of Innovative Technology and Exploring Engineering, 10*(2), 64-68. https://doi.org/10.35940/ijitee.A8198.1210220

Mogakwe, L. J., Ally, H., & Magobe, N. B. D. (2020). Reasons for non-compliance with quality standards at primary healthcare clinics in Ekurhuleni, South Africa. *African Journal of Primary Health Care & Family Medicine, 12*(1), 1-9. https://doi.org/10.4102/phcfm.v12i1.2179

Mogakwe, L. J., Magobe, N. B. D., & Ally, H. (2020). Facilitating compliance with quality standards at primary health care clinics through adequate health care resources. *Africa Journal of Nursing & Midwifery, 22*(1), 1-17. https://doi.org/10.25159/2520-5293/6569

Munyewende, P. O., Rispel, L. C., & Chirwa, T. (2014). Positive practice environments influence job satisfaction of primary health care clinic nursing managers in two South African provinces. *Human Resources for Health, 12*, Article 27. https://doi.org/10.1186/1478-4491-12-27

Navale, S., Ambale, R., Kapse, V., & Ghadge, S. (2022). To recommend the best hospital in an area using machine learning: Medic aid analysis. *International Journal of Research in Engineering, Science and Management, 5*(4), 156-158. https://journal.ijresm.com/index.php/ijresm/article/view/1978

Nurfaizah, Hariguna, T., & Romadon, Y. I. (2019). The accuracy comparison of vector support machine and decision tree methods in sentiment analysis. *Journal of Physics: Conference Series, 1367*, Article 012025. https://doi.org/10.1088/1742-6596/1367/1/012025

Pandesenda, A. I., Yana, R. R., Sukma, E. A., Yahya, A., Widharto, P., & Hidayanto, A. N. (2020, November 19-20). *Sentiment analysis of service quality of online healthcare platform using fast large-margin* [Conference Session]. International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia. https://doi.org/10.1109/ICIMCIS51567.2020.9354295

Paruchuri, H., Vadlamudi, S., Ahmed, A. A. A., Eid, W., & Donepudi, P. K. (2021). Product reviews sentiment analysis using machine learning: A systematic literature review. *Turkish Journal of Physiotherapy Rehabilitation, 32*(2), 2361-2367. https://doi.org/10.5281/zenodo.5534344

Pradhan, B., & Panda, P. S. (2023, February 17-18). *Long short term memory-based sentiment analysis of healthcare reviews* [Conference Session]. International Conference on Ambient Intelligence in Health Care (ICAIHC), Bhubaneswar, India. https://doi.org/10.1109/ICAIHC59020.2023.10431432

Prakash, C. O, Kumar, S. V., Pallavi, Abbas, J. K., Naser, S. J., & Haroon, N. H. (2024, April 4-5). *Enhancing natural language processing with deep learning for sentiment analysis* [Conference Session]. Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India. https://doi.org/10.1109/ICONSTEM60960.2024.10568829

Rahim, A. I. A., Ibrahim, M. I., Chua, S. L., & Musa, K. I. (2021). Hospital Facebook reviews analysis using a machine learning sentiment analyzer and quality classifier. *Healthcare, 9*(12), Article 1679. https://doi.org/10.3390/healthcare9121679

Ramjan, S., & Sunkpho, J. (2023). Principles and theories of data mining with RapidMiner. IGI Global. https://doi.org/10.4018/978-1-6684-4730-7

Sharma, S., Pandey, A., Kumar, V., Ohdar, D., Pillai, A. R., & Mahajan, M. (2023, March 24-25). *Recent trends in sentiment analysis using different machine learning based models: a short review* [Conference session]. 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India. https://doi.org/10.1109/ICAAIC56838.2023.10140954

Shimada, S. L., Petrakis, B. A., Rothendler, J. A., Zhao, S., Feng, H., Fix, G. M., Simon, S. R., Zirkle, M., Ozkaynak, M., Martin, T., Johnson, S. A., Tulu, B., Gordon, H. S., & Woods, S. S. (2017). An analysis of patient-provider secure messaging at two veterans health administration medical centers: message content and resolution through secure messaging. *Journal of the American Medical Informatics Association, 24*(5), 942-949-949. https://doi.org/10.1093/jamia/ocx021

Streiner, D. L., & Norman, G. R. (2006). "Precision" and "accuracy": Two terms that are neither. *Journal of Clinical Epidemiology, 59*(4), 327-330. https://doi.org/10.1016/j.jclinepi.2005.09.005

Tarieh, R. R. A., Zayyat, R., Naoufal, R. N., & Samaha, H. R. (2022). A case study exploring the impact of JCI standards implementation on staff productivity and motivation at the laboratory and blood bank. *Health Science Reports, 5*, Article e497. https://doi.org/10.1002/hsr2.497

Usman, M., Mujahid, M., Rustam, F., Soriano Flores, E., Vidal Mazón, J. L., de la Torre Díez, I., & Ashraf, I. (2024). Analyzing patients satisfaction level for medical services using twitter data. *PeerJ Computer Science, 10,* Article e1697. https://doi.org/10.7717/peerj-cs.1697

Xu, Y., Cao, H., Du, W., & Wang, W. (2022). A survey of cross-lingual sentiment analysis: Methodologies, models and evaluations. *Data Science and Engineering, 7*(3), 279–299. https://doi.org/10.1007/s41019-022-00187-3

Yacouby, R., & Axman, D. (2020, November 20). *Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models* [Conference Session]. First Workshop on Evaluation and Comparison of NLP Systems. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.eval4nlp-1.9

Yarimoglu, E., & Ataman, G. (2022). How service quality in hospitals varies based on hospital ownership and demographics: A study on Turkish patients living urban areas. *Total Quality Management & Business Excellence, 33*(7-8), 777-793. https://doi.org/10.1080/14783363.2021.1890576