# Forecasting the Business Performance of Restaurants on Social Commerce

**Supamit BOONTA[1], Kanjana HINTHAW[2]**

## Abstract

**Purpose:** This research delves into the various factors that influence the performance of restaurant businesses on social commerce platforms in Bangkok, Thailand. The study considers both internal and external factors, including but not limited to business characteristics and location. Moreover, this research also analyzes the effects of employing multiple social commerce platforms on business efficiency and explores the underlying reasons for such effects. **Research design, data, and methodology:** Restaurants can be classified into different price ranges: low, medium, and high. To further investigate, we employed natural language processing AI to analyze online reviews and evaluate algorithm performance using machine learning techniques. We aimed to develop a model to gauge customer satisfaction with restaurants across different price categories effectively. **Results:** According to the research findings, several factors significantly impact restaurant groups in the low and mid-price ranges. Among these factors are population density and the number of seats at the restaurant. On the other hand, in the mid-and high-price ranges, the price levels of the food and drinks offered by the restaurant play a crucial role in determining customer satisfaction. Furthermore, the correlation between different social commerce platforms can significantly affect the business performance of high-price range restaurant groups. Finally, the level of online review sentiment has been found to influence customer decision-making across all restaurant types significantly. **Conclusions:** The study emphasizes that restaurants' characteristics based on their price level differ significantly, and social commerce platforms have the potential to affect one another. It is worth noting that the sentiment expressed in online reviews has a more significant impact on customer decision-making than any other factor, regardless of the type of restaurant in question.

**Keywords :** Social Commerce, Machine Learning, Sentiment Analysis, Business Performance Prediction, Distribution Science

**JEL Classification Code**: C53, C55, C88

## 1. Introduction

Thailand's restaurant industry is a significant contributor to the country's economy. It is projected to reach a value of 4.35 trillion baht by 2023, with a growth rate of 7.1%. The industry is highly competitive and unpredictable, making it a challenging business. Despite the volatility experienced in the past year due to the global economic downturn and the

1 First Author. PhD candidate, Doctor of Philosophy Program in Business Administration and Digital Innovation, Mahasarakham Business School, Mahasarakham University, Thailand. Email: 63010990005@msu.ac.th

2 Corresponding Author. Advisor, Mahasarakham Business School, Mahasarakham University, Thailand. Email: kanjana.h@acc.msu.ac.th

COVID-19 pandemic's adverse effects, the restaurant business has shown resilience, experiencing steady growth as the situation improves. The evolving demographics fuel this growth, including a decrease in household sizes, urban expansion, and changing consumer lifestyles. Additionally, the fast-paced lifestyle and increasing demand for convenience have led to a rise in delivery services. Restaurants play a crucial role in the distribution of goods and services, connecting producers, suppliers, and customers. As a result of the evolution of digital technologies and changing consumer behaviors, new distribution channels and business models have emerged in the restaurant sector. Social commerce, which integrates social media and e-commerce, has become an increasingly important part of the modern distribution landscape for restaurants (Cho et al., 2019). Since restaurants are service businesses, both verbal and non-verbal communication strategies are crucial for effective service delivery, especially in counter-service restaurants (Choi, 2022). As a result, studying communication strategies to deliver services efficiently is essential for restaurant businesses. As of September 30, 2022, 18,096 restaurant businesses were operating in Thailand, accounting for 2.14% of the overall business landscape. Bangkok is the primary location for these establishments, accounting for 22.25% of the total. Despite the fact that the restaurant industry has been growing rapidly over the years, there are still a number of challenges that continue to plague the business. One of the biggest challenges is the intense competition in every segment and price level, which in turn affects the profitability and sustainability of the company. This is evident from the fact that a mere 35% of newly opened restaurants are able to sustain their business and continue operating beyond the three-year mark. These challenges require careful management and strategic planning in order to overcome the obstacles and succeed in this highly competitive industry.

The restaurant industry in Thailand is highly competitive, and consumers are increasingly reliant on online resources to make informed decisions about where to dine. Wongnai.com is a popular restaurant review website in Thailand that provides users with comprehensive information on food menus, restaurant location, type, price range, number of seats, and amenities such as parking, credit cards accepted, and delivery service. Users can also share their dining experiences by checking in, rating and reviewing, and leaving online comments to express their opinions and feelings. This electronic word-of-mouth (eWOM) concept significantly impacts online and offline businesses. According to a recent survey by Krungsri Research, Thai consumers who purchase products via social commerce tend to buy products primarily through Facebook, demonstrating the platform's effectiveness in accessing and engaging with websites or other platforms (Hirankasi & Klungjaturavet, 2022). As technology continues to rapidly evolve, restaurant businesses must adapt to stay competitive. Factors such as delivery services and communication channels through various social media platforms are becoming increasingly important in making business decisions. Social commerce platforms rely heavily on ratings and reviews to assess consumer satisfaction (Susskind, 2019). This research aims to study the internal and external factors affecting restaurant performance while analyzing whether there are any differences in performance among restaurants of different price levels. Critical factors on social commerce platforms include expanding communication channels through social media, such as Facebook Fan pages, and analyzing the levels of customer comments and feedback. Natural language processing and AI-powered sentiment analysis techniques can be employed to accurately evaluate customer reviews and opinions. A model can be developed by comparing the performance of forecasting algorithms and machine learning techniques to analyze the factors affecting restaurant performance on social commerce platforms. This will provide valuable insights for strategically managing restaurant businesses and help develop a decision support system for recommending business owners. Adopting such an approach will enable restaurant owners to evaluate the success or failure of their business before making decisions or during operations, increasing the likelihood of success and reducing the risk of business failure. As population, lifestyle, and technology continue to change, analyzing and forecasting restaurant performance on social commerce platforms is becoming increasingly critical for success in the industry.

## 2. Literature Review

### 2.1. Business Performance of a Restaurant

Running a successful restaurant business is a complex task that requires consideration of multiple factors. These factors include service quality, competitive pricing, and the presence of similar restaurants nearby, all of which can significantly impact the success or failure of the business (Wang, 2021). Location and geographic diversity also play a crucial role in determining the effectiveness of a restaurant business (Song, 2021). In large metropolitan areas, restaurants must pay close attention to their location and the surrounding commercial facilities and restaurants in the vicinity (Wu, 2021). In today's digital age, social media has become a powerful tool for consumers to share their opinions on the quality and various services of restaurants. Therefore, innovation activities have become an essential part of increasing the efficiency of the restaurant business

(Lee, 2018). Ultimately, the performance of a restaurant business depends on consumer satisfaction, which influences their intention to return and leads to business profits (Susskind, 2019). For social commerce, consumers can rate and review restaurants, reflecting their satisfaction with various factors.

## 2.2. Social Commerce

Social commerce is a modern approach to online shopping that combines the best of traditional e-commerce with the power of social media to reach and engage consumers within a network (Sohn, 2020). By creating virtual communities, social commerce can effectively promote the sale of products or services online, encouraging participation and sharing equally among users (Liao, 2021). This collaborative approach is precious for consumers who may be hesitant to purchase products they cannot physically touch (Boardman, 2019), as social commerce provides guidance and reduces the inherent uncertainty and risk associated with online shopping (Zhao, 2020).

In today's world, social media platforms have become an integral part of the consumer decision-making process. Before making any purchase, consumers tend to check out reviews and opinions from previous buyers. This trend is particularly prevalent in the restaurant industry, where customers look for feedback on the quality and service of restaurants before deciding where to dine. According to Gao (2018), consumer feedback plays a vital role in determining whether a product can meet the consumer's needs to a certain extent. Researchers have adopted data mining techniques from social commerce platforms to better understand restaurant business performance. Chang (2019) conducted research on predicting business performance based on location. The study aimed to identify insights for businesses that coexist and compete in geographic neighborhoods. The research used information from Yelp.com, a social commerce platform that provides member reviews and recommendations for businesses. The study analyzed 10,618 restaurant businesses in 2013 and found that the hybrid model was effective in supporting restaurants with their strategic and operational decisions. Leonard Gunawan (2023) researched the satisfaction levels of restaurant patrons. The study analyzed online reviews on TripAdvisor and used Support Vector Machine (SVM) techniques to classify the sentiment of the reviews. The results indicated that SVM was more accurate, with an accuracy value of 79%, compared to naive Bayes (NB). Overall, social media platforms serve as a data platform that enhances the quality of restaurant service through online reviews by consumers. They are a trusted source of information that helps consumers assess the quality of restaurants. Therefore, consumer reviews on the quality and service of restaurants are crucial.

## 2.3. Electronic Word of Mouth (eWOM)

Electronic word-of-mouth (eWOM) refers to the exchange of information or opinions, both positive and negative, through the Internet (Lim et al., 2022). This information is sourced from consumers, online social networks, or personal relationships (Chu, 2019), influencing the speed and rate of dissemination that consumers can choose from various platforms, such as online marketplaces or independent websites, and in various formats such as online reviews, blogs, and videos. Consumers can express their opinions through text, images, or ratings (Pyle, 2021). Consumers increasingly rely on electronic word-of-mouth to search for information about service providers and share personal experiences in using services (Golmohammadi, 2020). Consumer emotions are crucial in explaining post-purchase consumer behavior (Yan, 2018). Therefore, electronic word-of-mouth (eWOM) has become a significant reference for making purchase decisions and responding to consumer behavior in the digital age.

## 2.4. Emotional Intensity

Expressing emotions and feelings through online comments or reviews about products or services on various media platforms can vary significantly. Social networks like Facebook and review websites like wongnai.com are two such platforms that showcase different patterns of consumer behavior. Consumers tend to prioritize their self-image and interactions with other members on social media, often resulting in a positive bias in their comments. However, on review websites, where consumers are usually unfamiliar with each other, comments tend to align more with the genuine feelings of consumers (Liu, 2021). Conflicting online reviews, whether positive or negative, can help consumers decide whether they need more information about a product (Ruiz-Mafe, 2018). A study conducted by Li investigated the effect of emotional intensity on perceived usefulness. The study analyzed data from 600,686 reviews of 300 popular U.S. restaurants on Yelp using text mining and econometric analysis. The results showed that positive emotional intensity had a negative impact on the perceived usefulness of the review, whereas negative emotional intensity had a positive impact (Li, 2020). Evidently, consumers' perceptions and emotional responses play vital roles in their purchase intent, and online reviews can significantly enhance their confidence in decision-making.

## 2.5. Machine Learning

Machine learning (ML) is a powerful technique

involving training machines or computers to learn from large datasets, known as training datasets, to find answers and make accurate predictions. This process involves applying various methods and algorithms to the training data and then using these methods to predict the outcomes of new datasets (Raschka & Mirjalili, 2019). There are three main types of machine learning:

1) Supervised Learning involves using past data or variables with known answers as training data to predict unknown values or future data. This process involves adjusting parameters to minimize the difference between the target and computed outputs (Jo, 2021). If the variables are discrete, this process is called classification, while if the data or variables are continuous, regression or prediction includes classification with decision trees or neural networks.

2) Unsupervised learning, on the other hand, involves studying data without known answers or relationships between variables. This type of machine learning is used to find relationships or group data with similar characteristics (Kalita, 2022).

3) Reinforcement Learning (RL) is a type of machine learning that enables computers to learn by interacting with their environment through trial and error, similar to human learning. This type of learning consists of two main components: the agent and the environment, where the agent interacts with the environment to learn. RL is based on Markov decision processes (MDPs), which require agents to make decisions that account for environmental uncertainty and experience (Vamvoudakis et al., 2021). According to a predefined evaluation metric, the agent can receive a good score (Dong et al., 2020). RL aims to find the agent's policy by maximizing the desired outcome (Palmas et al., 2020).

## 2.6. Decision Tree

Decision Tree is a popular algorithm used in supervised learning. This algorithm works without parameters, making it a valuable tool for classification and prediction tasks. The decision-making process starts at the root node, which is the starting point for the algorithm's sequence of decisions. From there, branch nodes connect decision-making conditions, which divide the data into two or more branches based on the child nodes' conditions. Finally, the algorithm reaches the end nodes, also known as leaf nodes, representing the outcomes obtained after the classification process. Decision Tree is a versatile algorithm used in various applications, including finance, healthcare, and marketing (Sumathi et al., 2022). The process of the Decision Tree is shown in Figure 1.
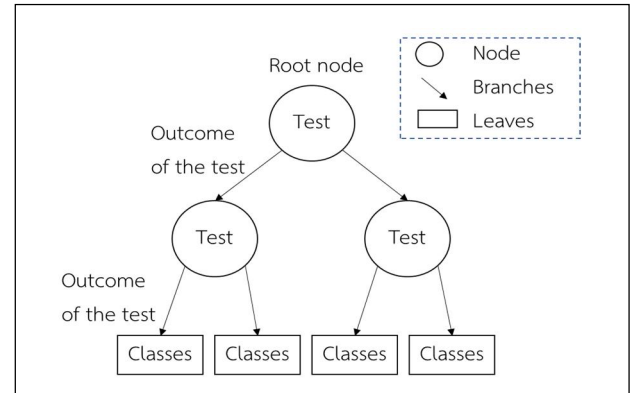


**Figure 1**: Decision Tree Structure

According to Figure1, the decision tree is a powerful tool known for its rule-based nature. In this approach, if-else rules are created based on the values of independent variables. The model does not rely heavily on equations dictating the relationships between independent and dependent variables. The approach is advantageous because it provides clear and easy-to-understand results, making it easy to apply in various scenarios.

## 2.7. Artificial Neural Network

Artificial neural networks (ANNs) are complex systems that possess the ability to learn and adapt from experience. They are designed to emulate the functioning of neural networks in the human brain, with operational units interconnected by an extensive network (Desai, 2021). An artificial neural network's performance largely depends on its learning process, which involves setting the network architecture and relevant parameters (Du, 2019). Typically, the architecture of a multilayer perceptron neural network comprises an input layer, hidden layers, and an output layer. To better understand this architecture, see Figure 2, which provides a visual representation.
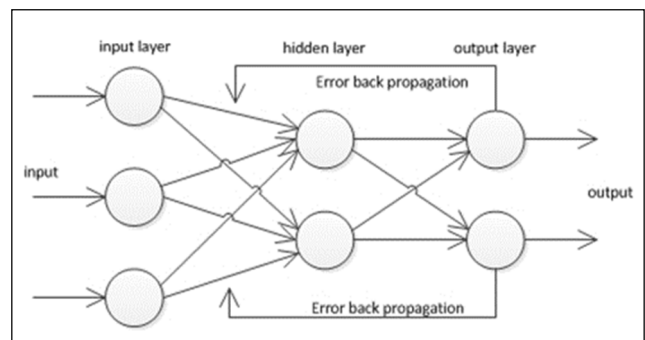


**Figure 2:** Architecture of the Multi-Layer Perceptron Neural Network

Each layer consists of a certain number of neurons, with data flowing continuously forward and the network error values spreading backward. The initial connection weights and criteria of the network start randomly. Then, a training process is employed by comparing the error values of the results with the actual values. The connection weights and criteria of the network are adjusted to reduce the error values until the results are close to the desired outcomes (Graupe, 2019).

## 2.8. Linear Regression Analysis

Linear Regression is a statistical method that helps to predict a dependent variable using one or more independent variables by establishing a linear relationship between the response variable and the predictor variables in a non-linear format. According to Ciaburro (2018), the primary objective of linear Regression is to determine the coefficients of the predictor variables (X) to showcase how these variables impact the response variable (Y), as depicted in the general equation of regression analysis in equation (1).

$$Y = a + b_1X_1 + b_2X_2 + b_3 + \ldots + b_nX_n \qquad (1)$$

When it comes to studying relationships between variables, it's important to understand the key terms involved. The independent variable, denoted as X, is the variable that is manipulated or changed in an experiment or study. On the other hand, the dependent variable, denoted as Y, is the variable that is being measured or observed. The intercept, denoted as a, represents the value of the dependent variable when the independent variable is equal to zero. Lastly, the parameter, denoted as b, represents the degree and direction of the relationship between the independent and dependent variables. It's important to keep these terms in mind when analyzing data and drawing conclusions.

## 2.9. Feature Selection

Feature selection is an essential pre-processing step in data analysis that involves reducing data size while maintaining the most appropriate data quality possible. There are several methods for feature selection, including Enter Regression, Forward Selection, and Backward Elimination. Enter Regression is a method of selectively introducing independent variables into the regression equation one at a time, using the criterion of the simple correlation coefficient. The process involves selecting the independent variable with the highest correlation with the dependent variable first and conducting a statistical test to determine if it can significantly predict the dependent variable. The process then continues by selecting the following independent variable in the order of relevance,

and the best criteria are applied in each round when combined with previously selected variables. This process continues until no further independent variables can be added to the equation if deemed appropriate. Forward selection is similar to enter Regression, where independent variables are selected for the equation one at a time using the criterion of the simple correlation coefficient. The process involves selecting the independent variable with the highest correlation with the dependent variable and conducting a statistical test to determine if it can significantly predict the dependent variable. The process then continues by selecting the following independent variable in the order of relevance, and the best criteria are applied in each round when combined with previously selected variables. This process continues until no further independent variables can be added to the equation if deemed appropriate (Bhadra & Bandyopadhyay, 2021).

Backward Elimination, on the other hand, is a method opposite to variable selection by stepwise addition. It starts with creating a regression equation that includes all independent variables. The p-value of each independent variable in the predictive model is then calculated (Nazarathy, 2021).

Subsequently, the variable with the highest p-value is considered, and its p-value is compared with the predefined significance level, such as $Pr = 0.05$. The variable is removed from the equation if the p-value exceeds the specified significance level. The process is repeated with the remaining variables until no p-value exceeds the designated significance level. The variable selection process is terminated at that point, and the equation is considered appropriate. Besides reducing the data size, feature selection reduces the learning process's complexity and overfitting. Overfitting occurs when a machine-learning model can make accurate predictions for the training dataset but fails to provide accurate forecasts for new data (Ba et al., 2023). This makes feature selection an essential step in data analysis and machine learning.

## 2.10. Method Evaluation

The confusion matrix is a valuable tool for evaluating the accuracy of binary classification or variables with only two groups of results. On the other hand, regression analysis uses different measurement methods, such as the R-squared value and mean absolute percentage error (MAPE) or root mean square error (RMSE), to estimate the relationship between a dependent variable and one or more independent variables.

Supervised learning is a type of machine learning that can measure accuracy because it uses a target as the starting point. This target value represents the expected outcome, and the model's forecast values can be compared against it

to measure the expected deviation from the initial data. Therefore, both classification and regression models can be evaluated for their accuracy, which indicates their efficiency.

### 2.10.1. Accuracy

The accuracy of a forecast model refers to its ability to make correct predictions. It is typically expressed as a percentage, calculated using equation (2).

$$Accuracy = \frac{TN+TP}{(TN+TP+FN+FP)} * 100 \qquad (2)$$

When evaluating the performance of a classification model, accuracy is one of the most commonly used metrics. It is calculated by taking into account four factors; true positive (TF), false positive (FP), true negative (TN), and false negative (FN). True positive refers to the number of correct positive predictions made by the model, false positive refers to the number of incorrect positive predictions made by the model, true negative refers to the number of correct negative predictions made by the model, and false negative refers to the number of incorrect negative predictions made by the model. By analyzing these factors, the accuracy of the model can be calculated, which gives an idea of how well the model is performing.

### 2.10.2. Recall

The term "recall" is used in machine learning to refer to a completeness score that indicates how well a model can select a significant amount of relevant information or answers in its predictions. This score is typically calculated using equation (3) and is an essential metric for evaluating the performance of a machine learning model. The recall score tells us how many relevant items were correctly predicted by the model out of all the relevant items in the dataset. A high recall score is generally desirable, as it indicates that the model can identify most of the relevant information or answers, even if it also produces some false positives.

$$Recall = \frac{TP}{(TP+FN)} * 100 \qquad (3)$$

### 2.10.3. Precision

Precision is a metric that determines the ability of a predictive model to filter out irrelevant answers or data. It serves as an indicator of the accuracy of the predictive model, as depicted by equation (4). Essentially, the higher the precision, the more effectively the predictive model is able to remove irrelevant information, leading to a more accurate and reliable prediction.

$$Precision = \frac{TP}{(TP+FP)} * 100 \qquad (4)$$

### 2.10.4. F-measure

The F-measure is a widely used evaluation metric that provides a comprehensive measure of the performance of a binary classification system. It is calculated by combining two critical metrics, precision and recall, into a single score. The resulting score provides an overall measure of the classification system's effectiveness in correctly identifying positive cases while minimizing false positives and false negatives. The equation (5) is used to compute the F-measure score.

$$F\text{-}measure = \frac{2.Precision.Recall}{Precision+Recall} * 100 \qquad (5)$$

## 2.11. Weight by Correlation

The Weight by Correlation operator is a highly useful tool for determining the relevance of attributes by computing the correlation value for each attribute of the input ExampleSet concerning the label attribute. This approach uses Correlation as a weighting scheme and returns the absolute or squared value of Correlation as the attribute weight.

The weight of an attribute is calculated concerning the label attribute using Correlation. The higher the weight of an attribute, the more relevant it is considered. Correlation is a numerical value that measures the degree of association between two attributes, X and Y. Correlation ranges from -1 to +1, where a positive correlation implies a positive association and a negative correlation value suggests a negative or inverse association. In the case of a positive correlation, large values of X tend to be associated with large values of Y, and small values of X tend to be associated with small values of Y. On the other hand, in a negative correlation, large values of X tend to be associated with small values of Y, and small values of X tend to be associated with large values of Y. Suppose two attributes are X and Y, with means X' and Y' and standard deviations S(X) and S(Y), respectively. The Correlation is computed as the summation from 1 to n of the product (X(i) X')·(Y(i) Y') and then dividing this summation by the product (n 1)·S(X)·S(Y), where X(i)-X')·(Y(i)-Y') and then dividing this summation by the product (n-1)·S(X)·S(Y) where n is the total number of examples and i is the increment variable of the summation. In a positive Correlation, if an X value is above average, the associated Y value is also above average. Then, the product (X(i)-X')·(Y(i)-Y') would be the product of two positive numbers that would be positive. If the X and Y values were below average, the product above would be of two negative numbers, which would also be positive. Therefore, a

positive correlation is evidence of a general tendency that large values of X are associated with large values of Y, and small values of X are associated with small values of Y. In a negative correlation, if an X value was above average, and the associated Y value was below average, the product (X(i) X')·(Y(i) Y') would be the product of a positive and a negative number, making the product negative. If the X value is below average, and the Y value is above average, the product above would also be negative. Therefore, a negative correlation is evidence of a general tendency that large values of X are associated with small values of Y, and small values of X are associated with large values of Y.

# 3. Research Methodology

## 3.1. Data Collection Methods

The researcher has conducted a comprehensive study, gathering data from various sources to use as input for analysis. Specifically, restaurant-related data was collected from the wongnai.com website within the Bangkok metropolitan area. This includes detailed information about restaurants, such as their categories, price ranges, parking availability, the number of reviews, credit card acceptance, Wi-Fi service, suitability for groups, provision of alcoholic beverages, website availability, delivery service, and review scores. Furthermore, Facebook information for the restaurants, such as check-ins, page likes, and followers, was also collected. In addition, population density data in the area was obtained from Bangkok's Geographic Information System (GIS) Technology Center. Google utilized natural language AI to analyze the sentiment of the customers' reviews on wongnai.com. The sentiment analysis will provide valuable insights into the opinions about the restaurants. Lastly, all the gathered information was recorded in a database for the following data preparation steps.

## 3.2. Data Preprocessing Methods

The methodology employed for data analysis in this research involves utilizing advanced machine learning techniques. A detailed workflow diagram, as illustrated in Figure 3, showcases the step-by-step process of this data analysis approach.
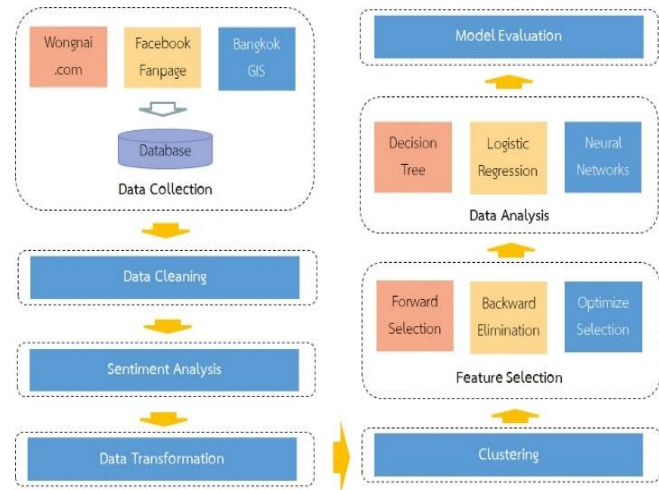


**Figure 3:** Data Processing and Analysis Steps

## 3.3. Data Cleaning

To ensure accurate analysis, it is important to carefully examine and address any outliers in the data set. For instance, if certain restaurants' values are missing due to incomplete information obtained from websites, these values must be replaced with appropriate substitutes. One common approach is to replace the missing factor values with 0, which allows for a more comprehensive and reliable analysis.

## 3.4. Sentiment Analysis

This involves analyzing customer feedback from online reviews, where customers who have used the restaurant services write their opinions on the wongnai.com website. Natural Language AI, a cloud service by Google, is used for this analysis. Machine learning can analyze unstructured text by understanding natural language (NUL). This is often used for sentiment analysis to understand customer opinions and obtain in-depth information that can be used to develop business products and services. The results are a score that can be interpreted as follows: 1) -1.0 to -0.25 represents negative sentiment, 2) -0.25 to 0.25 represents neutral sentiment, and 3) 0.25 to 1.0 represents positive sentiment.

## 3.5. Data Transformation

The researcher has transformed the data into a format suitable for analysis using machine learning techniques, as illustrated in Table 1.

**Table 1:** The Criteria for Data Transformation

| Factor | Description |
|---|---|
| Price | price range: <100 = 1; 101-250 = 2; 251-500 = 3; 501-1000 = 4; >1000 = 5 |
| Seats | seating capacity: 10 = 1; 40 = 2; 80 = 3; 150 = 4 |
| Parking | Parking availability: none = 1; available = 2 |
| Delivery | Delivery service: none = 1; available = 2 |
| CreditCard | credit card acceptance: none = 1; available = 2 |
| Wi-Fi | Wi-Fi service available: none = 1; available = 2 |
| Animal | pets allowed: none = 1; available = 2 |
| Group | suitable for group gatherings: none = 1; available = 2 |
| Alcohol | alcoholic beverages available: none = 1; available = 2 |
| SocialMedia | Other social media available: none = 1; available = 2 |
| Population Density | Population density in the area: <50,000 = 1; 50,000–100,000 = 2; 100,001–150,000 = 3; 150,001–200,000 = 4; >200,000 = 5 |
| Reviews | number of reviews: <50 = 1; 51-100 = 2; 101-150 = 3; 151-200 = 4; >200 = 5 |
| Sentiment | sentiment level: -1.0 -0.25 = 1; -0.26- 0.24 = 2; 0.25 1.0 = 3 |
| Check-in | Number of check-ins: <5,000 = 1; 5,000–10,000 = 2; 10,001–15,000 = 3; 15,001–20,000 = 4; >20,000 = 5 |
| Like | number of Facebook page likes: <25,000 = 1; 25,000–50,000 = 2; 50,001–75,000 = 3; 75,001–100,000 = 4; >100,000 = 5 |
| Followers | Number of Facebook followers: <25,000 = 1; 25,000–50,000 = 2; 50,001–75,000 = 3; 75,001–100,000 = 4; >100,000 = 5 |
| RatingScore | review score: 0 – 3.9 = (not good); >3.9 = 1 (good) |

## 3.6. Clustering

The researcher grouped restaurant types into three categories using the average price, a distinctive characteristic of each type. This was done to understand how each category has factors that influence the restaurant business's performance differently. This research chose to use the K-means Clustering technique for grouping.

## 3.7. Feature Selection

The data size reduction process implemented in this study was performed while ensuring that the optimal data quality for analysis was preserved. This was done by utilizing the evolutionary selection technique to calculate the weights of the specified factors. The weight assigned to each factor was determined based on relevance, with the higher-weighted factors being considered more significant. The inclusion or exclusion of a factor was based on its weight calculated during the process. Various methods were used to select the variables, such as forward selection, backward elimination, and enter regression. The weights of the specified factors were calculated to reduce the data size while preserving the optimal data quality for analysis using the evolutionary selection technique. The higher the weight of a factor in the calculation, the more relevant it is considered. If a factor's calculation results in a higher weight, it is retained. Conversely, if a factor's calculation results in a lower weight, the factor is excluded from the data analysis process. The methods employed in this study include forward selection, backward elimination, and enter regression for variable selection.

## 3.8. Data Analysis

This study compares the predictive model performance created using machine learning algorithms with RapidMiner software. The models include decision trees, artificial neural networks, and linear regression analysis. The data was divided into two parts for testing: the training dataset and the remaining data for testing, using the K-fold Cross-Validation method.

## 3.9. Method Evaluation

The performance of the prediction models was evaluated and compared by considering Accuracy and overall performance measured by the F-measure.

## 3.10. Weight by Correlation

Select factors influencing review scores reflecting the performance of restaurants by choosing a weight > 0.1.

## 4. Results and Discussion

## 4.1. Data Collection

We have conducted a comprehensive study and collected data from various sources using a program developed with C# .NET to retrieve web page content. The program has been designed to extract data from websites based on multiple factors observed on the web pages for a thorough

analysis. The study has encompassed the following factors:

1) Restaurant data within the Bangkok area, selected from restaurants with a Facebook fan page and information on the Wongnai website, totaling 1,750 establishments. After a thorough review of errors and missing data, 113 entries were identified, resulting in a final count of 1,637 restaurants. The study has included various aspects such as restaurant categories, price ranges, parking availability, number of reviews, credit card acceptance, Wi-Fi service, suitability for groups, availability of alcoholic beverages, website functionality, delivery service, and review scores.

2) Facebook fan page data for restaurants, including check-ins, page likes, and follower counts.

3) Population density data in the area obtained from Bangkok's Geographic Information System (GIS) Technology Center.

4) Sentiment level analyzed using Natural Language AI developed by Google, based on customer reviews for each restaurant on Wongnai.com.

The data will now be stored in a database for the data-cleaning process.

## 4.2. Clustering

To analyze the various types of restaurants, we classified them into three distinct groups. To achieve this, the researcher utilized the K-means Clustering technique, which involved grouping the restaurants based on similar characteristics and features. The results of this analysis were then presented in Table 2, which provided a clear and concise illustration of the different restaurant types and their respective features.

**Table 2:** The Data on Categorizing Restaurant Types Using the K-means Clustering Technique

| Cluster | Type of restaurant | Average price range | Price level |
|---------|-------------------|---------------------|-------------|
| Cluster_0 | noodles, coffee/tea shop, congee, street food, lunch box, À La Carte | 1.557 | low |
| Cluster_1 | shabu/sukiyaki/nabe, sushi, Izakaya, Japanese food, buffet, grill, seafood, fusion food | 3.527 | high |
| Cluster_2 | ramen, steak, chinese food, Thai food, isaan food, international food, bakery/cake, cafe | 2.55 | moderate |

## 4.3. Forecasting Model

The research employs RapidMiner software to develop three forecasting models: Decision Tree, Logistic Regression, and Artificial Neural Networks. Cross-validation is used to evaluate the performance of each type, where the data is split into ten equal parts using the 10-fold

cross-validation technique. Once the data is divided, variable selection is performed through Enter Regression, Forward Selection, and Backward Elimination. Finally, the performance of the best forecasting models from each type is compared. The findings of this research are as follows: the Decision Tree model performed the best, followed by the Artificial Neural Network and Logistic Regression models. However, all three models showed similar accuracy levels, and the difference in performance was not statistically significant.

**Table 3:** Comparison of the Performance of Forecasting Models for the Group of Low-priced Restaurants. (cluster_0)

| Forecasting Model | Feature Selection | Accuracy | F-measure |
|-------------------|-------------------|----------|-----------|
| Decision Tree | forward selection | 80.22% | 83.05% |
| | backward elimination | 80.65% | 82.79% |
| | enter regression | 80.43% | 83.47% |
| Logistic regression | forward selection | 81.52% | 84.10% |
| | backward elimination | 81.09% | 78.60% |
| | enter regression | 81.09% | 84.11% |
| Artificial neural networks | forward selection | 80.65% | 83.46% |
| | backward elimination | 84.13% | 86.10% |
| | enter regression | **85.22**% | **87.47**% |

Based on the data presented in Table 3, it is evident that the artificial neural network is the most efficient algorithm for forecasting. This algorithm has considered all variables (Enter Regression) and achieved an impressive accuracy rate of 85.22%. Furthermore, the artificial neural network has emerged as the most efficient algorithm when considering the overall prediction performance based on the F-measure. It has displayed an overall prediction performance of 87.47%, which is a remarkable achievement in forecasting. These findings suggest that the artificial neural network is highly dependable and trustworthy for making accurate predictions.

**Table 4:** Comparison of Correlation Coefficients for the Group of Low-priced Restaurants

| Factors | Weight |
|---------|--------|
| **Sentiment** | **0.6561** |
| **Seats** | **0.2117** |
| **Group** | **0.1640** |
| **Wi-Fi** | **0.1405** |
| **PopulationDensity** | **0.1071** |
| Like | 0.0954 |
| Reviews | 0.0890 |
| Delivery | 0.0679 |
| SocialMedia | 0.0573 |
| Alcohol | 0.0561 |
| Price | 0.0507 |
| CreditCard | 0.0358 |
| Parking | 0.0323 |
| Check-in | 0.0204 |
| Website | 0.0024 |

Table 4 displays the outcomes of the enter regression method used to select the factors that affect the review scores of low-priced restaurants on social commerce based on the artificial neural networks forecasting model. The analysis revealed that Sentiment, Seats, Group, Wi-Fi, and Population Density are the crucial factors that influence the performance of low-priced restaurants. These factors were chosen based on their weight, which was greater than 0.1 (> 0.1).

**Table 5:** Results of Comparing the Performance of Forecasting Models for the Group of High-priced Restaurants (cluster_1)

| Forecasting Model | Feature Selection | Accuracy | F-measure |
|---|---|---|---|
| Decision Tree | forward selection | 83.96% | 77.87% |
| | backward elimination | 84.32% | 79.61% |
| | enter regression | 83.78% | 78.85% |
| Logistic regression | forward selection | 84.31% | 77.36% |
| | backward elimination | 83.07% | 76.40% |
| | enter regression | 82.71% | 76.07% |
| Artificial neural networks | forward selection | 85.92% | 81.76% |
| | backward elimination | 87.51% | 83.02% |
| | enter regression | **91.08**% | **88.06**% |

According to the experimental results, the best algorithm for forecasting is the artificial neural network. This algorithm uses the "Enter Regression" method, where all variables are considered, resulting in an accuracy of 91.08%. Additionally, when considering the overall prediction performance (F-measure), the artificial neural network outperforms other algorithms with an overall forecasting performance of 88.06%.

**Table 6:** Comparison of Correlation Coefficients for the Group of High-priced Restaurants

| Factors | Weight |
|---|---|
| **Sentiment** | **0.6954** |
| **Price** | **0.2364** |
| **Alcohol** | **0.1880** |
| **CreditCard** | **0.1129** |
| **Like** | **0.1109** |
| Delivery | 0.0969 |
| Wi-Fi | 0.0894 |
| Group | 0.0685 |
| Seats | 0.0591 |
| Parking | 0.0393 |
| Check-in | 0.0326 |
| Website | 0.0139 |
| Reviews | 0.0085 |
| SocialMedia | 0.0070 |
| PopulationDensity | 0.0043 |

According to the data presented in Table 6, the artificial neural networks forecasting model has identified several key factors influencing the review scores of low-priced restaurants on social commerce platforms. These factors were selected using the enter regression method and were assigned a weight >0.1. The most significant factors were Sentiment, Price, Alcohol, CreditCard, and Like. These factors play a crucial role in determining the performance of low-priced restaurants on social commerce platforms. On the other hand, population density was identified as the least significant factor, which has a negligible impact on the review scores.

**Table 7:** The Results of Comparing the Performance of Forecasting Models for the Moderate-priced Restaurant Group (cluster_2)

| Forecasting Model | Feature Selection | Accuracy | F-measure |
|---|---|---|---|
| Decision Tree | forward selection | 80.91% | 79.89% |
| | backward elimination | 82.21% | 82.58% |
| | enter regression | 80.43% | 81.12% |
| Logistic regression | forward selection | 80.91% | 79.89% |
| | backward elimination | 82.22% | 82.82% |
| | enter regression | 81.57% | 82.12% |
| Artificial Neural Networks | forward selection | 80.10% | 80.20% |
| | backward elimination | **85.46**% | **85.89**% |
| | enter regression | 82.54% | 83.21% |

According to the data presented in Table 7, the artificial neural network algorithm has proven to be the most efficient in forecasting, outperforming all other algorithms by a considerable margin. The algorithm employs the backward elimination method for variable selection and has achieved a high accuracy rate of 85.46%. When assessing the overall forecasting performance (F-measure), the artificial neural network algorithm remains the most efficient, with an impressive overall prediction performance rate of 85.89%. In analyzing moderate-priced restaurants' performance in social commerce, the backward elimination method was used with the artificial neural networks forecasting model. The result is presented in Table 8, which shows the factors influencing review scores. The factors were ranked based on their weight by correlation, with a weight > 0.1 being considered significant. The three factors most significantly impacting review scores were Sentiment, Seats, and Price. On the other hand, the factor with the most negligible effect was CreditCard.

**Table 8:** The Results of Comparing Correlation Coefficients for the Group of Moderate-priced Restaurant.

| Factors | Weight |
|---|---|
| **Sentiment** | **0.6763** |
| **Seats** | **0.1528** |

| Factors | Weight |
|---|---|
| **Price** | **0.1085** |
| PopulationDensity | 0.0748 |
| Reviews | 0.0671 |
| Alcohol | 0.0614 |
| Website | 0.0471 |
| SocialMedia | 0.0401 |
| Parking | 0.0296 |
| Delivery | 0.0281 |
| Check-in | 0.0182 |
| CreditCard | 0.0018 |

# 5. Conclusions and Future Work

The study conducted on restaurants using artificial neural network (ANN) techniques revealed that these techniques provide the most accurate forecasting performance across all types of restaurants. Hence, they can be effectively employed to build models for forecasting the performance of restaurant businesses on social commerce platforms. The research has also identified the key factors that influence the review scores of restaurants, reflecting customer satisfaction levels within each restaurant category and price range. The selection process identified factors significantly correlated to each category's restaurant review scores. These factors can be summarized as restaurant business recommendations, which can help them improve their overall performance. Regarding internal factors, the study found that the price level had the most significant impact on customer satisfaction, especially for restaurant categories with mid-range and high price ranges. Seating capacity also affected low- and mid-range price categories, while population density influenced customer satisfaction. Therefore, adequate seating should be provided to accommodate large numbers of customers. Regarding other services, offering Wi-Fi is another factor that can attract customers to low-priced restaurants. For high-priced restaurants, consideration should be given to accepting credit cards and providing alcohol. Regarding social media marketing, the level of online review comments has the most significant impact on customer decision-making across all restaurant categories. Specifically, the number of likes on Facebook, which restaurants can use as a platform for publicity, influences customer satisfaction levels, particularly for high-priced restaurant categories.

In summary, the study found that the sentiment level is a crucial factor that influences the review scores of restaurants, reflecting customer satisfaction levels and ultimately impacting the performance of restaurant businesses on social commerce platforms at every price level. This underscores the significant role of word-of-mouth marketing in the restaurant business, particularly in the high-priced category. Using other social media platforms, such as Facebook, can effectively enhance consumer awareness, reducing restaurant operators' concerns regarding place-oriented marketing strategies and lowering operational costs to maintain competitiveness. Therefore, the research should focus on analyzing sentiments from various perspectives and categorizing reviews by content type to extract valuable insights for efficient utilization in the social commerce domain for restaurant businesses. Additionally, the variety of the types of foods may be another crucial factor to consider when evaluating restaurant services.

# References

Ba, J., Wang, P., Yang, X., Yu, H., & Yu, D. (2023). Glee: A granularity filter for feature selection. *Engineering Applications of Artificial Intelligence, 122*, 106080.

Bhadra, T., & Bandyopadhyay, S. (2021). Supervised feature selection using integration of densest subgraph finding with floating forward–backward search. *Information Sciences, 566*, 1-18.

Boardman, R. (2019). *Social Commerce: Consumer Behaviour in Online Environments*. Switzerland: Springer International Publishing.

Chang, X. (2019). Business performance prediction in location-based social commerce. *Expert Systems With Applications, 126*, 112–123.

Cho, M., Bonn, M. A., & Li, J. J. (2019). Differences in perceptions about food delivery apps between single-person and multi-person households. International Journal of Hospitality Management, 77, 108-116. https://doi.org/10.1016/j.ijhm.2018.06.019

CHOI, J. (2022). Distribution Strategies for Service Delivery: Focus on Verbal and Non-verbal Communication at Counter Service Restaurants. The Journal of Distribution Science, 20(3), 45-52, https://doi.org/https://doi.org/10.15722/jds.20.03.202203.45

Chu, S.-C. (2019). *Electronic Word of Mouth as a Promotional Technique: New Insights from social media*. New York: Routledge.

Ciaburro, G. (2018). *Regression Analysis with R*. Birmingham: Packt Publishing Ltd.

Desai, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth, 4*, 1-11.

Dong, H., Zhang, S., & Ding, Z. (2020). *Deep Reinforcement Learning*. Singapore: Springer Nature Singapore Pte Ltd.

Du, K.-L. (2019). *Neural Networks and Statistical Learning*. London: Springer.

Gao, S. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management, 71*, 19-32.

Golmohammadi, A. (2020). Negative online reviews and consumers' service consumption. *Journal of Business Research, 116*, 27-36.

Graupe, D. (2019). *PRINCIPLES OF ARTIFICIAL NEURAL NETWORKS*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Gunawan, L., Anggreainy, M. S., Wihan, L., Santy, Lesmana, G. Y., & Yusuf, S. (2023). Support vector machine based emotional analysis of restaurant reviews. *Procedia Computer Science, 216*.

Hirankasi, P., & Klungjaturavet, C. (2022, 29 September 2021). Social Commerce: The New Wave of E-commerce. Retrieved from https://www.kasikornresearch.com/TH/analysis/k-econ/business/Pages/Restaurant-Business-Y23-CIS3429-B-24-08-2023.aspx

Hussain, S. (2018). Consumers' online information adoption behavior: Motives and antecedents of electronic word of mouth communications. *Computers in Human Behavior, 80*.

Jo, T. (2021). *Machine Learning Foundations*. Switzerland: Springer International Publishing.

Kalita, J. (2022). *Machine Learning Theory and Practice*. UK: CRC Press.

Lee, C. (2018). Investigating the moderating role of education on a structural model of restaurant performance using multi-group PLS-SEM analysis. *Journal of Business Research, 88*, 298-305.

Li, H. (2020). Online persuasion of review emotional intensity: A text mining analysis of restaurant reviews. *International Journal of Hospitality Management, 89*, 1-13.

Liao, S.-H. (2021). Investigating online social media users' behaviors for social commerce recommendations. *Technology in Society, 66*, 1-14.

Lim, W. M., Ahmed, P. K., & Ali, M. Y. (2022). Giving electronic word of mouth (eWOM) as a prepurchase behavior: The case of online group buying. *Journal of Business Research, 146*, 582-604.

Liu, H. (2021). Social sharing of consumption emotion in electronic word of mouth (eWOM): A cross-media perspective. *Journal of Business Research, 132*, 208-220.

Nazarathy, Y., H. K. (2021). *Statistics with Julia: Fundamentals for Data Science, Machine Learning and Artificial Intelligence*. Gewerbestrasse 11 6330 Cham Switzerland: Springer.

Palmas, A., Ghelfi, E., Petre, A. G., Kulkarni, M., N.S., A., Nguyen, Q., ... & Basak, S. (2020). *The The Reinforcement Learning Workshop*. UK: Packt Publishing.

Pyle, M. A. (2021). In eWOM we trust: Using naïve theories to understand consumer trust in a complex eWOM marketspace. *Journal of Business Research, 122*, 145-158.

Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning : Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*: Packt Publishing Ltd.

Ruiz-Mafe, C. (2018). The role of emotions and conflicting online reviews on consumers' purchase intentions. *Journal of Business Research, 89*, 336-344.

Sohn, J. W. (2020). Factors that influence purchase intentions in social commerce. *Technology in Society, 63*, 1-11.

Song, H. J. (2021). The influence of board interlocks on firm performance: In the context of geographic diversification in the restaurant industry. *Tourism Management, 83*, 1-9.

Sumathi, S., Rajappa, S. V., Kumar, L. A., & Paneerselvam, S. (2022). *Machine Learning for Decision Sciences with Case Studies in Python*. Boca Raton: CRC Press.

Susskind, A. M. (2019). *The Next Frontier of Restaurant Management: Harnessing Data to Improve Guest Service and Enhance the Employee Experience*. New York: Cornell University Press.

Vamvoudakis, K. G., Wan, Y., Lewis, F. L., & Cansever, D. (2021). *Handbook of Reinforcement Learning and Control*. switzerland: Springer International Publishing.

Wang, Y. (2021). Interconnectedness between online review valence, brand, and restaurant performance. *Journal of Hospitality and Tourism Management, 48*, 138-145.

Wu, M. (2021). Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data. *Cities, 113*, 1-14.

Xu, J., Lu, W., Li, J., & Yuan, H. (2022). Dependency maximization forward feature selection algorithms based on normalized cross-covariance operator and its approximated form for high-dimensional data. *Information Sciences, 617*, 416-434.

Yan, Q. (2018). The influences of tourists' emotions on the selection of electronic word of mouth platforms. *Tourism Management, 66*, 348-363.

Zhao, Y. (2020). Electronic word-of-mouth and consumer purchase intentions in social e-commerce. *Electronic Commerce Research and Applications, 41*, 1-9.