Proceeding of the 7th TICC International Conference 2023
Toward Sustainable Development Goals: Transformation and Beyond
February 4-5th, 2023, Chiang Mai, Thailand

Sentiment Analysis on Social Networking: A Case Study of Chiang Mai Tourism

Sutheera Munsaen[1], Chat Chuchuen[2] and Jakkapong Sukphan[3]

Faculty of Business Administration, Maejo University, Chiang Mai, 50290, Thailand[1]
E-mail: thitinun@mju.ac.th
Faculty of Business Administration, Maejo University, Chiang Mai, 50290, Thailand[2]
E-mail: chat@mju.ac.th
Faculty of Business Administration, Maejo University, Chiang Mai, 50290, Thailand[3]
E-mail: jakkapong@mju.ac.th

Abstract

With the rise of the social networking epoch, there has been a surge of user-generated content. Microblogging sites have millions of people sharing their thoughts daily because of their characteristic short and straightforward manner of expression. We propose an investigate a paradigm to the sentiment from a popular real-time microblogging service, Twitter, where users post real-time reactions and opinions on Chiang Mai tourism through retweets or hashtags Chiang Mai review.

In this paper, we expound on a hybrid approach using the accuracy of three machine learning algorithms (Support Vector Machines, Naïve Bayes, and K-Nearest Neighbor) in predicting the sentiments of the retweets or hashtags in Chiang Mai review were investigated. The results found that the support vector machine provided the best results, with a maximum accuracy of 97.50 % on both positive and negative comments.

Subsequently, the comments were gathered to provide customer insights and suggestions to help increase tourism in Chiangmai. The results of this study suggest that it can help entrepreneurs to develop service quality and marketing strategies for customers in their tourist destinations, tourist attractions, restaurants, and nightlife.

KEYWORDS: Sentiment analysis, Social networking, Twitter, Chiang Mai tourism

1 INTRODUCTION

The widely of social media is bringing an increase in user-generated content (UGC). In this context, sentiment analysis is making growth in marketing and finance. It is effective computing research and is widely recognized as a Natural Language Processing (NLP), which broadly analyses users' opinions, and reviews (Poria et al., 2017). The sentiment analysis aims to evaluate a textual item expressed in a positive or negative opinion in general terms or any given entity (Nakov et al., 2016). The sentiment classification is well-defined words and phrases with a definitive score. Indeed, research has developed algorithms and methods to automatically detect the underlying text (Sun et al., 2017). Many companies deploy it to make better decisions and understand customer behavior and thoughts about their company or products. While several studies have examined the reactions toward products or companies (Paolanti et al., 2017).

As tourists have become more experienced and digitally literate, they have gained access to information and increased power of choosing by utilizing ICT, especially in Thailand (Valdivia Pérez et al., 2009). As a result, the Internet, especially social media, has changed how

people travel. People often seek travel experiences online during which the undertaking of the travel and the post-travel phase, during which they often share their experiences and journeys online through social media, which may cause others to make decisions to travel (Flores-Ruiz et al., 2021). Accordingly, social media has become the most popular source of inspiration for travelers searching for leisure and entertainment opportunities and new tourist destinations, as the tourism industry is one of the highest-income industries in Thailand. Foreign travelers and Thai people also enjoy traveling to Chiang Mai, the most popular destination due to being a significant hub in the Northern part of Thailand (Tourism Authority of Thailand, 2022). According to the sustainable tourism development strategies, one of the main objectives is to minimize the negative social and ecological impacts, improving visitors' experience and quality of life (Gursoy et al., 2019). Despite the potential of sentiment analysis, exploiting the massive reviews and posts by tourism, this is limited since the studies are costly and time-consuming if the entrepreneurs collect the data manually. Therefore, comprehensive extraction will receive practical knowledge from sentiment analysis.

Social media platforms, especially Facebook, Twitter, Instagram, and, Tiktok have overgrown in recent years, providing information for analyzing social trends and opinions. Social media platforms also benefit travelers who wish to understand the travel industry better. Twitter has over 368.4 million active users worldwide, while 11.4 million active users in Thailand in 2022, which is the second rank in Southeast Asia (Statista, 2022). Using specific hashtags on Twitter, it is possible to find tweets related to a particular topic and thereby analyze people's opinions on tourism in Chiang Mai, Thailand, as it has been mainly used in the area of tourism researchers utilize Twitter data (Mehraliyev et al., 2021). Moreover, The upsurge of social media, usage of digital devices, and the Internet of things embroils massive real-time data generation. Customer reviews are essential to other potential customers to understand customers' opinions and help them to make decisions.

However, It becomes challenging for businesses to obtain a comprehensive view of the datasets regarding customers' ideas or reviews. It requires proper analysis of customer reviews to enable potential users to have positive and negative opinions of products and services. Therefore it is highly desirable to produce an automatic analysis or summary of customer reviews that may impact gaining profits and achievement of enterprise on the worldwide market. Currently, sentiment analysis is a popular technique to understand the opinions and feelings of people based on text and a large amount of data [10]. As a result, many researchers try to use machine learning to perform sentiment and opinion mining. Therefore, this research aims to

1. To test the sentiment analysis model from social network comments
2. To compare Support Vector Machines (SVM) algorithms, Naïve Bayes algorithms, and K-Nearest Neighbor kNN algorithms in model testing to classify opinions in segmentation into learning datasets.
3. To evaluate the effectiveness of the model and the acceptance value of the analysis of opinions on social networking sites about tourism as well. Accuracy, Prediction (Precision), Probability (Recall)

2 LITERATURE REVIEW

The upsurge of social media, usage of digital devices, and the Internet of things embroils massive real-time data generation. Customer reviews are essential to other potential customers to understand customers' opinions and help them to make decisions. However, It becomes challenging for businesses to obtain a comprehensive view of the datasets regarding customers' ideas or reviews. It requires proper analysis of customer reviews to enable potential users to have positive and negative opinions of products and services. Therefore it is highly desirable to produce an automatic analysis or summary of customer reviews that may impact gaining profits and achievement of enterprise on the worldwide market.

The current trends of technologies, especially computer science, management sciences, and the social sciences, have contributed to opinion and social media analysis. For the past few years, most research to date has found that online customer reviews has attracted attention from researchers of data mining and natural language processing that was used to track the relationship between words featured in the comments (Bagheri et al., 2013; Drus & Khalid, 2019; Puschmann & Powell, 2018). As a result, sentiment analysis was used to help companies know what customers think about their products and service, improving the quality according to customers' opinions (Chumwatana, 2015).

2.1 Social Media

Social media can be categorized into four types: Content communities (Youtube, Instagram), Social networking (Facebook, LinkedIn), Blogs (Reddit, Quora), and Micro-blogs (Twitter, Tumblr) (Ali et al., 2017). Due to the nature of social media, monitoring customers' comments has become a critical task. Although the companies seem to engage consumers through social media, many companies are not familiar with the data mining processes (Dai et al., 2013). In contrast, successful companies should have the ability to process all available information. Sentiment analysis refers to the machine learning techniques that use to evaluate and classify the opinions on a specific topic of interest (Rambocas & Pacheco, 2018). Kalia (2013) stated that the social media is used for advertisement. The companies used it for lunching the promotions, searching and recruiting the employess, and electronic commerce. Therefore, business tends to use more social media for getting into consumer insight, market intelligence and present an opportunity to learn about customer review and perceptions. Thus, sentiment analysis has received much attention from market research as a practical approach for analyzing social media content. The advantage of sentiment analysis is it collects and analyzes online comments in real-time. This analysis is especially significant to researchers as it can automatically extract high-quality data on emotional expressions.

2.2 Twitter Sentiment Analysis

The sentiment analysis found in the comments or tweets to provide useful indicators for many different purposes. Saif et al. (2012) stated that sentiment could be categorized into two groups, which is negative and positive words. According to (Drus & Khalid, 2019), implementing sentiment analysis can be used either Lexicon-based method, Machine learning method, or a mix of both methods. The Lexicon-based method is known as an unsupervised learning method

that does not require any training data as it depends on the dictionary. Most of the studies adapted Sentiwordnet and TF-IDR methods. Due to the complexity of languages, this approach is not designed to cover all aspect of language especially slang, sarcasm and negation. Moreover, there are some problems such as some words or sentence have different meaning based on the application (Akter & Aziz, 2016; Khan et al., 2016; Nasukawa & Yi, 2003).

Machine learning method is known as an supervise learning as it requires training data in order to be processed. The most used method is the SVM and Naïve Bayes model. Nevertheless, machine learning performs poorly on Facebook as it requires a huge amount of training sample. Furthermore, it is time consuming (Hasan et al., 2018; Kanade et al., 2020; Shobana & Murali, 2021). Interestingly, Dhaoui et al. (2017) found that both types perform very similar in terms of accuracy. In this research, the machine-learning-based approach is suitable for testing text data on Twitter as it has been widely used.

2.3 Machine Learning

It is a type of artificial intelligence and computer science, which focuses on using data and building algorithms that allow computers to learn from sample data or environments. To create prototypes or use algorithms to predict new data. The aim is to develop or improve the performance of the better system (Hasan et al., 2018) 2.3.1. Supervised learning is learning from the characteristics of sample data that have specified desired results or types and then used to predict other data that do not know the answer. In this regard, supervised learning can be applied to the estimation of data, classification of data and forecasting data (Depa, 2022). 2.3.2. Unsupervised learning is the creation of a model that fits with data without specifying the desired effect and it can be used for clustering (Depa, 2022). Nowadays, many researchers use the machine learning to perform the sentiment and opinion mining In this research, the supervised learning is suitable for testing text data on Twitter as it has been widely used (Ahmad et al., 2017; Ahuja et al., 2019; Hasan et al., 2018).

Support Vector Machine (SVM)

Support Vector Machine (SVM) is an algorithm used to analyze data and classify data based on the principle of coefficient of equations to create a line to distinguish data groups that are entered into the training process for the system to learn. with an emphasis on distinguishing groups of data (Bowornlertsutee & Paireekreng, 2022; Vateekul & Koomsubha, 2016)

Naïve Bayes

Naïve Bayes algorithm is a probabilistic machine-learning algorithm, which has a simple algorithm for classifying data by learning the problems to create new data classification conditions (Lewis, 1998). Naïve Bayes's principle is based on the principle of probability. It assumes that the amount of attention depends on the probability distribution (Bowornlertsutee & Paireekreng, 2022). It is a technique for solving the problems with predictable outcomes. By analyzing the relationship between the variables to create a probability condition for each relationship suitable for the case of a large number of sample sets and the attributes of the samples are not independent of each other.

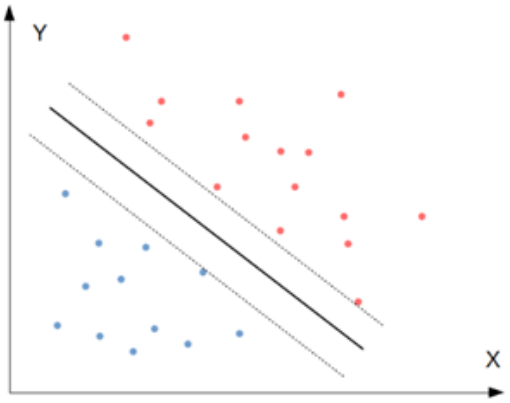

Figure 1: *The characteristics of Support Vector Machine (SVM)*

K-Nearest Neighbor (kNN)

K-Nearest Neighbor is a classing method for categorizing data. It has beed comparing information of interest with other information to see how similar they are. If the information you are interested in is the closest to which information, the system will give an answer that is similar to the answer to the nearest data. The behavior does not use the learning data set (training data) to create the model, as shown in Figture 2 (Sun & Huang, 2010).

Figure 2: *The characteristics of K-Nearest Neighbor (kNN)*

3 MATERIALS AND METHODS

3.1 3.1 Data collection

This research collects comments in the Thai language from social media users on Twitter in the form of retweets or hashtags ”Review Chiang Mai” from August - September 2022. The total of 317 tweets, or 12,174 words, is shown in Table 1. We selected the study mentioned above period because it was the period in which Chiang Mai began to relax from COVID-19 and tourism demand in Chiang Mai also increased.

Table 1: *The example of the comments*

N0	#Review Chiang Mai	Date
1	รีวิวห้องฟ้าจำลอง เชียงใหม่ถ่ายแบบรีบสุดๆเพราะไปเลท เราเลือกไปวันเสาร์เพราะเขามีบรรยากาศเรื่องดวงดาวกับเปิดสารคดีในโดม(นอนดูแบบเต็มตาสุดๆ) เข้าฟรีจอดรถแล้วมีรถกอล์ฟมารับไม่ต้องเดินให้เหนื่อย ใครอินเรื่องดาวต่างๆควรไปและเด็กเยอะนะเตือนไว้ก่อน #รีวิวเชียงใหม่	30/10/2565
2	📸แจกพิกัด 4 จุดถ่ายรูปยอดนิยมในเชียงใหม่ มุมไหนก็สวย มุมไหนก็ปัง ลานเนินนุ่น - Nimman Hill/จุดชมวิว คาดฟ้าเมฆ่า - One Nimman - ขัวเหล็กเสื้อผ้าของแบรนด์ VALINEE ใส่ถ่ายที่ไหนก็สวยแมทช์ ปังสุดๆ สามารถสั่งสินค้าผ่าน DM หรือ Shopee ก็ได้เน้อ💗	30/10/2565
	…..	….
317	อยากโหลดโปรตีน ปิ้งย่างสายเกาก็มาค่าาา 📍ลำโพชะ หัวละ 329.- ไม่รวมน้ำ เครื่องเคียงดี ไข่ตุ๋นเด็ด หมูจะสไลด์บางต่างจากร้านแม่ล้ำลำ แต่รสไม่แพ้กันเลยจ้า มาตรงมื้อโทรจองเน้อ เช็คเงื่อนไขกับเบอร์ในเพจร้านเน้อเจ้า #รีวิวเชียงใหม่	30/10/2565

3.2 Data processing

This step classified the labeled data into two main categories (positive and negative), as the text information on social media is not natural language. Therefore, this process must be converted into a format that the computer language can learn. As a result, the text processing process has steps shown in Figure 3.

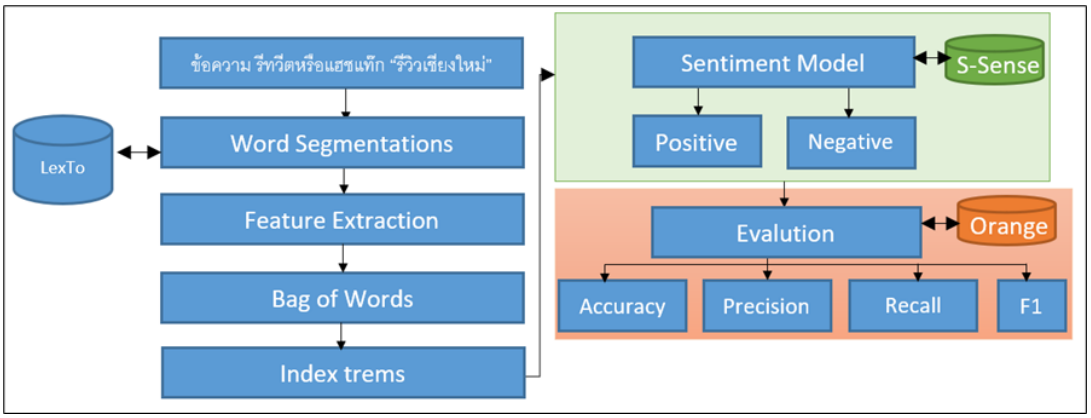

Figure 3: *Data processing.*

- Collecting the text from retweets or hashtags by "Review Chiang Mai" from Twitter.
- According to Thai, words are long sentences. In this research, we use a program for word segmentation of Thai words called the Lexto program, developed by the National Electronics and Computer Technology Center (NECTEC).
- Feature Extraction by tokenizing the words and filtering the length of words between 2 to 25 characters. At this stage, the Orange Data mining program was used.
- Bag of Words is the process used to describe a group of words without grammatical concern. In this process, we used the Orange Data mining program to create Indexing in order to evaluate the Term Weighting in single-word form.
- Sentiment Model, The S-Sense program was used to analyze opinions in three classification algorithms: Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbor for comparison.

- Measure the efficiency of analyzing opinions of tourism in Chiang Mai on Twitter by Accuracy, Precision, Recall, and F-measure.

4 RESULTS

To answer objective 1, We test the sentiment analysis model from social network comments. The results show that a total of 317 tweets, which is information from social media users on Twitter in the form of retweets or hashtags named "Review Chiang Mai," can be labeled of feelings into 2 Levels which are positive (Positive emotion) and negative (Negative emotion) as shown in Table 2 and Table 3.

Table 2: *Positive emotion*

Topics	Text	Sentiment	Score	Rank
Food	ร้าน SUNDAY BAKER ประทับใจมากก ร้านน่ารัก แสงในร้านสวย คุณเจ้าของร้านน่ารักมาก ที่สำคัญคืออาหารอร่อย ครัวซองต์หอมกรอบ แซลมอนบวกสเปสไข่ดีมากกินด้วยกันแล้วลงตัวไปหมด ของหวานก็ดี ชอบที่เค้าทำไอติมไม่หวานกินกับเบิร์นโทสต์ด้านล่างฟินสุด ดีจริงอยากให้ไปลองง #รีวิวเชียงใหม่	Positive	100	1
Tourist destinations	ใครมีแพลนไปเที่ยวเชียงใหม่ แล้วกำลังหาที่พักดีๆ นี่แนะนำโรงแรมริมปิง วิลเลจเลยค่ะ เป็นบูติกโฮเต็กสไตล์ล้านนาสวยๆ น่ารักอบอุ่น ชอบที่ห้องพักสะอาด มีหลายแบบให้เลือก พนักงานน่ารัก บรรยากาศก็ร่มรื่น แถมมีจักรยานให้เราปั่นเที่ยวเมืองเชียงใหม่ได้ฟรีเลย # รีวิวเชียงใหม่	Positive	99.98	2
Food	อีกร้านคือ SUNRAYS CAFE เป็นร้าน Breakfast&brunch ร้านน่ารัก บรรยากาศดี อาหารมีให้เลือกหลายอย่างดี ที่เราสั่งมาเป็นวาฟเฟิลไข่คนเบคอน (165.-) อีกจาน ABF ไข่คนครัวซองเบคอน (178.-) ละก็แฮชบราวน์ (84.-) รสชาติดีตามสไตล์เบรคฟาส แต่แต่งจานดีมาก ผักสด ราคาโอเคเลย #รีวิวเชียงใหม่	Positive	99.97	3
Food	Ausaa อุษา คาเฟ่เล็กๆที่ แม้ริมน่ารักอบอุ่นหัวใจมาก ตกแต่งน่ารัก ช่วงนี้มีต้นคริสต์มาสละด้วยย ใครชอบแนวนี้ไปเลยแนะนำ บรรยากาศดี ขนมก็อร่อยน้ำก็อร่อยย ชอบบบบบบบบบบบ #รีวิวเชียงใหม่	Positive	99.97	4
Tourist destinations	Aono Hotel แนะนำที่พักเชียงใหม่อีกที่ที่ประทับใจมากๆ โรงแรมเล็กๆมีไม่กี่ห้องแต่มุมถ่ายรูปดี มีสระเล็กๆแต่ลึก2เมตร ห้องน่ารัก สะอาดมาก ที่นอนนุ่ม ห้องน้ำดีมาก อยู่ใกล้คลองแม่ข่าแบบเดินไปได้ แต่ที่จอดรถน้อย เราจองในagoda ได้ราคา9xxไม่รวมอาหารเช้าแต่คุ้มมาก #รีวิวเชียงใหม่	Positive	99.93	5
..	…	Positive	..	…
Service Quality	noir cmi เพลงดี สเปซดี ค็อกเทลดี บางวันมีแจ๊สเล่นสดด้วย พนักงานสุภาพมาก ถามว่าชอบมั้ยชอบนะ เมื่อวานไปปิด เลยไปซ้ำวันนี้อีกรอบ อยากจะบอกว่าร้านปิดทุกวันพุธนะจ๊ะ ราคาก็แอบแรงนิดนึง #รีวิวเชียงใหม่	Positive	99.59	15
..	…	Positive	..	…
Tourist destinations	มาเชียงใหม่ห้ามพลาดที่นี่ ฟีลดี อากาศดี เกาหลีสุด ยิ่งหน้าหนาวนี่คือแบบดีมาก มาวิ่งตอนเช้าก็ได้ มาเดินเล่นอะไรได้หมด ตกเย็นก็มีคนเยอะ ก็ได้อีกฟีล	Positive	98.9	30

To answer objectives 2-3, we evaluated the model and compared the analysis by using 3 algorithms: Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbor. The experimental results can be summarized in Table 4 as follows:

Table 4 shows the model evaluations and model performance of 3 types of algorithms which can be summarized as follows.

Table 3: *Negative emotion*

Topics	Text	Sentiment	Score	Rank
Tourist destinations	แนะนำเส้นทางแทรคปั่นจักรยานสุดโหดเหี้ยมหลังห้วยตึงเฒ่า ปั่นทะลุไปได้ถึงศาลากาแฟทางแม่ริม ขาโหดต้องไปโดน ระมัดระวังอันตรายกันให้มาก #รีวิวเชียงใหม่	Negative	99.08	1
Tourist destinations	มาเชียงใหม่ครั้งนี้ผิดหวังอ่ะ ถนน-ไฟทางแย่มาก นิมมานมืดมากมืดแบบไม่น่าเชื่อว่าเป็นสถานที่ท่องเที่ยวสายไฟย้อยลงมาจนถึงระดับคนเดิน ถนนพังหลายเส้นมาก ประตูท่าแพมีแต่ขยะและที่สำคัญเมืองมืดมาก ไม่หมือนเมื่อก่อน #รีวิวเชียงใหม่	Negative	98.46	2
Tourist destinations	ที่พักในตัวเมืองเชียงใหม่ ราคาหลักร้อย สวย Maple Hotel Chiangmai เปิดตัวได้ไม่นาน มีที่จอดรถ ที่พักน่ารักค่ะ อยู่ใกล้ที่เที่ยวเยอะมาก เดินทางสะดวกมากเค้อ #รีวิวเชียงใหม่	Negative	96.97	3
Food	โดนพนักงานสุกี้ช้างเผือกบุลี่เรื่องอ้วน เขาเอาเก้าอี้เหล็กมาแล้วพูดใส่หน้าอูว่าเก้าอี้จะพักพี่ตัวใหญ่ แทนที่จะพูดว่าขอเปลี่ยนเก้าอี้นะตัวนี้มันไม่แข็งแรง ตอนเนี่ยอูยังโทษตัวเองอยู่เลยว่าอ้วนมันผิดมากขนาดนั้นเลยเหรอวะ #รีวิวเชียงใหม่	Negative	96	4
Service Quality	รีวิวสนามบินเชียงใหม่ คนเยอะมาก เที่ยวบินดีเลย์เพียบ มีประกาศว่า เพราะออกต้นทางช้ากว่ากำหนด เลยมาถึงนี่ช้า เผื่อมีประโยชน์กับคนที่กำลังเดินทาง ขณะนี้ #รีวิวเชียงใหม่	Negative	94.12	5

Table 4: *The model evaluations and model performance*

Sentiment	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-measure
Postitive	SVM	97.50	98.70	97.90	98.30
	Naïve Bayes	96.60	97.90	97.50	97.70
	kNN	95.20	95.60	98.00	96.80
Negative	SVM	97.50	93.50	91.70	89.40
	Naïve Bayes	97.10	88.10	86.70	87.40
	kNN	96.20	87.30	72.50	81.70

Positive (Positive emotion)

Support Vector Machines (SVM), the accuracy score was 97.50%, the prediction value (Precision) was 98.70%, the probability (Recall) was 97.90%, and the F-measure was 98.30%.

Naïve Bayes, the accuracy score was 96.60%, the prediction value (Precision) was 97.90%, the probability (Recall) was 97.50%, and the F-measure was 97.70%.

K-Nearest Neighbor (kNN), accuracy score was 95.20%, the prediction value (Precision) was 95.60%, the probability (Recall) was 98.00%, and F-measure 96.80%.

Negative (Negative emotion)

Support Vector Machines (SVM), the accuracy score was 97.50%, the prediction value (Precision) was 93.50%, the probability (Recall) was 91.70%, and the F-measure was 89.40%.

Naïve Bayes, the accuracy score was 97.10%, the prediction value (Precision) was 88.10%, the probability (Recall) was 86.70%, and the F-measure was 87.40%.

K-Nearest Neighbor (kNN), the accuracy score was 96.20%, the prediction value (Precision) was 87.30%, the probability (Recall) was 72.50% and F-measure 81.70%.

The Figure 4 show that when we considered the accuracy (Accuracy), prediction (Precision), probability (Recall), and F-measure in the simulation of all three algorithms, it was found that Support Vector Machines (SVM) classified provided the best results, with a maximum accuracy on both positive and negative emotions. Therefore, the SVM algorithm is an appropriate

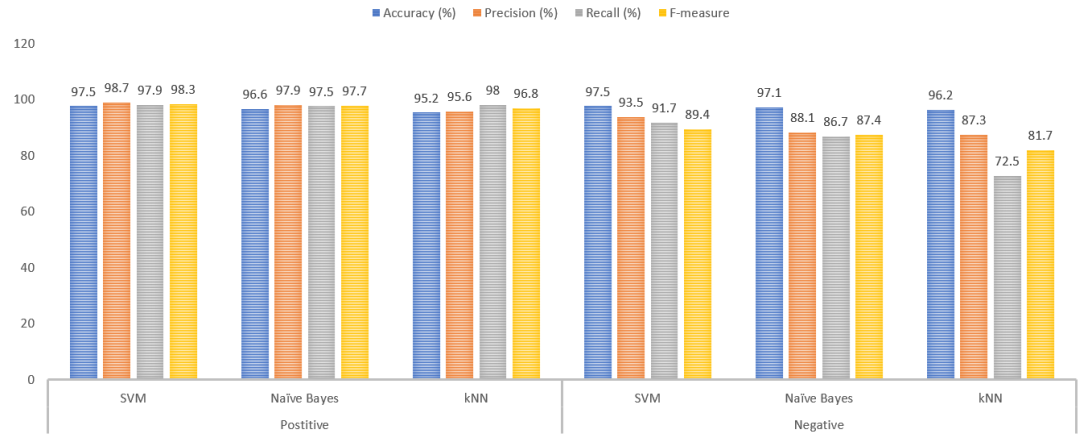

Figure 4: *Comparison between 3 algorithms*

algorithm that can effectively analyze Chiang Mai tourism opinions.

5 CONCLUSION AND DISCUSSION

This research presents a method for analyzing Thai tourism opinions in Chiang Mai from social media users on Twitter in the form of retweets or hashtags, the so-called "Review Chiang Mai." Based on Thai language tweets related to Chiang Mai tourism. Using feature extraction and a Thai Lexeme Tokenization and Normalization Tool (LexTo) developed by the National Electronics and Computer Technology Center (NECTEC). Moreover, word segmentations from the dictionary compared Longest Matching and Lexitron were combined. These technics are in line with the study of (Saensuk et al., 2020; Tesmuang & Chirawichitchai, 2020). Therefore, the Lexto program is suitable for extracting Thai text from comments from social media users. Similar to the S-sense program, it was used to analyze sentiments on social media in terms of the natural language processing techniques to analyze emotions from the text, as it has been widely used in other pieces of research. Haruechaiyasak et al. (2013) and Haruechaiyasak et al. (2018) used the S-sense program to analyze emotions and feelings from the text, and two sentiment levels were identified, positive and negative.

Three classification algorithms, Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbor, were used to test the classification model. It can be seen that the models can analyze two levels of emotional feelings and which support vector machine achieved the highest accuracy on all datasets. The result of this analysis was also in line with the studies of Ahmad et al. (2017) and Sontayasara et al.(2021) based on the optimistic view of the use of SVM classification in Sentiment analysis. Our research aligned with that of Kuhamanee et al. (2017) and (Leelawat et al., 2022), who found that among Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbor (kNN). However, for the number of analyzed terms unigrams can even provide higher accuracy than bigrams. The top 10 words implied that Twitter users have positive sentiments toward food, tourist destinations, and service quality in Chiang Mai, but negativesentiments toward the tourist destinations. It is likely that these negative impressions have greatly affected the tourism industry as (Leelawat et al., 2022) suggested.

6 LIMITATION AND SUGGESTION

The limitation of this study is that the S-Sense program analyzed only two levels of feeling: positive and negative, which may cause biased opinions. However, there will be some neutral opinions. Therefore, the analysis of three levels of feeling will make the model more accurate. The model should be performed to make predictions more efficiently by incorporating other algorithms into the data preparation process, such as sarcasm and Named-entity tagging to feature extraction, which includes emotional: positive, negative, and neutral, to perform a better model.

REFERENCES

Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment Analysis of Tweets using SVM. International Journal of Computer Applications, 177, 975–8887. https://doi.org/10.5120/ijca2017915758

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science, 152, 341–348. https://doi.org/10.1016/j.procs.2019.05.008

Akter, S., & Aziz, M. T. (2016). Sentiment analysis on facebook group using lexicon based approach. 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 1–4. https://doi.org/10.1109/CEEICT.2016.7873080

Ali, K., Dong, H., Bouguettaya, A., Erradi, A., & Hadjidj, R. (2017). Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework. 2017 IEEE International Conference on Web Services (ICWS), 660–667. https://doi.org/10.1109/ICWS.2017.79

Bagheri, A., Saraee, M., & de Jong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. Knowledge-Based Systems, 52, 201–213. https://doi.org/10.1016/j.knosys.2013.08.011

Bowornlertsutee, P., & Paireekreng, W. (2022). The Model of Sentiment Analysis for Classifying the Online Shopping Reviews. Journal of Engineering and Digital Technology (JEDT), 10(1), Article 1.

Chumwatana, T. (2015). Using Sentiment Analysis Technique for Analyzing Thai Customer Satisfaction from Social Media. 146.

Dai, Y., Kakkonen, T., Arendarenko, E., Liao, D., & Sutinen, E. (2013). MOETA: A novel text-mining model for collecting and analysing competitive intelligence. International Journal of Advanced Media and Communication, 5(1), 19–39. https://doi.org/10.1504/IJAMC.2013.053672

Depa. (2022). Machine Learning สิ่งใกล้ตัวแห่งโลกยุคใหม่. https://www.depa.or.th/th/article-view/article11-2563

Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: Lexicon versus machine learning. Journal of Consumer Marketing, 34(6), 480–488. https://doi.org/10.1108/JCM-03-2017-2141

Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. Procedia Computer Science, 161, 707–714. https://doi.org/10.1016/j.procs.2019.11.174

Flores-Ruiz, D., Elizondo-Salto, A., & Barroso-González, M. de la O. (2021). Using Social Media in Tourist Sentiment Analysis: A Case Study of Andalusia during the Covid-19 Pandemic. Sustainability, 13(7), Article 7. https://doi.org/10.3390/su13073836

Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. International Journal of Information Management, 49, 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2013). S-Sense: A Sentiment Analysis Framework for Social Media Sensing. Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP), 6–13. https://aclanthology.org/W13-4202

Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2018). S-Sense: A Sentiment Analysis Framework for Social Media Monitoring Applications. Information Technology Journal, 14(1), 11–22.

Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. Mathematical and Computational Applications, 23(1), Article 1. https://doi.org/10.3390/mca23010011

Kalia, G. (2013). A Research Paper on Social media: An Innovative Educational Tool. Issues and Ideas in Education, 1(1), 43–50. https://doi.org/DOI: 10.15415/iie.2013.11003

Kanade, A., Maniatis, P., Balakrishnan, G., & Shi, K. (2020). Learning and Evaluating Contextual Embedding of Source Code. Proceedings of the 37th International Conference on Machine Learning, 5110–5121. https://proceedings.mlr.press/v119/kanade20a.html

Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S., & Khan, K. H. (2016). Sentiment analysis and the complex natural language. Complex Adaptive Systems Modeling, 4(1), 2. https://doi.org/10.1186/s40294-016-0016-9

Kuhamanee, T., Talmongkol, N., Chaisuriyakul, K., San-Um, W., Pongpisuttinun, N., & Pongyupinpanich, S. (2017). Sentiment analysis of foreign tourists to Bangkok using data mining through online social network. 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), 1068–1073. https://doi.org/10.1109/INDIN.2017.8104921

Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. Heliyon, 8(10), e10894. https://doi.org/10.1016/j.heliyon.2022.e10894

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In C. Nédellec & C. Rouveirol (Eds.), Machine Learning: ECML-98 (pp. 4–15). Springer. https://doi.org/10.1007/BFb0026666

Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2021). Sentiment analysis in hospitality and tourism: A thematic and methodological review. International Journal of Contemporary Hospitality Management, 34(1), 46–77. https://doi.org/10.1108/IJCHM-02-2021-0132

Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2016). SemEval-2016 Task 4: Sentiment Analysis in Twitter. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 1–18. https://doi.org/10.18653/v1/S16-1001

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing (p. 77). https://doi.org/10.1145/945645.945658

Paolanti, M., Kaiser, C., Schallner, R., Frontoni, E., & Zingaretti, P. (2017). Visual and Tex-

tual Sentiment Analysis of Brand-Related Social Media Pictures Using Deep Convolutional Neural Networks. In S. Battiato, G. Gallo, R. Schettini, & F. Stanco (Eds.), Image Analysis and Processing—ICIAP 2017 (pp. 402–413). Springer International Publishing. https://doi.org/10.1007/978-3-319-68560-1_36

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Puschmann, C.,& Powell, A. (2018). Turning Words Into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media. Social Media + Society, 4(3), 205630511879772. https://doi.org/10.1177/2056305118797724

Saensuk, M., Songram, P., & Chompoowiset, P. (2020). Feature-Based Polarity Specification of SmartPhone on Facebook. Journal of Innovative Technology Management Rajabhat Maha Sarakham University, 6(2), 67–79.

Saif, H., He, Y., & Alani, H. (2012). Semantic Sentiment Analysis of Twitter. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. X. Parreira, J. Hendler, G. Schreiber, A. Bernstein, & E. Blomqvist (Eds.), The Semantic Web – ISWC 2012 (pp. 508–524). Springer. https://doi.org/10.1007/978-3-642-35176-1_32

Shobana, J., & Murali, M. (2021). An efficient sentiment analysis methodology based on long short-term memory networks. Complex & Intelligent Systems, 7(5), 2485–2501. https://doi.org/10.1007/s40747-021-00436-4

Sontayasara, T., Jariyapongpaiboon, S., Promjun, A., Seelpipat, N., Saengtabtim, K., Tang, J., & Leelawat, N. (2021). Twitter Sentiment Analysis of Bangkok Tourism During COVID-19 Pandemic Using Support Vector Machine Algorithm. Journal of Disaster Research, 16(1), 24–30. https://doi.org/10.20965/jdr.2021.p0024

Statista. (2022). Leading countries based on number of Twitter users as of January 2022. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Sun, S., & Huang, R. (2010). An adaptive k-nearest neighbor algorithm. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 1, 91–94. https://doi.org/10.1109/FSKD.2010.5569740

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information Fusion, 36, 10–25. https://doi.org/10.1016/j.inffus.2016.10.004

Tesmuang, R., Chirawichitchai, N. (2020). Sentiment Analysis of Thai Online Product Reviews using Genetic Algorithms with Support Vector Machine. Progress in Applied Science and Technology, 10(2), Article 2. https://doi.org/10.14456/past.2020.8

Tourism Authority of Thailand. (2022). Chiang Mai. https://www.tourismthailand.org/Destinations/Provinces/Chiang-Mai/101

Valdivia Pérez, A., Arteaga Pérez, L., Escortell Mayor, E., Monge Corella, S., & Villares Rodríguez, J. E. (2009). Analysis of complaints in primary care using statistical process control. Revista de calidad asistencial, 24(4), 155–161. https://doi.org/10.1016/s1134-282x(09)71799-3

Vateekul, P., & Koomsubha, T. (2016). A study of sentiment analysis using deep learning techniques on Thai Twitter data. 2016 13th International Joint

Conference on Computer Science and Software Engineering (JCSSE), 1–6. https://doi.org/10.1109/JCSSE.2016.7748849