ENHANCING HOTEL MANAGEMENT: A SENTIMENT ANALYSIS APPROACH TO ASSESSING CUSTOMER IMPRESSIONS ON ENVIRONMENT-BASED REVIEWS

Korakot Matarat[A]

ARTICLE INFO	ABSTRACT
	Purpose: This study aims to examine hotel reviews, with a specific emphasis on environmental aspects. Employing advanced sentiment analysis techniques, we delve into user sentiments expressed in evaluations. **Theoretical Framework:** The theoretical framework utilizes sentiment analysis to delve into hotel reviews, emphasizing the environment. Focused on key areas such as staff service (professionalism and friendliness), hotel environment and facilities (beautiful décor and cleanliness), and affordability with a friendly pricing strategy. **Design/Methodology/Approach:** This study involves the manual collection and analysis of 3,475 hotel review from Centara Hotel & Convention Center Udon Thani. The methodology includes systematic steps: Preprocessing (removing irrelevant charcters and stopword), Feature Extraction (identifying key elements), Classification (using Logistic Regression for binary sentiment analysis), and Prediction (applying the model to categorize new reviews). **Findings:** In our findings, Logistic Regression effectively categorized reviews into positive or negative sentiments, boasting a robust macro precision of 0.80. Notably, positive evaluations showed superior prediction results, with a high recall value of 0.84 percent, contributing to an impressive overall accuracy of 86 percent. These results highlight the efficacy of Logistic Regression in distinguishing sentiment categories, affirming its suitability for this analysis. **Research, Practical & Social Implications:** In summary, focusing on enhancing staff service, maintaining a pristine hotel environment, and offering friendly, affordable services has the potential to greatly boost customer satisfaction and overall business performance. These strategic efforts not only generate positive feedback but also lead to increased competitive advantages and expanded market share in the fiercely competitive hotel industry. **Originality/Value:** The findings have broader implications by shaping public perceptions, providing valuable insights for strategic decision-making, and positioning hotels for success in the highly competitive hospitality landscape. Doi: https://doi.org/10.26668/businessreview/2023.v9i1.4152

MELHORAR A GESTÃO HOTELEIRA: UMA ABORDAGEM DE ANÁLISE DE SENTIMENTO PARA AVALIAR AS IMPRESSÕES DOS CLIENTES SOBRE AS AVALIAÇÕES BASEADAS NO AMBIENTE

RESUMO
Objetivo: Este estudo tem por objetivo examinar as avaliações dos hotéis, com especial ênfase nos aspectos ambientais. Empregando técnicas avançadas de análise de sentimentos, aprofundamos os sentimentos dos usuários expressos em avaliações.

[A] *Master in Computer Science, Teacher. Faculty of Science and Technology, Sakon Nakhon Rajabhat University. Sakon Nakhon, Thailand. E-mail: korakot@snru.ac.th Orcid: https://orcid.org/0000-0001-9724-202X*

Intern. Journal of Profess. Bus. Review. |Miami, v. 9 | n. 1| p. 01-20 | e04152 | 2024.

1

Estrutura Teórica: A estrutura teórica utiliza a análise de sentimentos para aprofundar-se nas revisões do hotel, enfatizando o ambiente. Focado em áreas-chave, como serviço de pessoal (profissionalismo e simpatia), ambiente e instalações do hotel (bela decoração e limpeza), e acessibilidade com uma estratégia de preços amigável.

Design/Metodologia/Abordagem: Este estudo envolve a coleta e análise manuais de 3.475 avaliações de hotéis do Centara Hotel & Convention Center Udon Thani. A metodologia inclui etapas sistemáticas: Pré-processamento (remoção de caracteres irrelevantes e palavras irrelevantes), Extração de Recursos (identificação de elementos-chave), Classificação (uso de Regressão Logística para análise de sentimento binário) e Previsão (aplicação do modelo para categorizar novas revisões).

Constatações: Em nossas descobertas, a Regressão Logística efetivamente categorizou revisões em sentimentos positivos ou negativos, ostentando uma robusta precisão macro de 0,80. Notavelmente, avaliações positivas mostraram resultados de previsão superiores, com um alto valor de recordação de 0,84%, contribuindo para uma impressionante precisão geral de 86%. Esses resultados destacam a eficácia da Regressão Logística em distinguir categorias de sentimentos, afirmando sua adequação para esta análise.

Pesquisa, Implicações Práticas e Sociais: em resumo, o foco no aprimoramento do serviço de equipe, na manutenção de um ambiente hoteleiro intocado e na oferta de serviços amigáveis e acessíveis tem o potencial de aumentar significativamente a satisfação do cliente e o desempenho geral dos negócios. Estes esforços estratégicos não só geram uma reação positiva, como também conduzem a um aumento das vantagens competitivas e a uma maior quota de mercado na feroz concorrência do setor hoteleiro.

Originalidade/Valor: as descobertas têm implicações mais amplas ao moldar as percepções do público, fornecendo informações valiosas para a tomada de decisões estratégicas e posicionando hotéis para o sucesso no cenário de hospitalidade altamente competitivo.

Palavras-chave: Análise de Sentimento, Extração de Texto, Regressão Logística, Vantagens Competitivas, Intimidade do Cliente, Percepções Públicas.

MEJORA DE LA GESTIÓN HOTELERA: UN ENFOQUE SENTIMENTE DE ANÁLISIS PARA EVALUAR LAS IMPRESIONES DE LOS CLIENTES SOBRE LAS EVALUACIONES MEDIANTE EL MEDIO AMBIENTE

RESUMEN
Objetivo: El objetivo de este estudio es examinar las evaluaciones hoteleras, con especial énfasis en los aspectos medioambientales. Mediante el uso de técnicas avanzadas de análisis de opinión, profundizamos los sentimientos de los usuarios expresados en las evaluaciones.

Estructura Teórica: La estructura teórica utiliza el análisis de los sentimientos para profundizar en las revisiones del hotel, haciendo hincapié en el medio ambiente. Centrado en áreas clave, como el servicio de personal (profesionalismo y simpatía), el entorno y las instalaciones de los hoteles (decoración y limpieza de calidad), y la accesibilidad con una estrategia de precios amigable.

Diseño/Metodología/Enfoque: Este estudio incluye la recolección manual y el análisis de 3.475 evaluaciones hoteleras del Centro Hotel & Convention Center Udon Thani. La metodología incluye pasos sistemáticos: Pre-procesamiento (eliminación de caracteres irrelevantes y palabras irrelevantes), Extracción de recursos (identificación de elementos clave), Clasificación (uso de regresión logística para análisis de sentimientos binarios) y Previsión (aplicación de plantilla para categorizar nuevas revisiones).

Hallazgos: En nuestros descubrimientos, la Regresión Logística categorizó efectivamente las revisiones en sentimientos positivos o negativos, con una robusta precisión macro de 0.80. En particular, las evaluaciones positivas han mostrado mayores resultados de predicción, con un valor sin precedentes del 0,84 por ciento, lo que ha contribuido a una impresionante precisión general del 86 por ciento. Estos resultados resaltan la efectividad de la Regresión Logística al distinguir entre categorías de sentimientos, afirmando que son adecuados para este análisis.

Investigación, Implicaciones Prácticas y Sociales: En resumen, el enfoque en mejorar el trabajo en equipo, mantener un entorno hotelero intacto y ofrecer servicios amigables y asequibles puede aumentar significativamente la satisfacción del cliente y el rendimiento general de las empresas. Estos esfuerzos estratégicos no sólo generan una reacción positiva, sino que también conducen a un aumento de las ventajas competitivas y a una mayor cuota de mercado en la feroz competencia del sector hotelero.

Originalidad/Valor: Los descubrimientos tienen implicaciones más amplias al dar forma a las percepciones públicas, proporcionar información valiosa para la toma de decisiones estratégicas y posicionar a los hoteles para el éxito en el escenario de hospitalidad altamente competitivo.

Palabras clave: Análisis de Sentimientos, Extracción de Texto, Regresión Logística, Ventajas Competitivas, Intimidad del Cliente, Percepciones Públicas.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

2

INTRODUCTION

In an era dominated by the widespread use of the internet, hotel reservations have undergone a transformative shift towards online platforms. Unlike traditional methods involving travel agencies, users now have the responsibility of personally selecting and reserving hotel rooms. A pivotal aspect influencing user decisions in this digital landscape is the wealth of information available through hotel reviews. These reviews typically include a numerical rating as part of a user profile and an impression that encapsulates the guest's perspective on the hotel. The language used in reviews plays a crucial role in conveying sentiments. Positive reviews may feature words such as "great," "superb," "fab," and "good," while negative sentiments are often expressed through terms like "awful," "poor," "terrible," and "worse" (Samruddhi, et al., 2019). Hotel seekers increasingly rely on such user-generated content to inform their decisions, making reviews an integral component of the online reservation process. Consequently, both users and lodging service providers heavily depend on these reviews, with hotel administrators utilizing them as a tool to enhance services.

This paper delves into the realm of sentiment analysis, also known as opinion mining, to address the evolving dynamics of customer-driven service improvements. Sentiment analysis involves the detection, extraction, and classification of customer reviews, categorized into two levels. The first level focuses on document classification, discerning reviews as positive or negative, while the second level analyzes sentiments within individual review sentences. To formally introduce feature selection and extraction, our study presents a comprehensive analysis of hotel reviews, specifically employing Logistic Regression as a classification model to extract characteristic information. The process involves creating a dataset of Thai-language reviews, developing a classifier model, and subsequently analyzing the outcomes. The structure of the paper unfolds as follows: Section 2 delves into related work; Section 3 outlines the methods employed; Section 4 details experiments and their outcomes, and Section 5 provides conclusions while discussing avenues for future work.

LITERATURE REVIEW

Studying the theories and related research, it was found that many articles studied and analyzed confidence in hotel reviews. Applying machine learning methods such as the Naive Bayes, Maximum Entropy, and SVM yielded about 80% accuracy in this article. Ensemble- and hybrid-based Twitter sentiment analysis algorithms tended to outperform supervised machine learning techniques, since they were able to attain roughly 85% classification accuracy (Alsaeedi,

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

3

et al., 2019). Using an Enhanced Feature Acquisition Method, this research presents multilingual sentiment analysis of English and Hausa tweets (EFAM). The approach's success has been boosted by feature integration for multilingual sentiment analysis with an average precision of over 65 percent (Abubakar, et al., 2021). The sentiment analysis approach presented in this article uses three algorithms: Nave Bayes, Support vector machines, and decision trees. The experiment found that the decision tree is the best algorithm. While evaluations were carried out to determine the model's accuracy. The model's best accuracy level is 88.54 % (Apriliani, et al., 2020). This paper presents an approach to analyzing user reviews written in Thai based on techniques in natural language processing, topic modeling, and sentiment analysis on mobile. The result can provide users and developers with some insights into the analyzed mobile applications (Deewattananon and Sammapun, 2017). Using an aspect-based sentiment analysis technique, this project employs machine learning classification models such as Nave Bayes, Support Vector Machine, Logical Regression, Random Forest, and Decision Tree to classify user evaluations as positive or negative. With an accuracy of 84.55 percent, Nave Bayes outperformed all other algorithms (Dcunha, 2019). This study applies the multinomial Nave Bayes Classifier approach to classify positive or negative opinion reviews and compares models using preprocessing. The average F1-Score for the best experimental results employing preprocessing and feature selection with 10-fold cross-validation was greater than 91% (Farisi, et al., 2018). This paper examines the theory and logic behind various common feature selection and extraction approaches, as well as some of their applications. These methods are also presented with a numerical implementation. Finally, the approaches were put to the test on the MNIST dataset to see how well they performed. The accuracy of the original data without applying the extraction method was 53.50% (Ghojogh, et al., 2019). The goal of this study is to design an effective technique for Thai word segmentation, which the researchers have dubbed PTTSF (Parsing Thai Text with Syntax and Feature of Word). According to the findings, the derived mapping technique could address the problem of segmenting words that did not exist in the dictionary with an average accuracy of above 90% (Mahatthanachai, et al., 2016). In this study, a logistic regression technique for Arabic is proposed, along with term and inverse document frequency (TF*IDF). Experiments have revealed three key findings: 1) When used to predict good evaluations, the classifier is accurate. 2) The model is skewed when it comes to predicting unfavorable reviews. Finally, the low percentage of negative reviews in the corpus adds to the logistic regression model's ambiguity (Omari, et al., 2019). In this paper, he investigates Nave Bayes, K-Nearest Neighbors, Random Forest, Maximum Entropy, SVM, and Voted Perceptions, as well as other conventional

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

4

techniques used for the core Sentiment Analysis task of analyzing movie reviews. The processing of natural languages has enhanced Logistic Regression's performance (Panda, et al., 2020). In this research, proposes techniques for sentiment analysis of hotel reviews in the implementation, he trained six classifiers with an accuracy of 88 to 94 %. The top five most accurate classifiers were incorporated in the voted classifier, which had an accuracy of 93.57 % one in five is Logistic Regression (Patel, et al., 2020). This paper relies on use three approaches to classify a text based on whether a person's overall mentality is negative, positive, or neutral, and compare the results. Classification techniques 1) SVM, 2) Naïve Bayes (NB) 3) Logistic Regression using Hybrid Algorithm, which result provides greater precision (Pathuri, et al., 2019). This study provides a system that collects customer reviews and categorizes them according to the given remarks, making it easier to access information. The model is based on the Nave-Bayes algorithm, which classifies the reviews with an accuracy of 89iven remarks, making it easier to access information. The model is based on the Nave-Bayes algorithm, which classifies the reviews with an accuracy of 89% (Samruddhi, et al., 2019). This paper presents the idea of classifying the sentences into multiple categories, like happy, sad, hungry, in love, etc. This research uses Nave Bayes, which can give the highest accuracy when using machine learning algorithms or natural language processing algorithms (Sharad, et al., 2018). The focus of this study is on defining the nature of frequency as positive or negative. The proposed method is based on weighing techniques, with term frequency and inverse document frequency providing superior performance. The suggested classifier is based on the frequency and distribution of phrase occurrences. (Shinde and Deshmukh, 2016). This study provides a framework for categorizing a hotel review's opinion feature into three elements (location, service, and worthiness). In the hotel reviews, classify the opinion polarity into two classes using a hybrid concept of keyword knowledge and a classification technique (positive and negative). The accuracy of opinion feature recognition is 83.33 %, while the accuracy of opinion polarity classification is 81.47 % (Sungsri and Ua-apisitwong, 2017). This study compares a sentiment lexicon created from documents in a specific domain with a generic lexicon. The experimental results indicate that the specific-lexicon classifier performs better than the generic-lexicon classifier in terms of accuracy and f-measure. At the threshold value of 0.25, the f-measure for the negative sentiment class (respectively, the positive sentiment class) is 36.1% (respectively, 13.5%) higher (Tanantong, et al., 2020). This study presented based on typical expressions of impression remarks collected in online hotel reviews, and offered a feature extraction approach for numerical assessment criteria that are difficult to interpret just from numerical evaluation scores. By examining impression comments

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

5

from reviews in which numerical assessment items are evaluated equally (Tsujii, et al., 2015). The goal of this study is to develop a system for predicting customer opinion using supervised machine learning and the decision tree. The method for categorizing online hotel reviews as positive or negative using supervised machine learning and the decision tree approach produced results that show using the model produced from the balanced training set and filtering rare and frequent words local, room, breakfast, staff, and hotel produced the most accurate prediction (Yordanova and Kabakchieva, 2017).

However, in a reliable and effective analysis of both positive and negative reviews of a product or service, it is important to separate feature the review. Most of the research and tools currently focused on English text, we have sought to find reliable and efficient methods and algorithms to classification the characteristics of reviews that are positive or negative in this research based on review data hotel in the Thai language. Including the sorting of words that users say positively and negatively for related parties to consider improving the service. To improve hotel services, we'll increase the efficiency with which we gather evaluation data and look into new ways of presenting data.

DATA AND METHODOLOGY

This study focuses on hotel reviews from Centara Hotel & Convention Center UdonThani, which were manually collected for analysis. The selected reviews were then processed to extract relevant information, and the resulting dataset was saved in CSV format. The dataset comprises 3,475 text reviews, each accompanied by corresponding score ratings reflecting the overall service experience.

The methodological approach employed in this research involves a systematic process outlined in a Flow Chart (Figure 1) to effectively conduct sentiment analysis on the hotel reviews. The key phases of the methodology include Preprocessing, Feature Extraction, Classification, and Prediction using Logistic Regression.

Data Collection

Reviews from Centara Hotel & Convention Center UdonThani were manually collected.

The collected reviews, along with their score ratings, were compiled into a dataset for analysis.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

6

Preprocessing

Textual data undergoes preprocessing to enhance the quality of information for analysis.

Steps include removing irrelevant characters, stopwords, and performing tokenization.

Feature Extraction

Feature extraction involves identifying and selecting key elements from the preprocessed data.

Extracted features play a crucial role in training the sentiment analysis model.

Classification

The processed data is utilized to train a classification model using Logistic Regression.

Logistic Regression is chosen for its effectiveness in binary classification tasks, distinguishing between positive and negative sentiments.
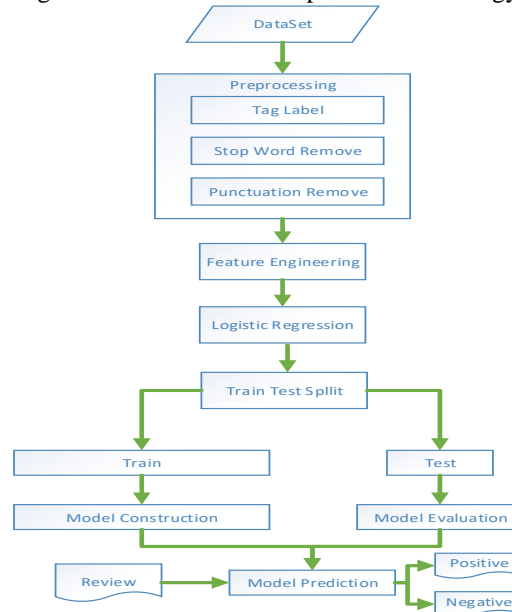
Prediction

The trained model is applied to predict sentiment in new, unseen hotel reviews.

The model categorizes reviews into positive or negative sentiments based on the features extracted during the training phase.

The proposed methodology, depicted in Figure 1, provides a structured and comprehensive framework for conducting sentiment analysis on hotel reviews. By employing these systematic steps, the study aims to uncover valuable insights into customer sentiments and preferences at Centara Hotel & Convention Center UdonThani.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

7

Figure 1: Flow chart of Proposed Methodology

Source: Prepared by authors elaboration using the https://app.diagrams.net

The classification model employed in this study aimed to discern the sentiments expressed by hotel guests in the provided dataset, categorizing them as either positive or negative. The determination of positive and negative polarities relied on the scores assigned to the reviews by their respective authors. This approach operates under the assumption that review ratings inherently capture the sentiments conveyed in the review content, with lower ratings indicating negative sentiments and higher ratings reflecting positive sentiments. To standardize the scoring system, a full score of 10 was assigned to each domain, ensuring consistency across the dataset. The sentiment identification process involved setting a criterion for the compound score, derived from the review ratings. If a review's compound score was greater than or equal to 6, it was classified as positive. Conversely, if the compound score was less than or equal to 5, the review was categorized as negative. This binary sentiment classification approach provides a straightforward distinction between positive and negative opinions expressed in the reviews. The rationale behind this methodology is rooted in the belief that the compound score, derived from the overall review rating, encapsulates both positive and negative aspects articulated by the reviewers. The outcomes of this sentiment classification process are illustrated in Figure 2, showcasing the distribution of positive and negative reviews based on the established compound score criteria. This analysis not only aids in understanding the prevailing sentiments within hotel reviews but also serves as a valuable tool for enhancing customer satisfaction and refining service quality in the hotel industry.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

8

Figure 2: The amount of positive and negative review
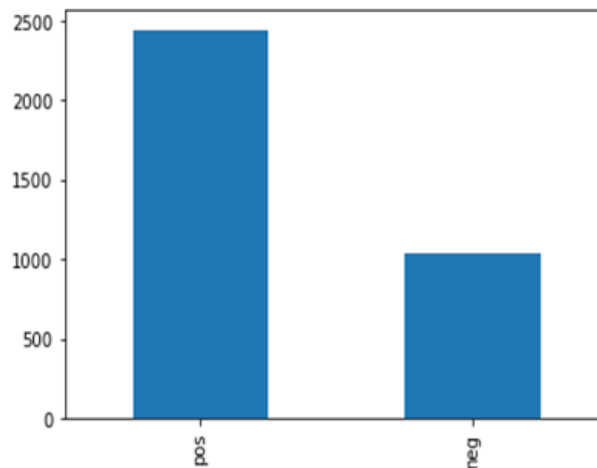
Source: Prepared by authors elaboration using the matplotlib.
Python Package

Pre-Processing

One of the most significant tasks in sentiment analysis is pre-processing. To prepare the data for categorization, it will clean the dataset by lowering its complexity. Firstly, datasets are tagged as positive or negative based on scores, if scores are less than or equal to 5 datasets are tagged as negative more than positive. Subsequently, it will remove words that are meaningless for processing the text, for example มี, เฉยๆ, เช่นใด, เพียงแต่, น้อยๆ, ข้างเคียง .The dataset was then tokenized to separate the words into tokens, which were subsequently stemmed to reduce the tokens to a single type. The stemming technique minimizes the number of superfluous words in a document. There are three tiers of cleaning strategies in pre-processing: tier 1, tier 2, and tier 3. By implementing these pre-processing steps, the dataset is refined, and its complexity is lowered, ensuring a more focused and meaningful basis for sentiment analysis. This meticulous cleaning process sets the stage for accurate categorization and meaningful insights into the sentiments expressed in hotel reviews.

Tier 1: Tag a series positive or negative. Datasets are tagged as positive or negative based on score. If the score is less than or equal to 5, the dataset is tagged as negative. More than 5 is positive as in example table 1. The outcome of the review after preprocessing is as follows:

Table 1 - Hotel review with tag as positive or negative

Score	Review	Tag
2	ห้องแย่ไม่คุ้มกับเงินที่จ่าย	neg
3	อาหารอร่อยแต่แอร์ไม่เย็นเลย	neg
4	พนักงานต้อนรับพูดจาและหน้าไม่ค่อยต้อนรับ	neg
5	บริการดีแต่เจอแมลงสาบในห้องพัก	neg

6	อยู่ใจกลางเมืองใกล้แหล่งช้อปปิ้งห้องพักค่อนข้างดี	pos
7	ที่พักสะอาดพอใช้ได้เข้าซอยนิดแต่อยู่ใกล้แหล่งช้อปปิ้ง	pos
8	โรงแรมสะอาดเดินทางสะดวกแต่เด็กเล็กก็มีคิดค่าอาหารเช้าด้วยโรงแรมน่าพัก	pos
9	โรงแรมใหม่สะอาดอาหารเช้าอร่อยพนักงานบริการดีค่ะ	pos

Source: Prepared by authors elaboration using the https://www.agoda.com/th-th/centara-hotel-convention-centre-udon-thani/hotel/udon-thani-th.html

Tier 2: Remove Stop Words: At this level, delete articles such as มี, ข้างเคียง, and เพียงแต่ they don't have any roles. Meaningless words refer to the types of words that do not affect the analysis. But it creates confusion between converting a text file. The experiment found that the data contained more than a hundred meaningless words such as special characters (A,", "", (), ?, !, @, #), date (13/10/21), and no words. Meaningful (pos+, pos-) at this stage, additional cleaning strategies are applied. In addition to deleting stop and words, numbers and meaningless words were also removed such as sentence numbers. Remove stopwords Thai shown in Table 2.

Table 2 **–** Thai Stopwords

Thai Stopwords
[มี,พอสม,ซึ่งก็,ฝ่าย,เฉยๆ,ทีๆ,เช่นใด,ทุกชิ้น,จำ,จัดหา,ทุกแห่ง,ถึงเมื่อ,เพียงแต่,ตลอดกาล,ตลอดทั่วทั้ง,น้อยๆ,ข้างเคียง,เป็นเพราะว่า,เขียน, ฯลฯ,อันที่จริง,ยังคง,ไง,มัน,ร่วมด้วย,จวนจะ,ด้วยเหตุว่า,เปลี่ยน,ทันทีทันใด,นิดๆ,ดั่งกับ,นับจากนี้, แบบ,เชื่อว่า,ส่วนใหญ่,ครบครัน,ข้าง,ทรง,ผ่าน,เกือบ,ที่ว่า,ทุกครั้ง,กำลังจะ,พบว่า,เป็นต้น,เมื่อเช้า,อาจเป็นด้วย,แล้วเสร็จ,เท่า,ยืนนาน,อันที่ จะ,ครั้งหลังสุด,ส่วนดี,พยายาม,กว่า,อัน,ส่วนน้อย,อย่างที่,ซึ่ง,ถ้าหาก,เล็ก,พอตัว,ล่าสุด,สู่,มั้ยเนี่ย,พวกแก,ช่วงแรก,คราที่,ทั้งสิ้น,ถูกๆ,ครั้งใ ด,เมื่อใด,ช่วงถัดไป,อันได้แก่,เสียแล้ว,ตาม,ยังโง้น,เยอะ,กลับ,ในเมื่อ,เท่าที่,หรือยัง,เป็นเพียง,ยังจะ,เหตุนี้,ทุกวันนี้,ทั้ง,รับรอง,เห็นควร,ภาย ใต้,ก่อน,ช่วงระหว่าง,นี่แน่ะ,น่ะ,ประการ,บอก,ยิ่งนัก,ไม่ค่อยเป็น,ถึงเมื่อไร,ด้วยกัน,ตามแต่,ยกให้,เท่ากับ,คล้ายกับว่า,......,ทุกอย่าง ,]

Source: Prepared by authors elaboration using the https://pythainlp.github.io

Tire 3: The dataset will then be cleaned by lowering its complexity in order to prepare it for classification. To begin, the dataset was tokenized to break down the words into tokens, and then stemming was used to condense the tokens into a single type, such as the word "มีสวนสาธารณะใกล้กับโรงแรม" will reduce to "มี / สวนสาธารณะ / ใกล้ /กับ / โรงแรม". In pre-processing, three-tier cleaning complete strategies were introduced show tokens in Figure 3.

Matarat, K. (2024)

Enhancing Hotel Management: a Sentiment Analysis Approach to Assessing Customer Impressions on Environment-Based Reviews

Figure 3: Split Word Tokens

3475 rows × 3 columns

Source: Prepared by authors elaboration using the word tokenize
Python Package

Feature Extraction Techniques

Term weighting was used as a feature extraction strategy in this investigation. The frequency with term is the number of occurrences in a text, as well as in all papers in the corpus. The TF-IDF is a calculation that determines the term's importance in a document and corpus, allowing the relevance of a document to all other papers in the corpus to be determined. The extraction and selection of attributes aim to select a subset of words that occur only in practice and using sets. Selecting an attribute using the information obtained as a criterion to assess the importance of an attribute. Engineering Features the Term Frequency-Inverse Document Frequency (TF-IDF) method can be used to choose attributes based on frequency in both training and testing datasets. The dataset and weight of the infrequent condition. TF-IDF is calculated using the following formula:

$$TF - IDF = TF_{t,d} * IDF_t \qquad (1)$$

Where t is a term (attribute) in a document (example) and d is the document (example) containing t.

Term Frequency (TF) is the ratio of the number of occurrences of a term in a document to the total number of terms in the document (n).

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

11

$$TF - IDF = \frac{n_t}{n} * \log_2 \frac{N_d}{N_t} \tag{2}$$

The Inverse Document Frequency (IDF) is the ratio of the total number of documents in the corpus (Nd) to the number of documents that contain the term t (Nt) (Yordanova and Kabakchieva, 2017).

Table 3 - Feature Selection from Frequent Words

Positive words	Frequency	Negative words	Frequency
โรงแรม	4103	แย่มาก	3966
เดินทาง	3411	เหม็น	3783
อาหารเช้า	3171	เสียงดัง	3736
ห้องพัก	3055	เก่า	3273
สะอาด	2795	สกปรก	2687
สะดวก	2790	ราคาแพง	2421
ยอดเยี่ยม	2244	น้ำซึม	1695

Source: Prepared by authors elaboration using the Count Vectorizer
Python Package

In this stage, a structured data frame is generated, cataloging positive and negative words, each associated with its respective sentiment and score. The objective of employing a Word Cloud is to visually extract and represent words that hold significance within the dataset. However, it is often impractical to display every single word, given that certain terms are frequently reiterated in reviews. The Word Cloud presented in Figure 4 serves as an illustrative snapshot, showcasing the most prevalent words in the positive assessment of datasets. This visual representation is designed to encapsulate the essence of the text data, emphasizing words that frequently emerge in favorable reviews. The Word Cloud not only offers an insightful overview of commonly used terms but also aids in discerning patterns and trends within the positive sentiments expressed by reviewers. This Word Cloud analysis contributes to a holistic comprehension of the sentiments conveyed in the dataset, offering a visually compelling portrayal of the most impactful words associated with positive assessments.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

12

Figure 4: Word Cloud Representation of a positive review

Source: Prepared by authors elaboration using the wordcloud Python Package

Logistic Regression

In this subsequent stage, the focus shifts towards calculating the positive and negative polarity of individual sentences. This polarity information serves as a pivotal metric, forming the basis for determining an individual's opinion regarding a particular aspect of the dataset. To effectively process and represent the text or documents, machine learning techniques such as Word Vectorizer or Bag-of-Words (BoW) are employed. These techniques are instrumental in transforming textual data into essential feature vectors, which are then utilized for text classification. Feature engineering involves calculating polarities, utilizing machine learning techniques for representation, and employing Bag-of-Words for feature extraction.

The Logistic Regression model is trained and tested to create a reliable sentiment classification model, contributing to a comprehensive sentiment analysis of hotel reviews.This meticulous feature engineering process and subsequent model training pave the way for a more nuanced understanding of sentiments expressed in hotel reviews, facilitating improved decision-making for hotel management, as shown in equation (3).

$$p(y = \pm 1 | x, w) = \frac{1}{1 + e^{-w^T h(x)}} \tag{3}$$

Where the chance of the review being positive class (+1) or negative class (-1) is dependent on the input (x) and its learnt coefficient (w). The equation computed by 1 divided by 1 plus exponential (e) multiplied by the dot product of the transposed learnt coefficient (w) for each feature (h) in the input data (x) (Omari, et al., 2019).

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

13

Performance Evaluation

We use the assessment process to see how well we're doing. The confusion matrix is a tool for determining how well a model performed. Equation 4, equation 5, equation 6, and equation 7 contain the formulas for this evaluation confusion matrix. Figure 5 shows the findings of these Evaluation Matrices.

Accuracy in this instance refers to the proportion of correctly identified cases, which may be computed using the formula.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{4}$$

Recall is the proportion of genuine positives that were accurately detected. It was determined using a formula.

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{5}$$

Precision is the proportion of correct affirmative identifications. It was determined using a formula.

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{6}$$

Where,

TP = True Positive, sentiments that are positive and are actually classified as 1
TN = True Negative, sentiments that are negative and are actually classified as 0
FP = False Positive, sentiments that are negative but are classified as 1
FN = False Negative, sentiments that are positive but are classified as 0

The F1 score is the harmonic mean of precision and recall. It is given by the formula (Dcunha, 2019).

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (7)$$

RESULTS AND DISCUSSION

Because of its strong efficacy in binary classification tasks, Logistic Regression was used to categorize reviews into positive or negative categories. Logistic Regression Estimates if the probability between two variables is positive or negative using a Logistic function label y and data features w given by input x. Detailed below are the macro and weighted averages of precision and recall. The macroprecision of the classifier is 0.80. When predicting positive evaluations, the classifier produced better results. Additionally, the data shows that the recall values for positive prediction are strong, resulting in a high macro-average 0.84 percent.

Figure 5: Precision and Recall Values

	precision	recall	f1-score	support
neg	0.67	0.81	0.74	257
pos	0.93	0.87	0.90	786
accuracy			0.86	1043
macro avg	0.80	0.84	0.82	1043
weighted avg	0.87	0.86	0.86	1043

Source: Prepared by authors elaboration using the sklearn Python Package

Figure 6 demonstrates that the classifier produced a ROC curve by plotting True Positive (TP) on the y-axis and False Positive (FP) on the x-axis. The curve is near the y-axis, oriented up towards the number 1, which is a good thing to align top left. The area under the curve is scored 0.84, indicating a successful classifier.

Figure 6: Receiver Operating Characteristic (ROC)
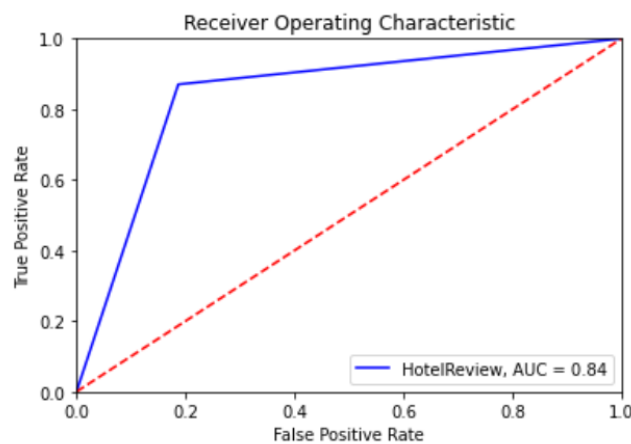
Source: Prepared by authors elaboration using the matplotlib Python Package

Figure 7 Text mining-specific criteria were used to assess the categorization performance of the show system. For the current system are calculated: precision, accuracy, recall, and F-measure. According to how to Logistic Regression, these values are produced based on a confusion matrix for all classified hotel reviews in the two classes (positive and negative).

Figure 7: Logistic Regression Confusion matrix
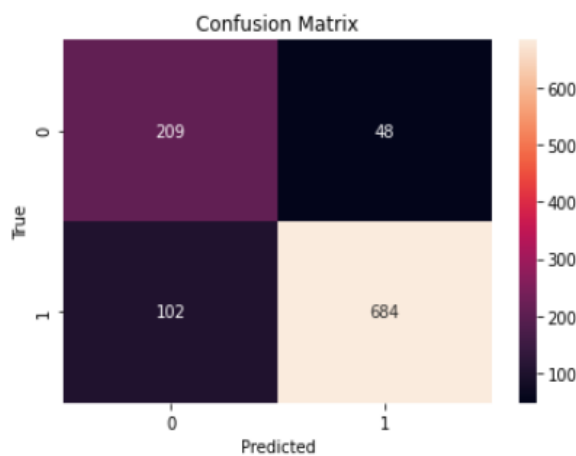
Source: Prepared by authors elaboration using the confusion matrix
Python Package

The confusion matrix show classification information according to the previous model. From a total of 3475 datasets, we split the data into two parts: the training portion of 70% and the testing portion of 30%.

The different values of the Confusion matrix would be as follows:

- True Positive (TP) = 684; meaning 684 positive class data points were correctly classified by the model.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

16

- True Negative (TN) = 209; meaning 209 negative class data points were correctly classified by the model.

- False Positive (FP) = 48; meaning 48 negative class data points were incorrectly classified as belonging to the positive class by the model.

- False Negative (FN) = 102; meaning 102 positive class data points were incorrectly classified as belonging to the negative class by the model.

Given the substantially larger number of true positive and true negative values, this proved out to be a quite effective classifier for our dataset.

CONCLUSION

This study addressed a crucial challenge in the domain of sentiment analysis for Thai language hotel reviews, leveraging a dataset of 3,475 reviews from Centara Hotel & Convention Center UdonThani. The experiments conducted yielded three key findings, shedding light on the classifier's strengths and potential areas for improvement.

Confidence in Positive Predictions

The classifier demonstrated a high level of confidence when predicting positive reviews. This confidence provides an opportunity for hotel managers to leverage insights from positive sentiments to enhance their services, focusing on the features highlighted by satisfied customers.

Bias in Negative Sentiment Prediction

Conversely, the classifier exhibited bias when forecasting negative sentiment in reviews. This identifies a specific area for improvement, urging a closer examination of the features influencing negative sentiments to rectify potential shortcomings.

Credibility of Logistic Regression

The high percentage of positive reviews within the corpus contributes to the overall credibility of the Logistic Regression model. This credibility is reflected in the model's accuracy rate of 86%, instilling confidence in its ability to effectively categorize sentiments.

The 86% accuracy of the model serves as a solid foundation for future research endeavors. To further enhance the classifier's performance, future work could involve expanding the dataset to encompass a broader range of services in the Thai language context.

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

17

Additionally, the exploration of various machine learning and deep learning classifiers may yield optimized results.

In summary, this study underscores the critical aspects for improvement identified in Centara Hotel & Convention Center UdonThani. By concentrating efforts on enhancing staff service skills, maintaining a beautiful and clean hotel environment, and offering friendly and reasonably priced services, hotels can elevate customer happiness and overall business performance. Focusing on these three factors has the potential to boost client satisfaction, generate positive feedback, and increase competitive advantages and market share in the highly competitive hotel industry. Looking ahead, future work aims to employ more powerful models for in-depth analysis of hotel features, contributing to comprehensive service improvements. The results obtained from this system can be harnessed to gauge public perceptions of specific hotels, facilitating strategies to enhance customer intimacy and satisfaction, thereby fostering success in the competitive hotel landscape.

REFERENCES

Alsaeedi, A. and Khan, M. Z. (2019). A Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications (IJACSA),Vol. 10, No. 2. https://doi.org/10.14569/IJACSA.2019.0100248

Abubakar, A. I., Roko, A., Bui, A. M. and Saidu, I. (2021). An Enhanced Feature Acquisition for Sentiment Analysis of English and Hausa Tweets. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 12, No. 9. https://doi.org/10.14569/IJACSA.2021.0120913

Apriliani, D., Abidin, T., Sutanta, E., Hamzah., A. and Somantri, O. (2020). Sentiment Analysis for Assessment of Hotel Services Review using Feature Selection Approach based-on Decision Tree. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 11. https://doi.org/10.14569/IJACSA.2020.0110432

Deewattananon, B. and Sammapun, U. (2017). Analyzing user reviews in Thai language toward aspects in mobile applications. 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), NakhonSiThammarat, Thailand, 2017, pp. 1-6, https://doi.org/10.1109/JCSSE.2017.8025903.

Dcunha, P. (2019). Aspect Based Sentiment Analysis and Feedback Ratings using Natural Language Processing on Irish Hotel. Masters thesis, Dublin, National College of Ireland. https://norma.ncirl.ie/4287/

Farisi, A. A., Sibaroni, Y. and Faraby, S. A. (2018). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. The 2nd International Conference on Data and Information Science, Vol. 1192. https://doi.org/10.1088/1742-6596/1192/1/012024

Ghojogh, B., Samad, N. M., Mashhadi, S. A., Kapoor, T., Ali, W., Karray, F. and Crowley, M. (2019). Feature Selection and Feature Extraction in Pattern Analysis: A Literature Review. Cornell University: arXiv: 1905.02845. https://doi.org/10. 48550/arXiv.1905.02845

Mahatthanachai, C. Malaivongs, K. Tantranont N. and Boonchieng, E. Development of thai word segmentation technique for solving problems with unknown words. International Computer Science and Engineering Conference (ICSEC), Chiang Mai, Thailand, 2015, pp. 1-6, https://doi.org/10.1109/ICSEC.2015.7401423

Omari, M. Al-Hajj, Hammami M. N. and Sabra, A. Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon. 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-5, https://doi.org/10.1109/ICCISci.2019.8716394

Panda, S., Gupta, S., Kumari, S. and Yadav, P. (2020). Sentiment Analysis Techniques and Approaches. International Journal of Engineering Research & Technology (IJERT), Vol. 9. https://doi.org/10.17577/IJERTV9IS060350

Patel, A., Jain, B., Chheda, B. and Gidwani, M. (2020). Sentiment Analysis of Customers Opinions on Hotel Stays using Voted Classifier. International Journal of Engineering Research & Technology (IJERT), Vol. 9. https://doi.org/10.17577/IJERTV9IS050615

Pathuri, S. K., ViswaGanesh, A., RaviTeja., P. and Rishi Chowdart, V. (2019). Feature-Based Opinion Mining for Amazon Product's using MLT. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8. https://doi.org/ 10.35940/ijitee.K1837.0981119

Samruddhi, K., Sowmyashree, A. And Ashok, K. R. (2019). Classification Model to Predict the Sentiment of Hotel Reviews. International Research Journal of Computer Science (IRJCS), Vol. 6. https://doi.org/10.26562/IRJCS.2019.JNCS10090

Sharad, D. A., Ashok, D. S., Dattatray, D. P. and Bhagwat, M. P. (2018). Sentiment Analysis of Hotel Review. International Journal of Advanced Research in Computer and Communication Engineering, Vol.7. https://doi.org/10.17148/IJARCCE.2018.71139

Shinde, G. and Deshmukh, N. S. (2016). Sentiment TFIDF Feature Selection Approach for Sentiment Analysis. International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), vol. 4. https://doi.org/10.15680/IJIRCCE.2016. 0407177

Sungsri, T. And Ua-apisitwong, U. (2017). The Analysis and Summarizing System of Thai Hotel Reviews Using Opinion Mining Technique. Proceeding of the 5th International Conference on Information and Education Technology, pp.167-170. https://doi.org/10.1145/3029387.3029391

Surbhi, S., Anant, D., Rupesh, P. and Reshma, P. (2023). Workplace Productivity Through Employee Sentiment Analysis using Machine Learning. International Journal of Professional Business Review, Vol. 8. No. 4. https://doi.org/10.26668/businessrevie-w/2023.v8i4.1216

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

19

Tanantong, T., Sanglerdsinlapachai, N. and Donkhampai, U. (2020). Sentiment Classification on Thai Social Media Using a Domain-Specific Trained Lexicon. 17th International Conference on Electrical Engineering / Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 580-583. https://doi.org/10.1109/ECTI-CON49241.2020.9158329

Tsujii, K., Takahashi, M. and Tsuda, K. (2015). Feature Extraction from Numerical Evaluation in Online Hotel Reviews. 19th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Vol. 60, pp. 1138-1145. https://doi.org/10.1016/j.procs.2015.08.164

Umawadee, U. and Wirapong, C. (2023).Trend and Knowledge Structure of Cryptocurrency Research in The Scopus Database. International Journal of Professional Business Review, Vol. 8. No. 4. https://doi.org/10.26668/businessreview/2023.v8i4.1640

Yazan, E. (2022). The Impact of Open-Book Management on Customer Satisfaction : The Mediating Role of Trust. International Journal of Professional Business Review, Vol. 7. No. 4. https://doi.org/10.26668/businessreview/2022.v7i4.e453

Yordanova, S. And Kabakchieva, D. (2017). Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning. International Journal of Computer Applications (IJCA), Vol. 158. https://doi.org/10.5120/ijca2017912806

Intern. Journal of Profess. Bus. Review. | Miami, v. 9 | n. 1 | p. 01-20 | e04152 | 2024.

20