Enhancing Sentiment Analysis Accuracy in Borobudur Temple Visitor Reviews through Semi-Supervised Learning and SMOTE Upsampling

Candra Agustina [1,]*, Purwanto Purwanto [2], and Farikhin Farikhin [3]

[1] Doctoral Program of Information System, School of Postgraduate Studies,
Diponegoro University, Semarang, Indonesia
[2] Department of Chemical Engineering, Diponegoro University, Semarang, Indonesia
[3] Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
Email: candra.caa@bsi.ac.id (C.A.); purwanto@live.undip.ac.id (P.P.); farikhin.math.undip@gmail.com (F.F.)
*Corresponding author

***Abstract*—The level of visitor satisfaction with tourist destinations can be known from reviews on social media. One method used is to carry out sentiment analysis on comments given by visitors on social media or related websites. This study was envisioned as a preliminary phase to bolster subsequent research concerning tourist destination recommendation systems around Borobudur Temple. We conducted a sentiment analysis using a semi-supervised learning approach. Within this approach, the dataset was partitioned into labeled and unlabeled data. The labeled data served as a reference for the automatic labeling process, which utilized the Multinomial Naïve Bayes algorithm. Specifically, the objective was to extract sentiments from visitors to Borobudur Temple. These extracted sentiments will later be employed as a variable in subsequent research. Dataset preprocessing steps encompassed data cleaning, sentence segmentation, tokenization, and stop word removal. We observed that the difference in labeling outcomes between datasets trained without Synthetic Minority Oversampling Technique (SMOTE) Upsampling and those trained with SMOTE Upsampling was a mere 0.18%. The labeled data not only plays a pivotal role in model training but is also instrumental in evaluating the accuracy of the Multinomial Naïve Bayes algorithm. Crucially, after implementing the SMOTE Upsampling technique, our model exhibited a significant improvement, achieving an accuracy rate of 83.68%. This noteworthy enhancement represents a substantial increase from the initial accuracy rate of 60.59%. Our in-depth analysis underscores the superior performance achieved when the training data undergo the SMOTE Upsampling process, indicating the effectiveness of this approach in refining sentiment analysis outcomes for tourist reviews.**

***Keywords*—Synthetic Minority Oversampling Technique (SMOTE) Upsampling, analysis sentiment, tourism, semi-supervised learning**

Manuscript received October 25, 2023; revised December 13, 2023; accepted December 25, 2023; published April 9, 2024.

I. Introduction

Borobudur Temple, recognized as a UNESCO World Heritage site, stands as a prominent tourism icon in Indonesia. This was corroborated by Waruwu *et al.* [1], which indicated that Borobudur Temple is a more popular destination compared to Lake Toba, Labuan Bajo, Mandalika, and Likupang. Borobudur Temple attracts millions of visitors annually, showcasing its significance as a major tourist attraction [2]. Apart from that, this temple complex is not only a silent witness to past civilizations but also a reflection of the dynamics of the contemporary tourism industry in this country. In today’s digital era, where every travel experience can be easily shared via social media, reviews, and other platforms, Borobudur Temple’s prominence and significance are continuously reaffirmed and amplified by global travelers who recount their experiences and share their awe of its grandeur [3].

Through this approach, we can find out the aspects that are most liked, appreciated, or resonated with the visitors and, conversely, the areas that might require improvement or attention to enhance the overall visitor experience [4], as well as aspects that might require improvement or modification to enhance visitor satisfaction. Consequently, stakeholders—ranging from the government and temple management to local tourism industry can devise more tailored strategies and policies [5]. The significance of a sentiment analysis extends beyond merely enhancing service quality. By comprehending visitor sentiments, we can also uphold a positive reputation and image. Over the long run, a profound grasp of visitor perceptions paves the way for sustainable growth within the tourism industry [6]. A sentiment analysis of visitors to Borobudur Temple using data sources from Tripadvisor had already been carried out, but the data processed were in the form of English texts. Apart from that, the method used was supervised learning, where all the review data had labels [4]. Therefore, if the data being processed have not yet had a label, they must be labeled first. This certainly this takes more time, energy and costs. An effective way to handle this is to use semi-supervised learning techniques, where some data will be labeled manually. Most of the others will be labeled automatically. To be accurate, labeling is carried out by experts or systems of which reliability has been tested, because careless labeling will result in invalid pseudo labeling.

The objective of this research was to gauge visitor sentiments towards the Borobudur Temple and categorize them as Positive, Neutral, or Negative. Additionally, the findings of this study will be incorporated as variables in research pertaining to the tourist destination recommendation system surrounding Borobudur Temple.

Sentiment analysis has been extensively applied in the tourism sector as it is a pivotal component of utilizing big data in tourism research. This is because sentiment analysis can discern emotions expressed by tourists derived from their personal experiences [7]. Sentiment analysis can also discern both negative and positive opinions on a large scale [8]. Moud *et al.* [9] described the integration of sentiment analysis with semantic clustering to develop a recommendation system. Their research revealed that the proposed system outperformed in terms of F-Measure. Hence, sentiment analysis techniques are highly suitable for supporting sustainable tourism. Not only are they beneficial from the tourist's perspective, but they also offer significant value for business professionals, especially in marketing [7].

II. Literature Review

Sentiment analysis, also known as opinion mining, allows one to determine whether the author's perception is positive, neutral, or negative [10]. Nandwani [11] demonstrated that the application of a Lexicon-Based approach yields favorable results in analyzing sentiments expressed by opinion writers. The fundamental steps of sentiment analysis are:

1) Input Dataset

The dataset, sourced from social media or the web, consists of customer opinions.

2) Preprocessing

The process involves data cleaning, where data that might diminish the accuracy of results are removed using specific methods.

3) Feature Extraction

At this stage, the document is segmented into sentences, which are then further broken down into words for subsequent processing.

4) Model Development

The model development stage involves selecting an appropriate algorithm based on the nature and structure of the existing data. After selecting the algorithm, the processed data are fed into the model for training. During the training process, the algorithm parameters are adjusted to get the best performance from the model. Once the model is trained with the training data, it is then tested using a test data set to assess its accuracy and effectiveness. The result of this stage is a sentiment analysis model that is ready to be used to interpret and classify new sentiments from data that have never been seen previously [12].

5) Model Assessment

The model assessment phase involves evaluating the performance of the utilized model.

According to Rintyarna *et al.* [13] in their research, Sentiment Analysis can be employed to:

a) Investigate customer perceptions concerning service quality.
b) Provide a framework to model service quality assessments.
c) Evaluate the accuracy or performance level of sentiment analysis with the proposed model.

When conducting sentiment analysis, researchers can employ the following algorithms:

A. Naïve Bayes and Support Vector Machine Algorithm

Watori *et al.* [14] carried out sentiment analysis research, comparing the Naïve Bayes algorithm with SVM to gauge public perceptions about relocating the country's capital. The findings revealed that the Naïve Bayes algorithm achieved an accuracy of 78.39%, while the SVM had an accuracy of 76.40% [15]. The same algorithm has also been demonstrated to effectively analyze the perceptions of mobile banking ID users in Malaysia.

B. Neural Network

In Apriliani *et al.*'s research [16] on hotel services in Indonesia, the Neural Network algorithm was utilized for sentiment analysis, achieving an accuracy rate of 88.99%.

C. Logistic Regression, Decision Tree, Maximum Entropy, K-Nearest Neighbours (KNN)

In his survey, Wankhade stated that algorithms such as Logistic Regression, Decision Tree, Maximum Entropy, and K-Nearest Neighbors (KNN) can be employed for sentiment analysis [17].

To determine the accuracy of the algorithm, one can use the following formula:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where:
TP (True Positives): Represents correct predictions.
FP (False Positives): Represents incorrect predictions.

In the realm of tourism, sentiment analysis stands as a pivotal element of Big Data technology. The most commonly employed analytical methods include Lexicon-Based and Machine Learning approaches. While its primary application has been in analyzing tourist attractions, hotels, and restaurants, the evolving landscape of sentiment analysis has expanded its use to gather opinions on various services, such as those offered by tour guides [8].

Ontology is a procedure used to detect whether a review from a respondent is a true opinion or a false opinion [18]. Albar *et al.* [19] categorized words commonly used in smartphone reviews. An ontology framework was developed using the Ontogen software. Through this approach, they identified several reviews that were not pertinent to the specific smartphone being discussed. Salaiwarakul [20] employed ontology to offer recommendations for historical tourist destinations tailored to elderly tourists.

D. Naïve Bayes Algorithm

According to Subashini *et al.* [21], the Naïve Bayes algorithm is the most frequently used algorithm in sentiment analysis. The underlying principle involves categorizing data into classes based on probability.

Advantages of using the Naïve Bayes Algorithm:

- It is a simple algorithm to understand and implement.
- It offers faster predictions for classifying data.
- It is suitable for small datasets.

Disadvantages of the Naïve Bayes Algorithm

- It faces the Zero Conditional Probability Problem, which can nullify emerging probabilities.
- It makes strong assumptions regarding the independence of feature classes, which can potentially decrease accuracy [22].

The formula can be expressed as follows [23]:

$$P(X \mid H) = \frac{P(X \mid H) \times P(H)}{P(X)} \tag{2}$$

where
X = Evidence
H = Hypothesis
$P(H|X)$ =Probability that Hypothesis H is True for Evidence X
$P(X|H)$ =Probability that Evidence X is True for Hypothesis H
$P(H)$ = Prior probability of Hypothesis H
$P(X)$ = Prior probability of Evidence X

1) Multinomial Naïve Bayes algorithm in text clustering

The Multinomial Naïve Bayes is a variant of the Naïve Bayes algorithm specifically designed for categorizing text and documents. Its working principle involves computing the probability that a given word belongs to a particular category [24].

$$P(c \mid d) \alpha P(c) \prod_{k \leq n_d} P(tk \mid c) \tag{3}$$

where
$P(c|d)$: Probability of document d being in class c
$P(c)$: Prior probability that the document falls into class c
$P(tk|c)$: How many t_k are in class c
$\{t_1, t_2, t_3, \ldots, t_n\}$: Tokens in document c

Advantages of Multinomial Naïve Bayes:

1. It is suitable for both continuous and discrete data.
2. It is capable of handling large datasets.
3. It is supporting multi-label data classification.
4. It is ideal for Natural Language Processing (NLP) models [25].

2) The role of ChatGPT in sentiment analysis

ChatGPT is a language model built on the transformer architecture and developed by OpenAI. It represents one iteration in the Generative Pre-trained Transformer (GPT) [26]. ChatGPT is designed to generate text responses in a conversational manner, exhibiting capabilities to understand and produce text with a quality akin to human-like conversation. The model is trained on vast amounts of text data, enabling it to grasp diverse topics, nuances, and linguistic styles. Consequently, it can respond to queries and tasks across a broad spectrum of knowledge domains [27]. From the initial release of GPT, OpenAI has introduced several iterations, with GPT-4 standing out as one of the most advanced. This model, having been trained on billions of words, is adept at understanding and generating text with human-like quality [28].

Research in Natural Language Processing (NLP) has experienced significant advancements with technologies like ChatGPT. Bender utilized the Generative Pre-trained Transformer (GPT) model, specifically the second and third iterations, as foundational elements in their training data. The inherent strengths of the GPT models, particularly in processing, enable researchers to achieve more precise and pertinent results. The deployment of GPT-2 and GPT-3 in such studies underscores the critical role of transformer-based models in analyzing and interpreting language data [29]. ChatGPT demonstrates its proficiency in extracting implicit meanings and connecting them to the provided information. Such a capability is vital, especially for applications demanding profound contextual understanding and drawing inferences from vast data or information.

One of ChatGPT's advantages in inference tasks may come from its training on very large and diverse datasets, which allows it to access extensive knowledge on a variety of topics and contexts. As a result, when given certain information or sentences, ChatGPT can use this knowledge to make logical guesses or conclusions.

In practice, ChatGPT's capabilities in inference can be utilized in various applications, ranging from recommendation systems, decision aids, to automatic data analysis. With further development and adaptation, the potential use of ChatGPT in tasks requiring deep inference capabilities will continue to grow [30].

One notable advantage of ChatGPT in inference tasks stems from its training on expansive and diverse datasets. This foundational training equips it with a breadth of knowledge across various topics and contexts. Consequently, when presented with specific details or sentences, ChatGPT can draw upon this vast reservoir of knowledge to formulate logical deductions or conclusions.

In separate research, Julianto [31] employed ChatGPT during the text preprocessing phase of sentiment analysis. Given that preprocessing plays a pivotal role in sentiment analysis—with the integrity of input data profoundly influencing the resultant analysis—the decision to deploy ChatGPT in this phase is crucial. Julianto harnessed ChatGPT for a variety of preprocessing activities, such as data cleaning, labeling, and text normalization. With ChatGPT's adeptness at grasping context and the nuances of language, preprocessing becomes markedly more efficient and precise. For instance, in data cleaning tasks, ChatGPT effectively identifies and excises extraneous elements like URLs, HTML tags, and non-alphabetic symbols, while astutely preserving sentiment-pertinent keywords. Julianto's findings revealed that incorporating ChatGPT during the preprocessing phase augmented the accuracy of sentiment analysis by 1.76%, compared to traditional preprocessing techniques. This observation underscores the value advanced generative language models like ChatGPT bring to sentiment analysis. Nevertheless, Julianto also stressed the importance of ongoing validation. Even with ChatGPT's commendable performance, there remain specific scenarios where human oversight or a rule-driven strategy might be indispensable for ensuring the utmost preprocessing quality [31].

3) SMOTE Upsampling in handling imbalanced data

The Synthetic Minority Over-sampling Technique (SMOTE) was specifically designed to address the issue of class imbalance in datasets. In numerous machine learning contexts, there is often an imbalance where one class, typically the minority class, has significantly fewer samples compared to other classes. This disparity can lead to suboptimal model performance for the minority class, as the model tends to be biased towards the majority class [32]. SMOTE creates synthetic samples from minority classes [33]. SMOTE selects (*k*) nearest neighbor samples. For each of these neighbors, the difference between the feature of the sample and that of the neighbor is computed and then multiplied by a random number between 0 and 1. This result is added to the sample feature to produce a synthetic sample [34]. In several studies, this technique has been proven to be able to be used with Support Vector Machine and Naive Bayes and proven to increase accuracy in the sentiment analysis process.

III. MATERIALS AND METHODS

A. Research Stages

The research conducted fell under Semi-Supervised Learning, where the data were divided into two groups: labeled data and unlabeled data [35, 36]. Labeled data were used as a basis for the algorithm to label groups of unlabeled data. Semi-Supervised Learning addresses the challenge of labeling large datasets, as the process is time-consuming and costly due to the need for expert input [37]. This approach is called Self-Training, and it often yields satisfactory results [32].

Fig. 1 shows the data labeling process. It explains the processing of labeled data to be used as a trainer for unlabelled data to produce pseudolabels. After that, the results of the labeling were combined with the labeled data to form a new dataset, as shown in Fig. 2.

The new resulting dataset was then resampled using SMOTE twice to equalize the number of samples based on labels. It was then processed using Naive Bayes, so the tool displayed the level of accuracy of the model in making predictions.

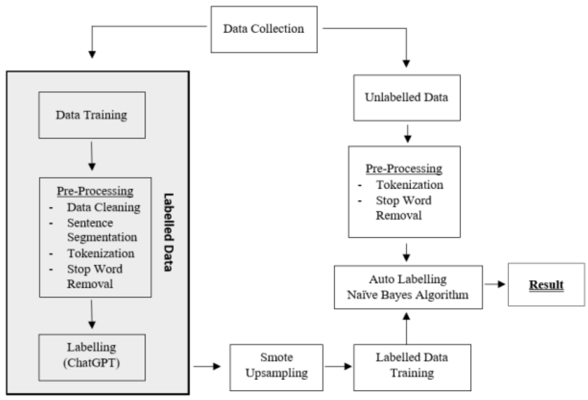

Fig. 1. Process of labeling Borobudur temple reviews.

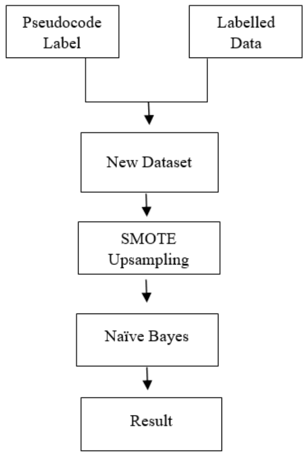

Fig. 2. Process of calculating performance.

B. Datasets

The dataset was sourced from the website (https://www.tripadvisor.co.id/Attraction_Review-g790291-d320054-Reviews-Borobudur_Temple-Borobudur_Magelang_Central_Java_Java.html) and consisted of 2,034 visitor reviews for Borobudur Temple. It was taken manually in September 2023.

The dataset was then bifurcated into two categories: training data and unlabeled data. The latter was automatically assigned a label in the subsequent steps.

C. Labeled Data/Training Data

Training data comprise reviews labeled as positive, negative, or neutral. Their primary function is to train the machine learning model to discern and predict sentiments with high precision for unlabeled data. Through these labeled training data, machine learning algorithms acquire the unique attributes and features of each sentiment, applying this understanding when assessing new data. Ensuring the training data are of high quality and are representative is pivotal, as it guarantees that the resulting model can adeptly generalize across diverse text types and scenarios.

From Fig. 3, we can see that the words ‘candi’ and ‘budha’ are the most frequently used by netizens, followed by the words ‘visit’ and ‘Yogyakarta’.

Fig. 3. Word cloud training data.

D. Unlabeled Data

Unlabeled data are a hallmark of semi-supervised learning. Labels will be assigned to these unlabeled data based on the learning outcomes of the chosen algorithm [36, 38].

E. Data Cleaning

Manual data cleaning is not recommended for big data processing because is really time-consuming [39]. The primary step in data cleaning involves removing stopwords. The list of Indonesian stopwords can be accessed at Indonesian Stoplist [40].

F. Sentence Segmentation

It involves breaking down a paragraph into individual sentences for subsequent processing [41, 42]. Segmentation is carried out before the tokenization process.

G. Tokenization

In its initial stage, text data are generally just a collection of characters. All procedures in text analysis depend on the words present in the dataset to ensure they are processed appropriately [11, 43].

H. Labeling of Training Data

Labeled training data serve as the ground truth used to train the model [44]. Typically, manual labeling is carried out by trained experts or graders to guarantee label accuracy and consistency [45]. However, this approach becomes impractical when dealing with large-scale datasets [46]. The labeling process leverages ChatGPT due to its proficiency in understanding and processing human language swiftly and accurately. By harnessing ChatGPT’s Natural Language Processing (NLP) capabilities, the labeling process not only becomes more efficient but also manages large volumes of data with greater speed [47].

I. SMOTE to Handle Imbalanced Data

The Synthetic Minority Oversampling Technique (SMOTE) is a preprocessing method introduced in 2002 that has since become the standard for addressing imbalanced data problems. Renowned for its effectiveness, SMOTE enhances model performance across numerous applications and domains. With time, this technique not only has resolved the data imbalance issue but also has spurred the development of innovative approaches in machine learning, such as multi-label classification and semi-supervised learning. Despite its age, SMOTE continues to be highly relevant and widely employed, serving as a point of reference in various studies concerning imbalanced data [48].

J. Data Processing

Data processing is conducted using the Multinomial Naïve Bayes algorithm. The labeled training data serve as a model, guiding the machine learning process to assign labels to the unlabeled datasets [49].

K. New Dataset

New dataset is a combination of labeled dataset and pseudolabel. Merging was done using Microsoft Excel. There was a total of 3765 sample data, all of which were labeled.

L. Multinomial Naïve Bayes Performance Analysis

To evaluate the model’s performance, metrics such as accuracy, precision, recall, and F1-score are used. Accuracy indicates the frequency with which the model correctly classifies the data. Precision, on the other hand, measures the proportion of instances predicted as positive by the model that are indeed positive [50]. Meanwhile, recall quantifies the proportion of actual positive cases that are correctly identified by the model. The F1-score is the harmonic mean of precision and recall, giving a balanced measure of the model’s overall performance [51].

In this analysis, comparing the results of manual labeling with the model’s predictions assists in gauging the model’s performance against human judgment. Discrepancies or inaccuracies in the predictions can be further analyzed to ascertain their root causes, such as overlooked features or gaps in the training data.

IV. RESULT AND DISCUSSION

The 2034 datasets, sourced from the Trip Advisor website, were formatted in Excel. These datasets were subsequently divided into training and testing data, facilitating automatic labeling by the system. From this collection, 300 reviews were selected as the training data. Subsequently, sentence segmentation and labeling were executed using ChatGPT, which transformed the 300 reviews into 973 labeled sentences. The breakdown is as follows:

TABLE I. RESULT OF LABELING USING CHATGPT

Labels	Total
Negative	135
Positive	570
Neutral	268

Based on Table I, one can observe the results of the labeling process. There was a marked disparity between the counts of negative and positive labels. Subsequent to this, testing was initiated to ascertain the accuracy of the Multinomial Naïve Bayes Algorithm in its predictive capabilities. Next, the accuracy of the labeled data were measured to get the best training data. The tool used was rapid miner. The process was dataset, then nominal to text, Process Document, and cross validation. The calculation process above had a sub-process, namely Process Document. Document processing had four processes, namely, tokenize, transfer cases, filter tokens so the text data were more easily processed by the system. The final subprocess was cross validation, in which the algorithm used was placed in followed by applying the model and using performance to see its accuracy. The initial test was conducted using the unmodified dataset. Subsequently, SMOTE was employed with automatic detection of the minority class. Consequently, the tool identified the negative class and increased its count from 135 to 570. The following test adjusted the minority class to 'Neutral', raising the total number of 'Neutral' label samples from 268 to 570. For the concluding test, samples labeled as 'Positive', 'Negative', and 'Neutral' were all equalized at 570 each. The results of these tests, which showcased the accuracy of the Multinomial Naïve Bayes Algorithm on the training data, can be observed in Table II:

TABLE II. ACCURACY OF LABELLED DATA

Testing	Accuracy
Labelled Data without SMOTE	60.95%
SMOTE Minority Class—Negative	77.13%
SMOTE Minority Class—Neutral	73.18%
SMOTE Minority Class—Negative & Neutral	83.51%

The process of measuring the accuracy of the labeled datasets without SMOTE Upsampling used a dataset with 135 negative, 570 positive, and 268 neutral labels, resulting in 60.95%. Then SMOTE Upsampling was applied, to detect samples with the smallest number, then the number was doubled until it reaching 570. Therefore, the resulting labeled data had 570 negative labels, 570 positive labels, and 268 neutral labels. The results of the data processing with this model generated accuracy of 77.13%. Next, SMOTE was set with the minority class being Neutral, so the total samples processed consisted of 135 negative, 570 positive and 570 neutral labels, resulting in 73.18%. Finally, the model used SMOTE Upsampling twice, namely 1 system detectsed negatives as a minority and the next detected neutral labels, so the resulting data consist of 570 negatives, 570 neutrals and 570 positives. It turned out that the accuracy reached 83.51%.

The following Fig. 4 depict this result:

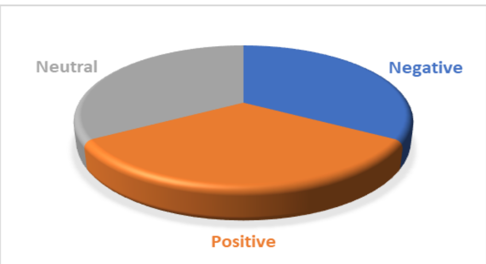

Fig. 4. Composition of labeled data after SMOTE Upsampling.

Therefore, the sample used the final model, then saved in Excel to be used as the training data. These training data were then used as a reference to provide pseudolabels to the unlabeled datasets.

The unlabeled data underwent processing twice: first without SMOTE Upsampling, and second, with the Negative & Neutral SMOTE Upsampling. The distinctions in the labeling outcomes between these two processes can be observed in Table III:

TABLE III. LABELLING OUTCOMES

Row Number	Labelled Data without SMOTE	SMOTE	Chat GPT
520	Negative	Positive	Positive
1032	Neutral	Positive	Positive
1528	Neutral	Positive	Positive
1704	Negative	Neutral	Positive

From Table III, it is evident that using SMOTE Upsampling successfully increased the correct predictions by 3 points compared to not using SMOTE Upsampling.

The results of the pseudolabel were then combined with the labeled data to see its accuracy for making predictions.

TABLE IV. CONFUSION MATRIX WITHOUT SMOTE UPSAMPLING

Labels	True Positive	True Neutral	True Negative	Class Precision
pred. Positive	1895	229	187	82.00%
pred. Neutral	443	208	54	29.50%
pred. Negative	438	79	142	21.55%
class recall	68.26%	40.31%	37.08%	

Based on Table IV, the results of the data processing showed a percentage of 61.09%, in which the highest corresponding prediction was in the group of positive labels. The new dataset consisted 383 negative labels, 516 neutral labels and 2,776 positive labels. We can see that there was a significant data imbalance, resulting in low accuracy. Therefore, the number of samples had to be equalized using the SMOTE Upsampling technique. After SMOTE had been carried out, the resulting confusion matrix is as Table V:

TABLE V. CONFUSION MATRIX WITH SMOTE UPSAMPLING

Labels	True Positive	True Neutral	True Negative	Class precision
pred. Positive	1894	63	17	95.95%
pred. Neutral	474	2432	33	82.75%
pred. Negative	408	281	2,726	79.82%
class recall	68.23%	87.61%	98.20%	

In Table V above, it can be seen that the accuracy obtained was 83.68%. It shows that using SMOTE Upsampling, the model performance increased by 22.6%. Table V also shows that the accuracy was lower compared to research by Singgalen, namely 96.36%. Nonetheless, these two researches used different focus and review data in which this research, processed reviews using Indonesian language. In fact, these results were still higher than Flores *et al.*'s research, with an accuracy of around 70% [52]. Future research needs to use other imbalanced data handling techniques such as ADASYN, Ensamble, etc., to obtain higher accuracy. Apart from that, it is also necessary to consider the selection of algorithms to be combined with these techniques.

V. Conclusion

The number of samples greatly affects the accuracy of predictions. In semi-supervised learning, SMOTE Upsampling can be employed twice, first during the preparation of training data, which can yield better quality training data, thus leading to more accurate pseudo labels..By employing the SMOTE Upsampling method, accuracy pseudo label can be enhanced by 0.18%. Overall, the model accuracy with the new dataset has an accuracy of 61.09%. After applying the SMOTE Upsampling method, accuracy increased by 22.6% to 83.68%. Thus, SMOTE Upsampling can be applied to semi-supervised learning to perform sentiment analysis, with good results. Semi-supervised learning will make it easier for researchers to obtain results without having to label all the data. Making it time and cost-saving. The results of this research can also be used to determine tourists' perceptions of tourist attractions. This is very useful for tourism business managers in developing their products.

The application of SMOTE in semi-supervised learning involves a lengthy process, starting with SMOTE implementation in the pseudolabeling phase and extending to prediction. This complexity can be considered a limitation of the research. The prolonged process demands significant computational resources, potentially hindering scalability and efficiency. Furthermore, the algorithm's complexity, sensitivity to parameters, and the potential for overfitting pose challenges. The intricate nature of the method may also affect the interpretability of the resulting model. Clear identification and communication of these limitations are crucial for providing context to readers in evaluating the research findings.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

In writing this paper, Candra Agustina was tasked with collecting and processing data, as well as compiling a draft paper. Purwanto collected references and literature reviews. Meanwhile, Farikhin reviewed the results of the data processing and the paper prior to submission. All authors had approved the final version.

Acknowledgment

I would like to express my heartfelt gratitude to Diponegoro University for providing me with an excellent academic environment and resources that have greatly contributed to the completion of the academic pursuit. My sincere appreciation extends to the esteemed faculty members whose guidance and support have been invaluable throughout my educational journey.

References

[1] D. Waruwu, I. W. R. Junaedi, R. T. P. B. Santoso, and I. G. B. R. Utama, "The gap between legal status and market potential of indonesia five super priority tourist destinations," *Revista De Gestão Social E Ambiental*, vol. 17, no. 5, Jul. 2023. doi: 10.24857/rgsa.v17n5-015

[2] R. Mustajab. (2022). The number of tourist visits to Borobudur Temple will increase in 2022. Data Indonesia. [Online]. Available: https://dataindonesia.id/pariwisata/detail/jumlah-kunjungan-wisatawan-ke-candi-borobudur-melonjak-pada-2022 (in Indonesian)

[3] Z. Lin and S. M. Rasoolimanesh, "Sharing tourism experiences in social media: A systematic review," *Anatolia*, 2022. doi: 10.1080/13032917.2022.2120029

[4] Y. A. Singgalen, "Analysis of tourist sentiment through borobudur temple review data on Tripadvisor using the Naïve Bayes classifier algorithm," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022. doi: 10.47065/bits.v4i3.2486 (in Indonesian)

[5] P. Sahid. Fauziah Eddyono. [Online]. Available: https://www.penerbituwais.com

[6] F. Borrajo-Millán, M. Alonso-Almeida, M. Escat-Cortes, and L. Yi, "Sentiment analysis to measure quality and build sustainability in tourism destinations," *Sustainability (Switzerland)*, vol. 13, no. 11, Jun. 2021. doi: 10.3390/su13116015

[7] H. Irawan, G. Akmalia, and R. A. Masrury, "Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling," *ACM International Conference Proceeding Series*, no. 1, pp. 7–12, 2019. doi: 10.1145/3361821.3361829

[8] M. Sigala, R. Rahimi, and M. Thelwall, *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications*, Springer, 2019. doi: 10.1007/978-981-13-6339-9

[9] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Syst. Appl.*, vol. 167, 114324, 2021. doi: 10.1016/j.eswa.2020.114324

[10] I. Onyenwe, S. Nwagbo, N. Mbeledogu, and E. Onyedinma, "The impact of political party/candidate on the election results from a sentiment analysis perspective using #AnambraDecides2017 tweets," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, 2020. doi: 10.1007/s13278-020-00667-2

[11] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–19, 2021. doi: 10.1007/s13278-021-00776-6

[12] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 104, 101822, Apr. 1, 2020. doi: 10.1016/j.artmed.2020.101822

[13] B. S. Rintyarna *et al.*, "Modelling service quality of internet service providers during COVID-19: The customer perspective based on twitter dataset," *Informatics*, vol. 9, no. 1, pp. 1–12, 2022. doi: 10.3390/informatics9010011

[14] N. A. Rahman, S. D. Idrus, and N. L. Adam, "Classification of customer feedbacks using sentiment analysis towards mobile banking applications," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 4, pp. 1579–1587, 2022. doi: 10.11591/ijai.v11.i4.pp1579-1587

[15] J. Watori, R. Aryanti, A. Junaidi, and A. Yani, "Use of classification algorithms for capital relocation sentiment analysis using automatic labeling," *Jurnal Informatika*, vol. 7, no. 1, pp. 85–90, 2020. doi: 10.31311/ji.v7i1.7528 (in Indonesian)

[16] D. Apriliani, T. Abidin, E. Sutanta, A. Hamzah, and O. Somantri, "Sentiment analysis for Indonesia hotel services review using optimized neural network," *J. Phys Conf. Ser.*, vol. 1538, no. 1, 2020. doi: 10.1088/1742-6596/1538/1/012060

[17] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, 2022. doi: 10.1007/s10462-022-10144-1

[18] D. U. Vidanagama, A. T. P. Silva, and A. S. Karunananda, "Ontology based sentiment analysis for fake review detection," *Expert Syst. Appl.*, vol. 206, no. June, 117869, 2022. doi: 10.1016/j.eswa.2022.117869

[19] A. S. Akbar, E. Sediyono, and O. D. Nurhayati, "Ontology-based sentiment analysis at the sentence level to measure product perception," *Jurnal Sistem Informasi Bisnis*, vol. 5, no. 2, pp. 84–97, 2015. doi: 10.21456/vol5iss2pp84-97 (in Indonesian)

[20] A. Salaiwarakul, "A historical tourism recommendation system for the elderly tourist using natural language processing and the ontology technique," *ICIC Express Letters*, vol. 16, no. 4, pp. 409–417, Apr. 2022. doi: 10.24507/icicel.16.04.409

[21] L. D. C. S. Subhashini, Y. Li, J. Zhang, A. S. Atukorale, and Y. Wu, "Mining and classifying customer reviews: A survey," *Artificial Intelligence Review*, vol. 54, no. 8, 2021. doi: 10.1007/s10462-021-09955-5

[22] P. Kaviani and S. Dhotre, "International journal of advance engineering and research short survey on Naive Bayes algorithm," *International Journal of Advance Engineering and Research Development*, vol. 4, no. 11, pp. 607–611, 2017.

[23] S. Raschka, *Naive Bayes and Text Classification I—Introduction and Theory*, 2014, pp. 1–20.

[24] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*, Cambridge: Cambridge University Press, 2021.

[25] Shriram. Multinomial Naive Bayes explained: Function, advantages and disadvantages, applications in 2022. [Online]. Available: https://www.upgrad.com/blog/multinomial-naive-bayes-explained/

[26] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, vol. 3. KeAi Communications Co., pp. 121–154, Jan. 1, 2023. doi: 10.1016/j.iotcps.2023.04.003

[27] OpenAI. ChatGPT. [Online]. Available: https://chat.openai.com/

[28] T. B. Brown *et al.* (2020). Language models are few-shot learners. [Online]. Available: https://commoncrawl.org/the-data/

[29] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proc. the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2021*, Association for Computing Machinery, Inc, Mar. 2021, pp. 610–623. doi: 10.1145/3442188.3445922

[30] Z. Qihuang *et al.*, "Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT," arXiv preprint, arXiv:2302.10198, 2023. doi: 10.48550/arXiv.2302.10198

[31] I. T. Julianto, D. Kurniadi, Y. Septiana, and A. Sutedi, "Alternative text pre-processing using ChatGPT open AI," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 1, pp. 67–77, Mar. 2023. doi: 10.23887/janapati.v12i1.59746

[32] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of Synthetic Minority Oversampling Technique (SMOTE) for imbalanced learning," *Mach. Learn.*, 2023. doi: 10.1007/s10994-022-06296-4

[33] D. V. Ramadhanti, R. Santoso, and T. Widiharih, "Comparison of smote and adasyn on imbalance data for classification of poor households in Temanggung district using the k-nearest neighbor algorithm," *Jurnal Gaussian*, vol. 11, no. 4, pp. 499–505, Feb. 2023. doi: 10.14710/j.gauss.11.4.499-505 (in Indonesian)

[34] Y. A. Singgalen, "Consumer sentiment analysis of food, services, and value in popular restaurants and eateries in Makassar city based on Tripadvisor recommendations Using the CRISP-DM and SERVQUAL methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, Mar. 2023. doi: 10.47065/bits.v4i4.3231 (in Indonesian)

[35] Y. Guo, H. X. Mao, J. Yin, and Z. H. Mao, "Gaussian transformation enhanced semi-supervised learning for sleep stage classification," *J. Big Data*, vol. 10, no. 1, Dec. 2023. doi: 10.1186/s40537-023-00758-9

[36] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, Feb. 2020. doi: 10.1007/s10994-019-05855-6

[37] Y. C A P. Reddy, P. Viswanath, and B. E. Reddy, "Semi-supervised learning: A brief review," *International Journal of Engineering & Technology*, vol. 7, no. 1.8, pp. 81–85, Feb. 2018. doi: 10.14419/ijet.v7i1.8.9977

[38] H. Zou and Z. Wang, "A semi-supervised short text sentiment classification method based on improved Bert model from unlabelled data," *J. Big Data*, vol. 10, no. 1, Dec. 2023. doi: 10.1186/s40537-023-00710-x

[39] F. Ridzuan and W. M. N. W. Zainon, "A review on data cleansing methods for big data," *Procedia Computer Science*, vol. 161, pp. 731–738, 2019. doi: 10.1016/j.procs.2019.11.177

[40] O. R. Hartono. Indonesian Stoplist. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist

[41] E. Saravia. Fundamentals of NLP-Chapter 1-Tokenization, lemmatization, stemming, and sentence segmentation. [Online]. Available: https://dair.ai/notebooks/nlp/2020/03/19/nlp_basics_tokenization_segmentation.html

[42] J. Petrus, Ermatita, Sukemi, and Erwin, "An adaptable sentence segmentation based on Indonesian rules," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1491–1499, Sep. 2023. doi: 10.11591/ijai.v12.i3.pp1491-1499

[43] V. Mohan, "Text mining: Open source Tokenization tools—An analysis," *Advanced Computational Intelligence: An International Journal (ACII)*, vol. 3, no. 1, pp. 37–47, Jan. 2016. doi: 10.5121/acii.2016.3104

[44] P. T. Ha, R. D'silva, E. Chen, M. Koyutürk, K. Koyutürk, and G. U. Karakurt, "Identification of intimate partner violence from free text descriptions in social media," *J. Comput. Soc. Sc.*, vol. 5, pp. 1207–1233, 2022. https://doi.org/10.1007/s42001-022-00166-8

[45] V. O. Tama, Y. Sibaroni, and Adiwijaya, "Labeling analysis in the classification of product review sentiments by using multinomial Naive Bayes algorithm," *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2019. doi: 10.1088/1742-6596/1192/1/012036

[46] A. Elyashar, J. Bendahan, and R. Puzis. (Aug. 2017). Has the online discussion been manipulated? Quantifying online discussion authenticity within online social media. [Online]. Available: http://arxiv.org/abs/1708.02763

[47] M. K. Mampuono. 10 benefits of ChatGPT that we need to know. [Online]. Available: https://bbpmpjateng.kemdikbud.go.id/10-manfaat-chatgpt-yang-kita-perlu-tahu/

[48] A. Fernández, S. García, F. Herrera, and N. V Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, 2018.

[49] M. F. Chen, B. Cohen-Wang, S. Mussmann, F. Sala, and C. Ré, "Comparing the value of labeled and unlabeled data in method-of-moments latent variable estimation," in *Proc. International Conference on Artificial Intelligence and Statistics*, 2021, vol. 130, pp. 3286–3294.

[50] D. J. Hand, P. Christen, and N. Kirielle, "F*: An interpretable transformation of the F-measure," *Mach. Learn.*, vol. 110, no. 3, pp. 451–456, Mar. 2021. doi: 10.1007/s10994-021-05964-1

[51] D. Chicco and G. Jurman, "The advantages of the Matthews Correlation Coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Jan. 2020. doi: 10.1186/s12864-019-6413-7

[52] A. C. Flores, R. I. Icoy, C. F. Peña, and K. D. Gorro, "An evaluation of SVM and Naive Bayes with SMOTE on sentiment analysis data set," in *Proc. 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, 2018. doi: 10.1109/ICEAST.2018.8434401