 ISSN 2579-7298

Comparison of Ensemble Learning Methods for Mining the Implementation of the 7 Ps Marketing Mix on TripAdvisor Restaurant Customer Review Data

Budi Sunarko [a,1,*], Uswatun Hasanah [b,2], Syahroni Hidayat [c,3]

[a] Department of Informatics and Computer Engineering Education, Faculty of Engineering, Universitas Negeri Semarang, Sekaran, Semarang 50229, Indonesia.
[b] Department of Computer Engineering, Faculty of Engineering, Universitas Negeri Semarang, Sekaran, Semarang 50229, Indonesia
[c] Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang, Sekaran, Semarang 50229, Indonesia
[1] budi.sunarko@mail.unnes.ac.id[1], uswatun_hasanah@mail.unnes.ac.id[2]; syahroni.hidayat@mail.unnes.ac.id[3]
* corresponding author

ARTICLE INFO

Article history
Received 29 Jul 2023
Revised 08 Sep 2023
Accepted 12 Des 2023

Keywords
7P marketing mix
Ensemble learning
Online review
Sentiment analysis
Text classification

ABSTRACT

The 7P marketing mix encompasses various business facets, notably the Process element governing internal operations from production to customer service. With the surge in online customer feedback, assessing machine learning efficacy, especially ensemble learning, in classifying 7P-related customer review data has gained prominence. This research aims to fill a gap in existing literature by evaluating ensemble learning's performance on 7P classification, an area not extensively explored despite prior sentiment analysis studies. Employing a methodology merging Natural Language Processing (NLP) with ensemble learning, the study processes restaurant reviews using NLP techniques and employs ensemble learning for precision and accuracy. Findings demonstrate that DESMI yielded the highest performance metrics with accuracy at 0.697, precision at 0.699, recall at 0.697, and an F1-score of 0.684. These outcomes underscore ensemble learning's potential in handling complex datasets, signifying its relevance for marketers and researchers seeking comprehensive insights from customer reviews within the 7P marketing mix domain. This study sheds light on how ensemble learning outperforms its foundational methods, indicating its prowess in extracting meaningful insights from diverse and intricate customer feedback.

1. Introduction

Marketing mix covers four main elements (4P), namely Product, Price, Place, and Promotion. It is known as Marketing mix 4P. However, in the developing intelligent business era, the marketing mix develops to 7P, including three other Ps: Process, Physical Evidence, and People. Furthermore, the Marketing 7P has been established as classical work frame for marketing management [1]. The 7P marketing mix 7P emphasizes a variety of customer needs to recognize customer behavior from production to after sale services. Deep insight of customer needs and preferences are needed to result in appropriate products, good prices, and strategic location [2]. In the intelligent business era, data is important for companies to understand customers' behaviors well. According to Kotler and Keller, effective marketing depends on deep understanding of customer's needs, desire, and behaviors [3]. Intelligent business has a key playing role in deriving knowledge from data [4].

The 7P marketing mix includes the Process element involving internal business management from production to customer services [5]. In this case, companies are possible for companies to integrate data and business process analysis for identifying weaknesses and efficiency. This can impact on the overall performance of companies. The intelligence business supports in collecting, storing, and analyzing operational data to create opportunities for increasing efficiency and improving business process. In the case of the marketing mix, the Price and Promotion elements determine to what extent a product competes in the market. By using intelligence business, companies can analyze market trends, consumers' behaviors, and competitors' strategies to get price an optimal price and its promotion. The use of intelligent business analysis can support companies to make an appropriate decision for determining price and effectively developing promotion strategies [6].

In case of the marketing mix, the Price and Promotion elements determine to what extent a product compete in the market. By using intelligence business, companies can analyze market trends, consumers' behaviors, and competitors' strategies to get price an optimal price and its promotion. The use of intelligence business analysis can support companies to make an appropriate decision for determining price and effectively developing promotion strategies [7]. The 7P marketing mix emphasizes the importance of physical evidence and people interaction in branding and providing a positive customer's experience. The Intelligence Business supports companies to learn the interaction of physical environment and customer influencing the brand. It can also potentially provide insight into customer interactions, help companies compile physical evidence and build strong relationships with customers [8].

In a digital era, abundant information, the use of technology to optimize marketing strategies has become increasingly important. While there are many opportunities to increase the effectiveness of marketing strategies through the integration of the 7P marketing mix with BI, challenges also arise [9]. One of the main challenges is the complexity of managing and analyzing data obtained from various sources, especially for companies in the hotel and restaurant sector. As the number of customers sharing their experiences online increases, evaluating the performance of machine learning, especially ensemble learning, on 7P classification of customer review data is becoming a major concern. The application of this technique is expected to provide in-depth insights related to various aspects such as Product, Price, Place, Promotion, Process, Physical Evidence and People. The importance of customer review data cannot be ignored in the hotel and restaurant industry. Modern customers tend to search and leave reviews online before choosing a place to eat or stay. Therefore, companies in this sector are increasingly realizing the immense value contained in customer review data. Analysis of these reviews allows businesses to extract valuable insights that can improve service quality and build a positive reputation [10].

In line with technological developments, the use of machine learning in analyzing sentiment from customer review data has become an increasingly in-depth research topic. This method involves machine learning algorithms to classify customer reviews as positive, negative, or neutral. Research has shown that this technique can provide sentiment analysis results with a high degree of accuracy, allowing companies to quickly respond to customer feedback. Linchi Kwok has conducted research analyzing the 7Ps by applying the SVM machine learning algorithm to hotel customer review data sourced from Air BnB. Classification results reaching more than 80% are the best results [11]. Therefore, opportunities for improvement can still be made by applying better machine learning algorithms such as ensemble learning. The use of ensemble learning has been proven to increase the accuracy of machine learning and is implemented in many areas such as health [12], agriculture [13] and environmental [14], electrical [15], education [16], and many others.

Research in sentiment analysis of customer review has been conducted; however, literature gap in evaluation of machine learning performance remains, especially that of ensemble learning involving 7P classification. Deep research in classification of customer review into 7Ps using ensemble machine model is still limited. Thus, this research proposes evaluation of ensemble learning performance in 7P classification for hospitality and restaurant. The significant novelty of this research is implementation of ensemble learning to improve 7P classification of customer review. Ensemble learning combines some of the machine learning models to increase accuracy and overall performance. Then, it can be used to predict 7P marketing aspects accurately and deeply.

2. Method

The methodology employed in this research involves merging NLP (Natural Language Processing) with ensemble learning. NLP is utilized in processing customer restaurant reviews data, while ensemble learning is employed for learning and achieving its accuracy. Figure 1 illustrates the steps employed in this research.

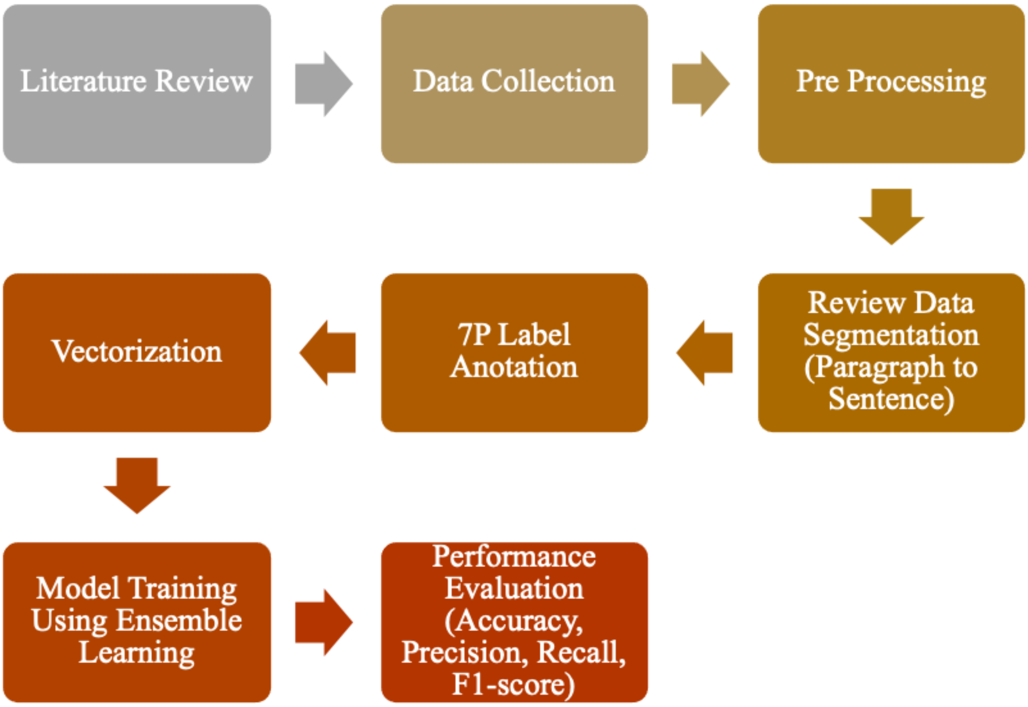

Fig. 1. Research step

2.1. Literatur Review

Table 1 displays comparable research carried out by other scholars. The studies pertain to reviewing data mining in the realms of management and marketing. Among these studies, merely two were dedicated to exploring the intricacies of the marketing mix. The prevalent classification algorithms predominantly remain single-method classifiers, signifying the absence of research centered on mining the marketing mix through ensemble methods. This study endeavors to bridge this void and offer novel perspectives on employing ensemble learning within the sphere of management and marketing.

Table 1. Literature review of previews related works

Researcher	Research Object	Machine Learning Research Method	Evaluation Performance
Linchi Kwok et.al [11]	Mining 7P mix marketing for hotel review from Air BNB	SVM	F1-Score
Mingming Cheng et.al [17]	The sentiment analysis of 7P mix marketing on AirBNB hotel review	-	Positif and Negative sentiment
Junegak Joung et.al. [18]	New product development based on customer review segmentation	Decision Tree, Random Forest, LGBM, xgboost, cat-boost, and ANN	F1-score

Ali Ahani et.al. [19]	Segmentation of SPA costumer based on the review on TripAdvisor	SOM Clustering, CART	Cluster Coefficient Determinant
Bee Shin et.al. [20]	The sentiment analysis of Restaurant based on Google Review	-	Positif and Negative sentiment
Aksh Patel et.al. [21]	The sentiment analysis of Airlines Feedback	Logistic Regression, KNN, Naïve Bayes, SVM, Decision Tree, Random Forest, Adaboost, BERT	Accuracy, Precision, Recall, F1-Score
Ziedhan Alifio Diekson et.al. [22]	The sentiment analysis of Traveloka apps	Logistic Regresssion, Naïve Bayes, and SVM	Precision, Recall, F1-Score

2.2. Dataset

The dataset used originates from Tripadvisor's New York hotel reviews and contains restaurant review data which obtained from data.world website [23]. It comprises 3716 paragraphs representing restaurant reviews, yet without the 7Ps label, despite containing ratings. Hence, there's a need to add the 7Ps annotations to these reviews.

2.3. Pre-processing

In this phase, an initial processing step is implemented to acquire a refined review dataset, removing symbols, extra spaces, and most punctuation except periods and commas. It also involves the removal of meaningless sentences resulting from incomplete sentence beginnings due to imperfect dataset crawling.

2.4. Review Data Segmentation

At this phase, data review segmentation aims to break down the original paragraph-shaped review dataset into separate sentences, utilizing the period (.) as a delimiter. Every resulting sentence retains its initial associated information, including review timing, reviewer identity, initial rating, and more.

2.5. Annotation

In the annotation process, each sentence derived from the paragraph division undergoes annotation. To facilitate this, we compiled a keyword list for each P category, following a method akin to that of Linchi Kwok [11]. Presented below is a table listing these keywords, serving as a reference for annotation, along with accompanying examples in Table 2.

Table 2. List of keywords and its example

No	Categories	Keywords	Review Examples
1	Product	name of food, menu, list of menus, drink, etc.	*My friend and I had the lychee mocktails, the pork bun, pork dumplings and chicken soy ramen and everything was great and perfect.*
2	Process	service, waiting time, fast, reservation, booked/booking, take out, warm.	*Service is very professional.* *The moment we came in, we were greeted warmly.*
3	Physical Evidence	the view, interior, atmosphere, ambience, music, table, cozy, little, decor, review, space, crowded.	*Great place, good atmosphere.* *It is a large space.*
4	People	name of staff, staff, owner, the management.	*The staff were so friendly.* *Also, [name of someone] is the best.*
5	Promotion	recommend, recommended, will come, worth visit, visit, return, gem.	*I totally recommend.* *Gem of a find.*
6	Place	Location, street, name of place.	*Its on a bustling corner, near old Village streets.* *location on 2nd avenue and 82nd street.*
7	Price	pricey, price, $, USD, budget, expensive.	*Reasonable pricing.* *Lunch was under $30.*
8	Traveler	everything that not related to the 7P's above.	*I come with my family, my friends, my husband, my wife. First time visit. birthday. anniversary.*

2.6. Vectorization

The vectorization stage involves preparing the dataset for classification by employing the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm. This algorithm calculates the importance of each word within the review sentences. TF measures a word's frequency within a sentence, while IDF evaluates its occurrence across similar sets of review sentences. The TF-IDF formulation, expressed in formula (1), includes parameters such as N, representing the total document count, DF indicating the count of documents containing a specific word, and IDF as the inverse of DF [17].

$$TF - IDF = TF \ x \log(\frac{N}{DF}) \tag{1}$$

2.7. Ensemble Learning

Ensemble learning, a form of machine learning, integrates various fundamental methods during its training. This method proves potent in enhancing system performance by leveraging the strengths of these basic methods. There are several types of ensemble learning, including boosting, bagging, voting, stacking, and dynamic approaches. In this research, the dynamic ensemble learning method from the Deslib 0.4 library [18] will be utilized. This library encompasses various ensemble methods like dynamic classifier selection (DCS), dynamic ensemble selection (DES), and static methods. Further specifics about the applied dynamic ensemble method in this study are detailed in Table 3.

Table 3. The type of dynamic ensemble learning used in the research.

Type	Metode Name
Dynamic Classifier Selection (DCS)	Overall Local Accuracy (OLA)
	Local Class Accuracy (LCA)
	Multiple Classifier Behaviour (MCB)
Dynamic Ensemble Selection (DES)	k-Nearest Oracle Union (KNORA-U)
	k-Nearest Oracle-Eliminate (KNORA-E)
	Dynamic Ensemble Selection performance (DES-P)
	Dynamic Ensemble Selection-knn (DES-KNN)
	DES Multiclass Imbalance (DES-MI)

Overall Local Accuracy (OLA)

This method is part of dynamic ensemble learning which assesses the ability of each individual classifier, namely by selecting the classifier that is most competent in predicting labels for each test sample. This method will evaluate the accuracy of every single classifier based on its RoC. The single classifier with the highest competency will be selected and used by OLA as a predictor for each subsequent test sample [18].

Local Class Accuracy (LCA)

This method, in its process, will involve every single method to make predictions on new test data and then save the predicted class. The accuracy for each prediction in k neighboring areas will be evaluated. Then the LCA method will only select and return prediction results from the model that has the best capabilities for each subsequent new test sample [18].

Multiple Classifier Behaviour (MCB)

This method selects a basic method based on an assessment of how competent an individual classifier is based on its accuracy in its specific competency area. k-Nearest Neighbors (k-NN) and Behavioral Knowledge Space (BKS) are generally used as choices for defining competency areas. MCB attempts to select the most capable classifier for each test sample by considering its accuracy within this local area of competence [18].

k-Nearest Oracle Union (KNORA-U)

This dynamic ensemble selection method estimates the likelihood of a sample using the k-Nearest Neighbors (k-NN) algorithm. It picks the most capable classifier for each test sample by assessing its accuracy within its specific competence area. KNORA-U seeks to boost the ensemble classifier's efficiency by dynamically picking the best-suited base classifier for each test sample, ultimately enhancing the overall classification accuracy [19].

k-Nearest Oracle-Eliminate (KNORA-E)

This dynamic ensemble selection method aims to find a local Oracle, a base classifier that correctly classifies all samples within a test sample's competence region. It selects all classifiers that perform perfectly within this area as the local Oracle. If no such classifiers exist, KNORA-E eliminates the sample farthest from the test sample within its competence region and repeats the process [19].

Dynamic Ensemble Selection-Performance (DES-P)

It serves as an assessment of the base classifiers' capability within dynamic ensemble selection. Its role involves evaluating each classifier's effectiveness and determining the ideal choice for predicting specific test samples. DES-P fundamentally aims to appraise how well base classifiers perform on test data and opt for the most proficient one for predictions. Enabled by a dynamic ensemble selection framework, this method allows adaptability to distinct input data traits, thereby boosting performance in classification tasks. DES-P operates within DESlib, a Python library dedicated to dynamic classifiers and advanced ensemble selection techniques [18]. Essentially, DES-P stands as a performance measure within dynamic ensemble selection, evaluating base classifier competency and choosing the best fit for predicting test samples. This strategy has shown superiority over static ensemble selection approaches and monolithic classifiers across diverse classification tasks [19].

Dynamic Ensemble Selection-KNN (DES-KNN)

A method that utilizes the K-Nearest Neighbors (KNN) algorithm in a dynamic ensemble selection framework. The KNN algorithm is used to estimate the competence of each classifier in the ensemble, and the most competent classifier is selected to make predictions on a given test sample. This approach allows the ensemble to adapt to the specific characteristics of the input data, resulting in improved performance in classification tasks [20]. The DESlib library provides an implementation of the KNORA algorithm with dynamic ensemble selection, where each class can be used as a scikit-learn model directly. This library offers a suite of dynamic selection techniques, including DES-KNN, and allows scikit-learn's full suite of data preparation, modeling pipelines, and model evaluation techniques to be used directly [18]. The DES-KNN method is very useful for handling large-scale data sets and is insensitive to local data distribution, making it a valuable approach for dynamic ensemble selection in various real-world applications [21].

DES Multiclass Imbalance (DES-MI)

A method specifically designed to address the challenges of imbalanced datasets in the context of multiclass classification. The DES-MI method is a novel approach that utilizes dynamic ensemble selection to address the problem of imbalanced class distribution in a multiclass setting. The DES-MI method aims to improve the performance of classification models when faced with imbalanced data sets containing many classes. By dynamically selecting a set of classifiers based on the characteristics of the input data, DES-MI can effectively reduce the impact of class imbalance and improve the overall prediction accuracy for multiclass problems. Research and studies conducted on DES-MI have demonstrated its effectiveness in handling imbalanced multiclass datasets, making it a promising approach for real-world applications where imbalanced data and multiclass classification are prevalent [22].

Dynamic ensemble learning has parameter tuning options to get the best performance. Parameters that can be tuned include the value of the number of neighbors k to estimate the competency of the base classifier [22], the application of dynamic friend pruning (DFP) [23], the algorithm used to estimate the competency area (knn_classifier) [18], and the metrics used by knn_classifier classifier to estimate distance (knn_metric) [18]. In the Deslib 0.4 library, respectively, the k value is a positive integer value, the DFP value is True or False, the knn_clssifier options are 'knn', 'faiss', and None, and the knn_metric options are 'minkowski', 'cosine', ' mahalanobis'. In this study, tuning parameters were used as shown in table 4.

Table 4. Tuning parameter for dynamic ensemble learning

Name of parameter	Value
k neighbor	19
DFP	True
knn_classifier	‘knn’
Knn_metric	‘minkowski’

2.8. Single Classifier

A single classifier in this case is a machine learning method that stands alone in carrying out classification. Ensemble learning can work by combining several single classification methods simultaneously. Some basic classification methods commonly used by ensemble learning include logistic regression, SVM, KNN, Decision Tree, and so on. In this study, basic classification methods were used in the form of Logistic Regression, SGD Classifier, Ridge Classifier, Linear SVC, and SVC.

Logistic Regression

Logistic Regression is a method used to classify based on the probability value of an instance occurring. The probability value here is then used as a benchmark for entering each instance into a certain class. To obtain probability values, Logistic Regression uses the sigmoid function with a value range between 0 and 1. This method can be applied to binary or multiclass data so it is very suitable for the conditions of the restaurant review dataset with the 7P label [24].

Stochastic Gradient Descent (SGD) Classifier

The SGD method is an algorithm that is included in the optimization method category. However, in machine learning SGD can be applied to carry out classification for both binary and multiclass data. The way SGD works is by minimizing the objective function through iterative adjustments to model parameters based on randomly taken data samples. In each iteration SGD will stochastically update the parameters based on the gradient of the objective function. This allows the SGD method to increase its convergence speed and allows it to handle very large datasets such as customer review data. SGD has a parameter tuning option by providing a learning rate value with a certain value so that the model can be optimal [25].

Ridge Classifier

The Ridge method is a method that is very similar to logistic regression, but the addition of the L2 regularization component allows it to avoid overfitting results so that the stability of the model formed will increase. In the Ridge Classifier process there are two components to minimize the objective function. The first component is the log-loss function used in logistic regression and the second component is L2 regularization. Like SGD,,Ridge also has parameter tuning options. The Ridge tuning parameter is a regularization strength value (α) that controls how strong regularization will be applied to each training. The higher the α value, the stronger the regularization [26].

Linear SVC

Linear SVC is part of the Support Vector Machine (SVM) algorithm whose implementation is specifically for handling cases with Linear Decision Boundaries by utilizing a linear kernel. For large datasets with a high number of features, such as restaurant customer review data, Linear SVC computing will be more efficient. Another advantage of the Linear SVC method is that it uses the main formulation of the SVM objective function when solving problems with a high number of features and a large number of samples [24].

SVC

Like Linear SVC, SVC is also a part of the SVM algorithm. The difference is that the kernel applied to SVC is aimed at solving problems of both linear and non-linear relationships between features and labels. Apart from that, SVC is also suitable for solving problems with a smaller number of datasets but with higher feature dimensions. Both SVC and Linear SVC are suitable for application to datasets with binary and multiclass classes [24].

2.9. Performance Evaluation

The performance of classification results can be measured based on accuracy, precision, recall and f1-score values. These values can be obtained if we know the confusion matrix from the classification test results. For binary problems, the confusion matrix is shown in table 5. The values of TP, TN, FP, and FN can be obtained by formulas (2), (3), (4), and (5), successively.

Table 5. Confusion Matrix of binary class

	Predictive Positive	**Predictive Negative**
Actual Positif	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

$$TP = \frac{TP}{TP + TN} \tag{2}$$

$$TN = \frac{TN}{FP + TN} \tag{3}$$

$$FP = \frac{FP}{FP + TN} \tag{4}$$

$$FN = \frac{FN}{TP + FN} \tag{5}$$

The TP, TN, FP, and FN values are used to obtained accuracy, precision, recall, and f1-score. Those values can be obtained using formulas (6), (7), (8), and (9). We can see that the Recall value is the same as the True Positive (TP) value [27], [28]. Accuracy (Acc) measures how often the model correctly predicts the correct class. Precision (Pr) measures how accurate the model is in predicting positives. Recall (Rc) measures how well the model identifies all positive samples. F1 score provides an overview of the overall performance of the model. For cases where the dataset classes are balanced, the average value of accuracy, precision, recall, and f1-score can represent system performance. However, for the case of unbalanced classes, the weighted average value of accuracy, precision, recall and f1-score is used.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\mathrm{Pr} = \frac{TP}{TP + FP} \tag{7}$$

$$Rc = \frac{TP}{TP + FN} \tag{8}$$

$$F1\,score = 2 * \left(\frac{Pr * Rc}{Pr + Rc}\right) \tag{9}$$

3. Results and Discussion

In this research, the restaurant review dataset initially contained 3716 rows, which increased to 13265 rows after cleaning and segmentation into sentences. However, only 7165 rows were utilized for manual annotation, aligning with keywords specified in Table 6. The annotated dataset subset corresponding to these keywords is presented in Table 6.

Table 6. The original dataset resulting from manual annotation

index	review_text	label
0	*Two friends and I had lunch at the new uptown venue.*	traveler
1	*The food was absolutely delicious (we had three different pastas,perfectly prepared) and so was the wine.*	product
2	*The service was extremely attentive.*	process
3	*we ate outdoors in a well sheltered area and benefitted from the heaters.*	physicalevidence
4	*The staff were so friendly.*	People
…	…	…
7160	*and I approached the chef asking him to have a gluten free pasta with alfredo sauce.*	product
7161	*He confirmed my request and asked me to order it at the cassier.*	process
7162	*I did as his suggestion and ordered my.*	promotion
7163	*Every single time weve been here, it has blown us out of the water.*	traveler
7164	*Best pasta, best service, best appetizers, best price.*	product

We observed the distribution of each label. As depicted in Figure 2, the distribution ranges from the most abundant to the least in the following order: Product, Promotion, Traveler, Process, Physical Evidence, People, Place, and Price. The distribution of each label is uneven; the number of data instances with the 'Product' label is almost three times greater than the second highest, 'Promotion.' Meanwhile, the smallest datasets, 'Place' and 'Price,' are more than 15 times smaller in comparison. As for the distribution of 'Promotion,' 'Traveler,' 'Process,' and 'Physical Evidence,' the differences are not notably significant.

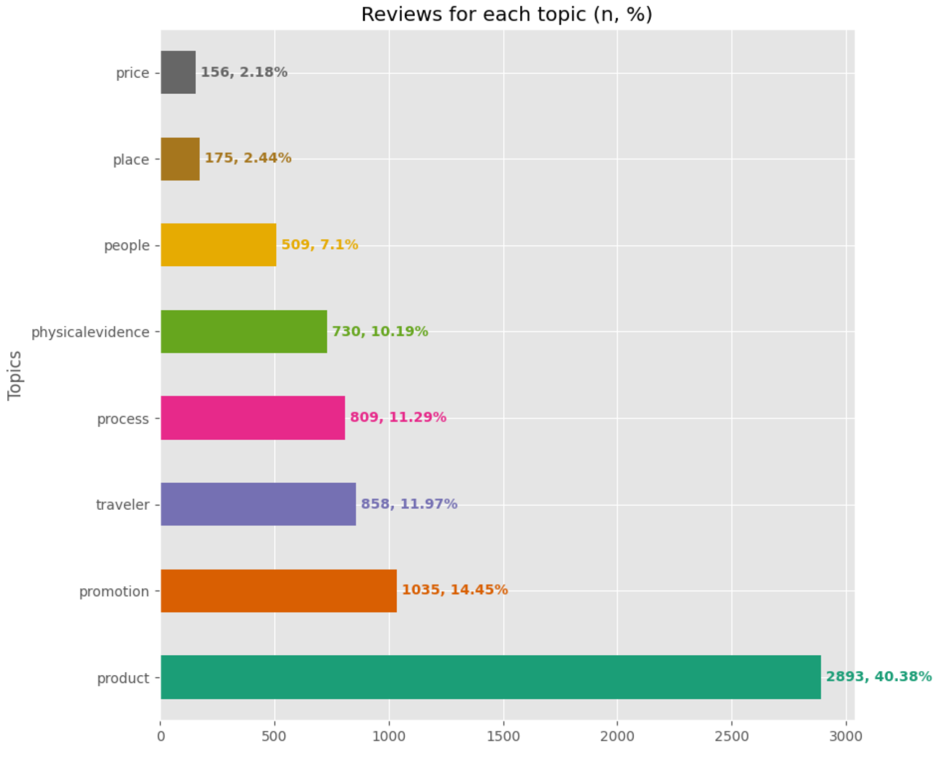

Fig. 2. Data distribution base on each label

3.1. The Best Ensemble Learning Model Performance

To pinpoint the optimal ensemble learning model for the mix marketing 7P classification, we scrutinized the performance metrics of each approach in figure 3. Among the ensemble methods evaluated, DESMI and KNORAU emerge with the highest classification accuracies, scoring 0.697 and 0.693, respectively. Metrics like precision, recall, and F1 scores bolster the effectiveness of these two methods. DESMI showcases well-rounded performance, boasting a precision score of 0.699, recall score of 0.697, and an F1 score of 0.684. This indicates its ability not only to predict positives accurately but also to minimize false positives and false negatives, rendering it a robust choice for the mix marketing 7P classification task. Following closely, KNORAU yields a precision score of 0.699, recall score of 0.693, and an F1 score of 0.679. While slightly lagging behind DESMI in accuracy, KNORAU demonstrates robust predictive capabilities. The selection between DESMI and KNORAU might hinge on specific business demands and the relative significance of precision and recall within the given context.

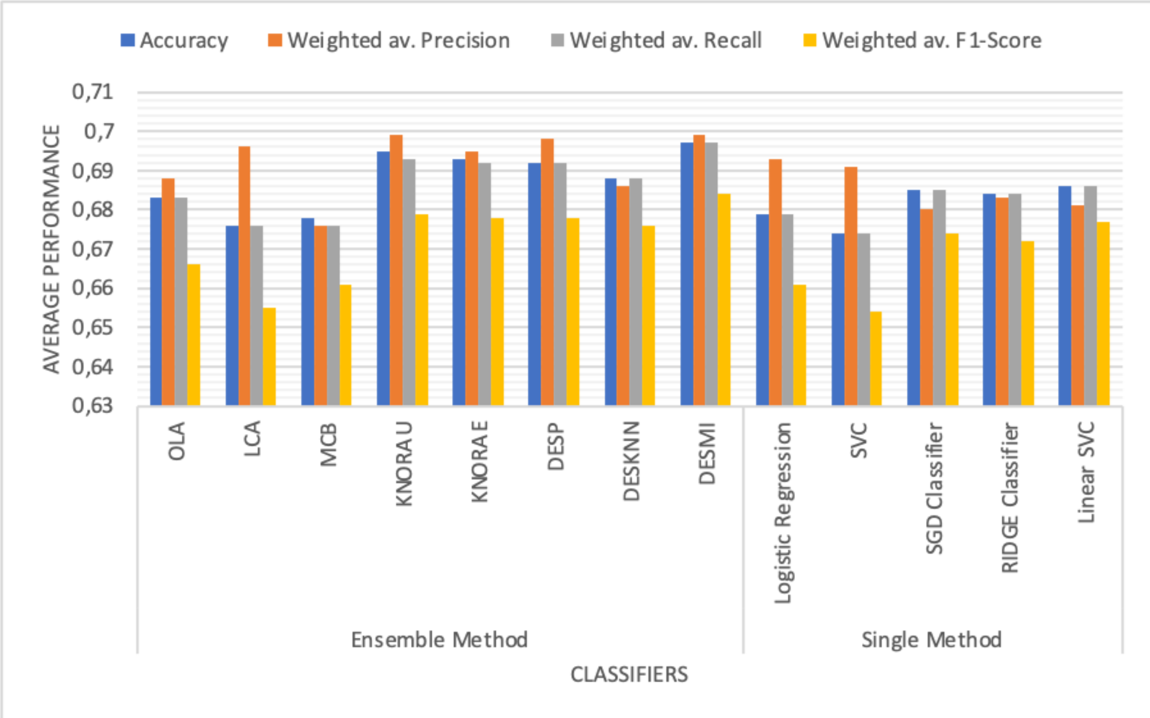

Fig. 3. Performance of the dynamic ensemble learning implementation on mining 7 P mix marketing

Judging from accuracy, precision, recall, and F1-score metrics, the use of dynamic ensemble learning has notably elevated the system's overall performance. In comparison to its basic (single method) counterparts, ensemble learning showcases superior performance across all evaluated metrics. Among the dynamic ensemble learning techniques employed in categorizing the 7P marketing mix based on restaurant customer reviews, DESMI emerges as the most effective. This is primarily attributed to the dataset's imbalanced nature concerning multiclass variables, where DESMI aptly resolves this challenge.

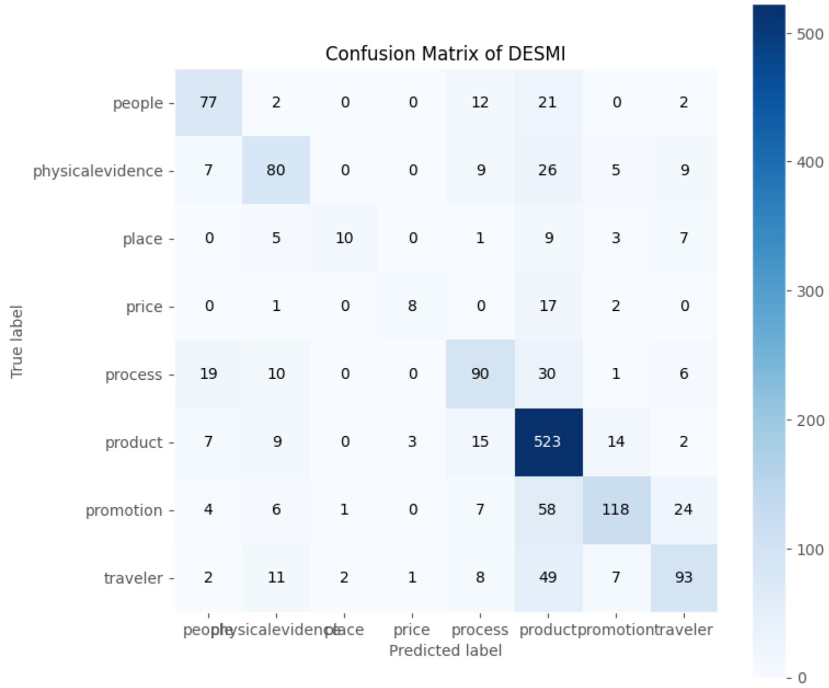

Fig. 4. Performance of the dynamic ensemble learning implementation on mining 7 P mix marketing based on confussion matrix

Upon scrutinizing the confusion matrix in figure 4, especially concerning DESMI as the best classification result, it becomes evident that several labels are frequently misclassified as 'product.' This stems from the manual labeling process based on predefined annotation keywords, uncovering numerous combined reviews encapsulated within a single sentence. Customers tend to amalgamate reviews encompassing various aspects of the 7P marketing mix into concise sentences, challenging the effectiveness of single manual annotations. Additionally, in Table 7, the unigram table per label category reveals certain words recurrent across multiple labels, such as 'recommend' and its derivatives. The potential appearance of similar words across various labels contributes to the system's relatively robust performance in extracting information about the 7P marketing mix from restaurant review data.

Table 7. The unigram table of the words represent the most informative features for prediction 7 Ps

Product	Process	Physical Evidence	People	Promotion	Place	Price	Traveler
menu	wait	space	owner	come	found	fair	anniversary
steak	attentive	seating	bartender	return	across	priced	night
chicken	impeccable	small	greg	again	blocks	worth	visited
recommend	minutes	place	waitress	would	central	money	lunch
mozzarella	greeted	music	nisar	back	walking	reasonably	dinner
pasta	table	beautiful	waiter	**recommended**	midtown	cheap	here
fresh	fast	clean	helpful	will	street	reasonable	friends
pizza	reservations	tables	server	definitely	park	expensive	went
delicious	reservation	cozy	friendly	highly	near	prices	birthday
food	service	atmosphere	staff	**recommend**	location	price	my

3.2. Ensemble Learning vs. Single Classifier Performance

The outcomes provided offer an insight into the effectiveness of ensemble learning versus individual classifiers for the mix marketing 7P classification task. Various ensemble methods like OLA (Online Bagging), LCA (Leveraging Cluster Analysis), MCB (Majority Class Boosting), DESMI (Dynamic-Ensemble Selection via Mutual Information), KNORAU (K-Nearest Oracles Union), KNORAE (K-Nearest Oracles Edition), DESP (Dynamic Ensemble Selection via Proximity), and DESKNN (Dynamic Ensemble Selection via K-Nearest Neighbors) were assessed based on classification accuracy, precision, recall, and F1 score. Contrasting these ensemble techniques against single classifiers—Logistic Regression, Support Vector Classifier (SVC), Stochastic Gradient Descent (SGD) Classifier, Ridge Classifier, and Linear SVC—it's evident that ensemble learning consistently outperforms individual classifiers in classification accuracy and other metrics. Ensembles collectively exhibit higher accuracy, precision, recall, and F1 score compared to any standalone classifier. This superiority stems from the diversity inherent in ensemble models, which contributes to enhanced generalization and robustness. By effectively capturing various facets of the intricate data relationships, ensemble methods generate a more comprehensive and precise predictive model.

3.3. Implications of Mix Marketing 7P Mining using Ensemble Learning

Utilizing ensemble learning to extract mix marketing 7P insights from restaurant reviews carries substantial implications for both academic exploration and practical implementation. The 7P mix encompasses Product, Price, Place, Promotion, People, Process, and Physical evidence, forming a comprehensive framework for understanding and optimizing restaurant marketing strategies. Ensemble learning emerges as a potent tool for navigating the intricacies of 7P classification in marketing. The amalgamation of multiple models within ensemble learning enhances the capacity to capture nuanced data aspects pertinent to each marketing element. This proves especially valuable in the realm of restaurant reviews, where diverse sentiments and opinions regarding various business aspects are expressed. Employing ensemble learning in mining 7P mix marketing facilitates a more nuanced and accurate comprehension of how customers perceive and evaluate different facets of their restaurant experience. For instance, it unveils patterns in customer sentiments regarding product quality, price competitiveness, promotional efficacy, staff impact, process efficiency, and physical attributes such as ambiance and cleanliness. Furthermore, the ensemble approach aids in averting overfitting risks and enhances the model's adaptability to novel, unseen data. This attribute is critical in establishing a sturdy and dependable predictive model, particularly in dynamic environments where customer preferences and market dynamics evolve continuously.

4. Conclusion

This study outlines the application of ensemble learning in classifying online text reviews based on the 7P marketing mix label. By utilizing labels that represent the 7P aspects, this research suggests a new perspective that ensemble learning outperforms its base methods. In our experiments, the DESMI method demonstrated the highest performance in terms of accuracy (0.697), precision (0.699), recall (0.697), and F1-score (0.684). The efficacy of ensemble learning in this scenario underscores its capacity to manage intricate and multifaceted datasets, positioning it as an asset for marketers and researchers aiming to derive significant insights from customer reviews and feedback within the domain of mix marketing 7P. During this experiment, we identified a significant pattern in the sequential data of online reviews. Subsequent research holds greater promise for business practitioners to extract valuable insights from the time-series data embedded in online text reviews.

Acknowledgment

This research was conducted with support from Ministry of Education and Culture of the Republic of Indonesia through the Basic Research scheme with contract number of 54.20.6/UN37/PPK.10/2023.

Declarations

Author contribution. The contribution or credit of the author must be stated in this section.
Funding statement. This research was conducted with support from DRTPM through the Basic Research scheme with contract number 54.20.6/UN37/PPK.10/2023.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] K. Jarek and G. Mazurek, “Marketing and artificial intelligence,” *Cent. Eur. Bus. Rev.*, vol. 8, no. 2, pp. 46–55, 2019.

[2] J. Paul and E. Mas, “Toward a 7-P framework for international marketing,” *J. Strateg. Mark.*, vol. 28, no. 8, pp. 681–701, 2020.

[3] P. Kotler and K. L. Keller, *Marketing Management*. Pearson, 2016.

[4] M. Mustak, J. Salminen, L. Plé, and J. Wirtz, “Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda,” *J. Bus. Res.*, vol. 124, no. January, pp. 389–404, 2021.

[5] S. Fan, R. Y. K. Lau, and J. L. Zhao, “Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix,” *Big Data Res.*, vol. 2, no. 1, pp. 28–32, 2015.

[6] M. Holmlund *et al.*, “Customer experience management in the age of big data analytics: A strategic framework,” *J. Bus. Res.*, vol. 116, no. February, pp. 356–365, 2020.

[7] K. K. Halim, S. Halim, and Felecia, “Business intelligence for designing restaurant marketing strategy: A case study,” *Procedia Comput. Sci.*, vol. 161, pp. 615–622, 2019.

[8] A. Rosário and R. Raimundo, “Consumer marketing strategy and e-commerce in the last decade: A literature review,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 7, pp. 3003–3024, 2021.

[9] M. Stone *et al.*, “Artificial intelligence (AI) in strategic marketing decision-making: a research agenda,” *Bottom Line*, vol. 33, no. 2, pp. 183–200, 2020.

[10] J. Saidali, H. Rahich, Y. Tabaa, and A. Medouri, “The combination between Big Data and Marketing Strategies to gain valuable Business Insights for better Production Success,” *Procedia Manuf.*, vol. 32, pp. 1017–1023, 2019.

[11] L. Kwok, Y. Tang, and B. Yu, “The 7 Ps marketing mix of home-sharing services: Mining travelers’ online reviews on Airbnb,” *Int. J. Hosp. Manag.*, vol. 90, no. August 2019, 2020.

[12] B. Sunarko *et al.*, “Penerapan Stacking Ensemble Learning untuk Klasifikasi Efek Kesehatan Akibat,” *Edu Komputika J.*, vol. 10, no. 1, pp. 55–63, 2023.

[13] M. Murad, S. Sukmawaty, A. Ansar, R. Sabani, and S. Hidayat, “Sistem Pendeteksi Kerusakan Buah Mangga Menggunakan Sensor Gas Dengan Metode DCS - LCA,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 186–194, 2021.

[14] F. D. Vidal Bezerra, F. Pinto Marinho, P. A. Costa Rocha, V. Oliveira Santos, J. Van Griensven Thé, and B. Gharabaghi, “Machine Learning Dynamic Ensemble Methods for Solar Irradiance and Wind Speed Predictions,” *Atmosphere (Basel).*, vol. 14, no. 11, p. 1635, 2023.

[15] Y. Yang, F. Jinfu, W. Zhongjie, Z. Zheng, and X. Yukun, “A dynamic ensemble method for residential short-term load forecasting,” *Alexandria Eng. J.*, vol. 63, pp. 75–88, 2023.

[16] R. Rismayati, I. Ismarmiaty, and S. Hidayat, “Ensemble Implementation for Predicting Student Graduation with Classification Algorithm,” *Int. J. Eng. Comput. Sci. Appl.*, vol. 1, no. 1, pp. 35–42, 2022.

[17] M. Cheng and X. Jin, “What do Airbnb users care about? An analysis of online review comments,” *Int. J. Hosp. Manag.*, vol. 76, no. April 2018, pp. 58–70, 2019.

[18] J. Joung and H. Kim, “Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews,” *Int. J. Inf. Manage.*, vol. 70, no. February, p. 102641, 2023.

[19] A. Ahani, M. Nilashi, O. Ibrahim, L. Sanzogni, and S. Weaven, “Market segmentation and travel choice prediction in Spa hotels through TripAdvisor’s online reviews,” *Int. J. Hosp. Manag.*, vol. 80, no. March 2018, pp. 52–77, 2019.

[20] B. Shin, S. Ryu, Y. Kim, and D. Kim, “Analysis on Review Data of Restaurants in Google Maps through Text Mining: Focusing on Sentiment Analysis,” *J. Multimed. Inf. Syst.*, vol. 9, no. 1, pp. 61–68, 2022.

[21] A. Patel, P. Oza, and S. Agrawal, “Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model,” *Procedia Comput. Sci.*, vol. 218, pp. 2459–2467, 2022.

[22] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, “Sentiment analysis for customer review: Case study of Traveloka,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 682–690, 2022.

[23] Crawl Feeds, "opensnippets/new-york-hotel-reviews-dataset-from-tripadvisor," *Data.World*, 2021. [Online]. Available: https://data.world/opensnippets/new-york-hotel-reviews-dataset-from-tripadvisor. [Accessed: 17-Oct-2023].

[24] E. M. Mercha and H. Benbrahim, "Machine learning and deep learning for sentiment analysis across languages: A survey," *Neurocomputing*, vol. 531, pp. 195–216, 2023.

[25] B. Gaye, D. Zhang, and A. Wulamu, "Sentiment classification for employees reviews using regression vectorstochastic gradient descent classifier (RV-SGDC)," *PeerJ Comput. Sci.*, vol. 7, pp. 1–27, 2021.

[26] A. A. Sayed, E. Elgeldawi, A. M. Zaki, and A. R. Galal, "Sentiment Analysis for Arabic Reviews using Machine Learning Classification Algorithms," in *Proceedings of 2020 International Conference on Innovative Trends in Communication and Computer Engineering, ITCE 2020*, 2020, no. February, pp. 56–63.

[27] S. García, Z. L. Zhang, A. Altalhi, S. Alshomrani, and F. Herrera, "Dynamic ensemble selection for multi-class imbalanced datasets," *Inf. Sci. (Ny).*, vol. 445–446, pp. 22–37, 2018.

[28] P. R. G. Cordeiro, G. D. C. Cavalcanti, and R. M. O. Cruz, "Dynamic Ensemble Algorithm Post-Selection Using Hardness-Aware Oracle," *IEEE Access*, vol. 11, no. July, pp. 86056–86070, 2023.