International Journal of Asian Business and Information Management
Volume 14 • Issue 1

Thai Commercial Banks on Twitter:
Mining Intents, Communication Strategies, and Customer Engagement

Mathupayas Thongmak, Thammasat Business School, Thammasat University, Thailand*

ABSTRACT

Twitter is a social media (SM) platform that rapidly generates electronic word of mouth (e-WOM). Marketer-generated content (MGC) is controllable and could enhance the positive e-WOM. Hence, in this study, the author examined the characteristics of MGC and reactions from followers based on Thai banks' Twitter accounts. The author collected a total of 10,000 tweets from nine banks in Thailand—both high- and low-performing banks. The author conducted research with natural language processing (NLP) to uncover intents using an open application programming interface (API). The author used three data-mining techniques—association, clustering, and classification. The Twitter strategies of banks with high and low performances are quite similar. The sentiment is the intent type that dominates Thai banks' intent strategies. Several intents could be combined to draw e-WOM in terms of favorites (FAV) and retweets (RT). Six intent patterns (clusters) were extracted. Some of these clusters are classifiers for FAV and non-FAV tweets. This study guides the application of data mining in business research and suggests MGC strategies for marketers.

KEYWORDS

Association, Banking, Classification, Clustering, Decision Tree, Intent Mining, Social Media Strategy, Twitter

1. INTRODUCTION

Information communication technologies (ICTs) disrupt many sectors, including financial services that rely heavily on ICTs to provide excellent services to customers (Anshari et al., 2020). Four forces (financial crises, changes in customer behavior, the pace of innovation diffusion, and the emergence of non-banks), which directly or indirectly connect to a social media (SM) phenomenon, cause banking to digitally transform. SM has a strong influence on people's activities and banking with no exception (Kirakosyan, 2014). SM gradually evolves the commercial and social exchange of information (Akhlaq et al., 2021). It is the most effective medium for communication that could lead to brand reputation, customer satisfaction, and customer loyalty (Ajina, 2019; Kirakosyan, 2014). Customers tend to search for product or service information as well as reviews from others through SM (Jaman et al., 2020).

DOI: 10.4018/IJABIM.321730
*Corresponding Author

They are also interested and demand to interact with banks through SM (Mitic & Kapoulas, 2012). They are looking for a brand's SM to form brand judgments (Culotta & Cutler, 2016).

SM could create benefits for internal purposes, for customer-related purposes, and for external partners/ suppliers (Botchway et al., 2020). It is attractive for banks seeking to increase their competitiveness and build a rapport with their stakeholders (Cosimato & Troisi, 2015; Mitic & Kapoulas, 2012). Thus, banks need a clear strategy to profitably use SM as marketing tools, such as implementing SM as a new financial delivery channel, managing relationships with customers, promoting digital relationships with huge anonymous users, promoting products or services, and performing public relations (Bohlin et al., 2018; Jaman et al., 2020; Khajeheian & Mirahmadi, 2015; Kirakosyan, 2014; Mitic & Kapoulas, 2012; Ocampo et al., 2021; Shabbir, 2020). Many companies view SM as an effective channel for spreading the electronic word of mouth (e-WOM) (Jaman et al., 2020).

Firms apply Twitter to accelerate communication with individuals (customers, potential customers, and investors) (Chahine & Malhotra, 2018; Surucu-Balci, et al., 2020). Business sectors, particularly those in the service sectors, widely use microblogs such as Twitter to update customers about the latest news or redirect links from their websites (Ocampo et al., 2021). Twitter is a SM platform that allows users to send and read short messages called tweets (Garg & Rani, 2017; Jaman et al., 2020; Mucan & Özeltürkay, 2014). Users can also simply post by retweeting others' tweets, encouraging them to post more often (Alboqami et al., 2015; Mosley, 2012). So, it is suggested to be focused to create e-WOM (Alboqami et al., 2015; Bohlin et al., 2018; Fitri et al., 2019; Mucan & Özeltürkay, 2014; Weerawatnodom et al., 2017). Brand-focused e-WOM on Twitter also becomes an important source of marketing information, which deserves researchers' and practitioners' attention (Chu & Sung, 2015). In the United States, Twitter is the most frequently used platform in all industries except retailing. Facebook and Twitter are also tied in the financial service sector (Mucan & Özeltürkay, 2014). Using Twitter is thus important for the banking industry (Alboqami et al., 2015; Mucan & Özeltürkay, 2014; Senadheera et al., 2011).

Although online content marketing dominates the online strategy of companies nowadays and SM is standard practice for several banks, they face the challenges to harness Twitter's power to manage customer relationships and engage customers (Ajina, 2019; Chikandiwa et al., 2013; Kirakosyan, 2014; Ocampo et al., 2021). Banks are not so active in SM communication, although customer engagement could lead to trust, loyalty, commitment, and advocacy (Ajina, 2019; Cosimato & Troisi, 2015). Besides, marketer-generated content (MGC) is spread rapidly among SM users for some content, but other content receives an inadequate reaction, no reaction, or is not spread. The results of MGC in SM such as Twitter could be tracked by the number of retweets (RT) and favorites (FAV), which are different among different tweets (Alboqami et al., 2015). Creating interactive and relevant content is thus crucial for SM engagement (Mitic & Kapoulas, 2012).

Enthusiasm for the SM power to enhance customer relationships is not equally shared among all banks (Mitic & Kapoulas, 2012; Mucan & Özeltürkay, 2014). Banks in different countries have different approaches or levels of SM use (Kirakosyan, 2014). Understanding the key characteristics of MGC posts (e.g., Twitter tweets) affecting e-WOM success (e.g., RT and FAV) is an interesting issue (Alboqami et al., 2015; Fitri et al., 2019; Weerawatnodom et al., 2017). A tweet's characteristics, such as the fluency of messages, content type, the existence of a link, and the existence of a call-to-action, significantly influence its engagement rate (Surucu-Balci et al., 2020). Intent categories, as one of the tweets' characteristics, are identified by the studies of Wang et al. (2015), Hollerit et al. (2013), and Pandey et al. (2018). But user intent mining has been a relatively new area in social media research (Pandey et al., 2018; Wang et al., 2015). These studies also do not link a tweet's characteristics with the expected results such as e-WOM or engagement. Companies normally analyze sentiment to collect feedback from customers, but they do not explore the sentiment of their tweets and their impact (Mosley, 2012; Sahu et al., 2015; Senadheera et al., 2011), particularly in the banking context such as the study of Botchway et al. (2020).

Asian countries are ranked in the fourth position in SM use globally (Jaman et al., 2020). In Thailand, there are 52 million active SM users, which accounted for 75% of the total population (We Are Social Inc. & Hootsuite, 2020). Twitter dominates 13.96% of SM in Thailand (StatCounter, 2021).

Thus, Twitter is an excellent platform to be analyzed because of its popularity, relevancy, public social connection, and well-organizing (Culotta & Cutler, 2016; Ligiarta & Ruldeviyani, 2022). One hundred thousand tweets are generated daily (Alboqami et al., 2015; Garg & Rani, 2017). However, according to a survey of finding trends in data-mining techniques for social media analysis by Nanayakkara et al. (2021), there are only a few studies conducted in Thailand and Southeast Asian countries.

Big data analytics plays a crucial role in businesses' marketing decisions, but analyzing Twitter data is relatively rare in the marketing literature (Culotta & Cutler, 2016). SM data such as Twitter data, representing a large part of big data, make the analysis tasks more difficult (Benslama & Jallouli, 2020). Therefore, several methods and techniques, such as data mining, visualization, and machine learning, could help in this context (Benslama & Jallouli, 2020). For banks, big data analytics provides various advantages for banks and their customers, including fraud detection and prevention, customer segmentation, risk management, and future predictions (More & Moily, 2021). SM analytics with data-mining techniques could be applied to extract trends, patterns, and rules from the SM pool (Nanayakkara et al., 2021). In the banking field, data mining is used to detect fraud, assess risks, and analyze trends and profitability (Nanayakkara et al., 2021). Data mining can also help banks to improve decisions about customers in marketing and relationship management (Khder et al., 2021; Park & Javed, 2020; Voican, 2020). According to the literature review by Benslama and Jallouli (2020) regarding mining (clustering) SM data and marketing decisions, there is only a study in the banking context to extract knowledge to help the banking industry.

In light of these research gaps, this study posits the following research questions focusing on the MCG-based e-WOM:

- **RQ1:** What are the tweets' characteristics of Thai commercial banks' official accounts (high-performing vs low-performing banks)?
- **RQ2:** What characteristics in terms of intent categories are associated with tweets being retweeted and favorited?
- **RQ3:** What patterns or strategies can be found in banks' Twitter in the aspect of tweet intents?
- **RQ4:** What bank tweets' characteristics (controllable and uncontrollable characteristics) indicate followers' engagement success (RT and FAV)?

This study proposes to extract tweet characteristics from Twitter data posted by the official Twitter accounts of Thai commercial banks, to identify tweet intents from those tweets, to compare the association of intents between RT tweets and FAV tweets, to find the strategies of tweet intents, and to explore the intent strategies together with other characteristics to predict RT or FAV on each tweet, using data-mining techniques (association rules, clustering, and classification).

2. RELATED RESEARCH

Hamzah and Hidayatullah (2018) tried to cluster Twitter data from the official account of higher education institutions based on hashtags. The results showed that Indonesian higher education institutions mostly used Twitter to post general information, news, agenda, announcement, information for new students, and achievements. Pedrood and Purohit (2018) proposed a novel learning method for identifying help intent for a new disaster on Twitter. The method transferred the knowledge of help intent from labeled messages of past disasters using the new sparse coding feature representation, which could increase the performance gain up to 15%. Bohlin et al. (2018) investigated the use of three SM platforms—Facebook, Twitter, and YouTube—by 100 leading global banks. Findings indicated that banks offered SM services in nine areas: marketing, financial education and advice, information support, customer support, sales representativeness, customer engagement, online recruitment, survey and polling, and other services.

Saura et al. (2019) examined user-generated content (UGC) on Twitter during the Black Friday event. The results revealed that consumers had positive perceptions of the topics "exclusive promotions and smartphones," but had negative perceptions of the topics "fraud, insults and noise, and customer support." Ajina (2019) determined the role of SM on customer loyalty in banks in Saudi Arabia. Findings pointed out that the use of SM to engage customers impacted customer loyalty.

Olaleye et al. (2020) applied sentiment analysis to explore bank customers' satisfaction. Findings showed a slight difference between the polarities of customer tweets in the International Authorization Banking group and the National Authorization group. Khruahong et al. (2020) explored comments on Facebook and YouTube to analyze the framework of car purchase decisions of Thai consumers. The results indicated that positive sentiment should be more than 69.85%, while negative sentiment should not exceed 30.15%. Surucu-Balci et al. (2020) examined Twitter post characteristics leading to higher stakeholder engagement in the container shipping market. The result from Decision Tree showed that tweet fluency, the tangibility of company resources in the tweet, vividness, content type, the existence of a link, and the existence of a call-to-action significantly affected stakeholder engagement.

Park and Javed (2020) aimed at recognizing customers' sentiments and insights for strategic decision-making in Saudi Arabia's financial industry by collecting data from Facebook and Twitter in three financial companies and using data mining to analyze sentiments. Findings revealed that all companies had similar likes and followers. One company had no negative sentiments in posts in English, while two companies had more Internet penetration than one company. Botchway et al. (2020) conducted a sentiment analysis on large-scale tweet data relating to Ecobank Group, a bank in Sub-Saharan Africa. The results showed that the outperformed sentiment lexicon was Valence Aware Dictionary and sEntiment Reasoner (VADER) in terms of accuracy and computational efficiency, compared with the other three lexicons.

Asali (2021) investigated Indonesian banks' consumer insights in terms of their sentiments on SM toward mobile banking features, including payment, block, opening a new bank account, log-in, transaction report, bank balance, top-up, transaction, and transfer. More than 5,000 tweets were classified by sentiments, in which negative tweets had the biggest proportion followed by neutral tweets. Negative tweets were also greater than positive tweets in all services. Illia et al. (2021) empirically examined which conditions of Tweets promoted negative sentiments and created negative impacts on Italian bank outcomes using autoregressive time series models. Findings indicated that a tweet impacted a bank's outcomes only for a tweet embedded in a larger conversation about the bank, but not a simply repetitively shared tweet.

Ghobakhloo and Ghobakhloo (2022) tried to identify customer needs by extracting opinions and analyzing sentiments to design a recommendation system to provide suitable services based on their satisfaction, sentiments, and experiences. The opinions and experiences of customers were obtained from tweets with hashtags of titles or headings of banking services. Classification methods with opinion mining and terminal-designed systems were combined to provide personalized services of the banking system for customers. Ligiarta and Ruldeviyani (2022) examined customer satisfaction regarding mobile banking using Twitter data and sentiment analysis. The support vector machine (SVM) was applied to predict positive or negative sentiments. Data from several mobile banking were analyzed. One platform received more positive sentiments from customers than others, while others had issues with reliability, usefulness, and responsiveness.

3. THEORETICAL BACKGROUND

Brand marketers continuously use SM such as Twitter to communicate with customers and promote their products or services. E-WOM is one dimension encompassed by social media marketing (SMM). SMM has a significantly positive effect on overall brand equity (Hafez, 2021). Twitter is a good platform for e-WOM, especially favorites (likes) and retweets (shares) that can help spread messages and enhance interaction among customers (Alboqami et al., 2015; Chu & Sung, 2015; Kim et al., 2014; Mucan & Özeltürkay, 2014; Weerawatnod et al., 2017). Both RTs and FAVs have significant and positive correlations with user participation on Twitter (Grover & Kar, 2020). As same as RTs, the recommendation from a group chat on another social media (i.e., WeChat) significantly affects the expected intention of users, such as online purchase intention (Shaheen, 2022). The majority of banks use Twitter to offer customer services, customer supports, and financial education and services (Bohlin et al., 2018; Chikandiwa et al., 2013; Mucan & Özeltürkay, 2014; Senadheera et al., 2011). Marketers of banks can use Twitter to publicize news, product/service information, promotions, and so on (Alotaibi, 2013; Weerawatnodom et al., 2017). There should be some MGC characteristics that play a crucial role in getting more reactions in terms of FAVs and RTs (Alboqami et al., 2015). This study classifies MGC as marketer-controllable characteristics and fundamental characteristics.

3.1 Marketer-Controllable Characteristics

Characteristics of MGC posted on Twitter could be classified as contextual, entertainment, informational, and brand (Alboqami et al., 2015; Weerawatnodom et al., 2017). These characteristics are proposed to affect RTs and FAVs afterward. Tweet types, hashtags, and mentions are contextual characteristics in the study of Alboqami et al. (2015). Mentions, tweets, retweets, comments, likes/favorites, and hashtags are major types of interactions on Twitter (Reyes-Menendez et al., 2020). Retweets, tweets, and mentions also support sharing and conversations (Senadheera et al., 2011).

3.2 Tweet Type

Twitter allows users to create microblogs called tweets. In addition to tweeting, users can also retweet, favorite, or reply to a tweet. Retweeting is forwarding a message to others, while replying is an interaction with a tweet sender. Favorite shows users' impression of a tweet (Abunadi, 2015; Boyd et al., 2010; Farina et al., 2014; Icha, 2015; Mosley, 2012; Mucan & Özeltürkay, 2014; Purohit et al., 2013; Rantanen et al., 2019; Weerawatnodom et al., 2017).

3.3 Hashtag and Mention

Another innovative feature of Twitter is the hashtag. A user can use a hashtag to track or write on a given topic by putting a # sign in front of the topic (Abunadi, 2015; Farina et al., 2014). The hashtag feature creates collaboration effectiveness that enables any user to write about a user-defined topic or event. Anyone who views a hashtag is able to see others' views of the same hashtag (Abunadi, 2015; Icha, 2015; Purohit et al., 2013). Thus, messages can be grouped by topics using hashtags (Mosley, 2012; Purohit et al., 2013). Hashtags make tweets easier to find with the Twitter search feature (Rantanen et al., 2019). Hashtags, as one Twitter-specific feature, strongly influence information diffusion (Sridevi et al., 2020).

On Twitter, direct conversations called mentions usually involve the use of @ symbol to refer to others and address messages to them; these messages are visible to the public (Boyd et al., 2010; Mosley, 2012). Messages with @user are attention-seeking (Boyd et al., 2010). Mentions decreased RTs in the study by Weerawatnodom et al. (2017). But several studies in the literature reveal the positive effect of mentions on message dissemination (Lahuerta-Otero et al., 2018).

Both hashtags and mentions significantly affected information diffusion in past research (Sridevi et al., 2020). Users writing mentions and hashtags significantly got likes and RTs in the study by Lahuerta-Otero et al. (2018). Adding a hashtag in the tweet significantly decreased favorites, while embedding a mention in the tweet significantly lowered retweets (Alboqami et al., 2015). The use of hashtags and user mentions was significantly negatively associated with the RT count in the study by Nanath and Joy (2021). However, functional interactivities, such as hashtags and mentions, significantly generated more public engagement on Twitter in the study by Zhang et al. (2022). There was also a strong positive interaction effect when mixing hashtags, mentions, and links in the study by Sridevi et al. (2020).

3.4 Media Type

SM enables companies to build a brand by distributing photos and videos (Jaman et al., 2020). Data generated from Twitter is heterogeneous because users can post texts, images, and audio/videos in any format (Garg & Rani, 2017). Using additional media such as photos or videos is one technique considered as a best practice for a Twitter account (Hougaard, 2017). Marketers can apply texts, graphics, or videos to create more effective strategies (Ocampo et al., 2021). Pictures, videos, or only texts are media types that have been explored for their impacts on FAVs and RTs in the study by Alboqami et al. (2015). Having a picture in the tweet has a significantly positive effect on both FAVs and RTs. Pictures and videos, as SM post characteristics presenting vividness, significantly affect likes, comments, and shares; these types of posts also drive likes and shares (Kordzadeh & Young, 2022). In the study by Wang and McCarthy (2021), for Singapore, posts with videos led to more comments and shares, while posts with photos created more likes and emojis. In Australia, posts with photos and videos helped to enhance engagement in different forms than those with text only. In the study by Abbas et al. (2021), photo posts on Instagram received fewer comments than video posts. Video was one of the significant drivers for the RT model in the study by Weerawatnodom et al. (2017). A combination of these media types (e.g., texts and photos or texts and videos) yielded better results in generating behavioral engagement in the study by Tafesse and Wien (2018).

3.5 Day and Time

The time of the day and day of the week are SM post characteristics in terms of timing that influence user engagement in past studies (Kordzadeh & Young, 2022). According to Boyd et al. (2010), some people retweet time-sensitive materials and breaking news. The day of the week was also a control variable in the conceptual model for brand-post engagement on SM in the study by Menon et al. (2019). Prior research identified weekends as the optimal timing for tweets (McShane et al., 2021). Timing in terms of the day of the week, day of the year, month, weekend, or weekday, and the year of the tweet was also controlled in the study by McShane et al. (2021). All of them affected likes and almost all of them except for the day of the year significantly affected RTs.

3.6 Intent

Intent is a purposeful action that could help to identify actionable information (Purohit, et al., 2015). SM enables marketers to analyze the intention of users (Hamroun & Gouider, 2020). Intention detection on SM is a valuable source of information for online businesses. Detecting users' purposes and goals from their actions is called intent mining (Mishael & Ayesh, 2020). Analyzing incorporating intent mining could provide better guidelines to create effective content (Wang et al., 2022).

Examples of mining intent in the context of buying and selling intention are bidding, buying, cheap, purchasing, and selling (Hollerit et al., 2013). Wang et al. (2015) classified intent tweets of Groupon as food and drink, travel, career and education, goods and services, event and activities, and trifle. Past research extracted three user query intents consisting of navigational, informational, and transactional intent. The study by Pandey et al. (2018) detected policy-affecting intent as specific intent categories as follows: accusational, validational, sensational, or no intent. Past research indicated that common topics in tweets concerning brands were asking questions, describing interests, expressing attitudes/opinions, and sharing information, news, or updates on daily activities (Chu & Sung, 2015). Several studies in the banking context detected sentiments of UGC (Cahyonoa et al., 2020; Khanum et al., 2016; Alamsyah & Indraswari, 2017; Shakeel et al., 2020), rather than the sentiments of MGC, which could affect the engagement as well. Hence, this study detects MGC tweets from banks as requests, sentiments, questions, and announcements.

3.7 Fundamental Characteristics

3.7.1 Bank Revenue and Bank Code

Banks could use financial innovations to enhance the efficiency of financial activities (Anshari et al., 2020). But different banks have different sizes. Large companies tend to adopt SM faster than smaller ones because of their abundant wealth. Big companies can also hire more professional teams to support fluently using SM compared with small firms with limited capital resources. Thus, Jaman et al (2020) proposed that company size could increase SM adoption. In addition, Senadheera et al. (2011) indicated that smaller banks are less active on Twitter. Only a few of them are active or have verifiable Twitter accounts. However, Chahine and Malhotra (2018) found that market reactions to the launching of a Twitter platform are more positive for small-size firms and those firms with losses.

Different banks also perform differently. Shabbir and Zeb (2020) evaluated the performance of conventional banks compared with Islamic banks, and their findings showed that Islamic banks have marginal bank spread but bear higher operational costs than conventional ones. Shabbir and Zeb's (2020) study also revealed that bank types affect customers' trust differently. According to Ahmad and Khan (2021), the technical efficiency of private banks has an edge over public banks. In contrast, Kaura's (2013) study showed no significant difference between public and private banks relating to the positive IT impact on customer satisfaction. In terms of service quality, there were quality gaps between public banks and private banks in the study by Mishra et al. (2010) regarding reliability and empathy, and Singh's (2013) study showed quality gaps regarding reliability, tangibility, and assurance. Kaur and Kiran (2015) also revealed a significant difference in the service quality among private, public, and foreign banks.

3.8 Follower

Each Twitter account has followers who receive tweets and updates from the account through their timelines (Abunadi, 2015). The number of followers presents brand engagement (Hoffman & Fodor, 2010), and followers are subscribers of other users' tweets (Mosley, 2012). Generally, the majority of Twitter posts are visible to those followers on a Twitter page (Chu & Sung, 2015). Retweets are forwarding a message to other followers (Abunadi, 2015; Mosley, 2012). The number of followers as one of the content features could affect the speed and spread of Twitter content (Nanath & Joy, 2021). Gunarathne et al. (2015) pointed out that a company tends to respond to a tweet sent by a customer with a high number of followers, while Gunarathne et al. (2018) revealed that an airline responds to a complaint sent by a customer with a high number of followers as well. Ehrmann and Wabitsch (2022) investigated both English and German samples and found that tweets from accounts with more followers had more likelihood to get RTs and FAVs.

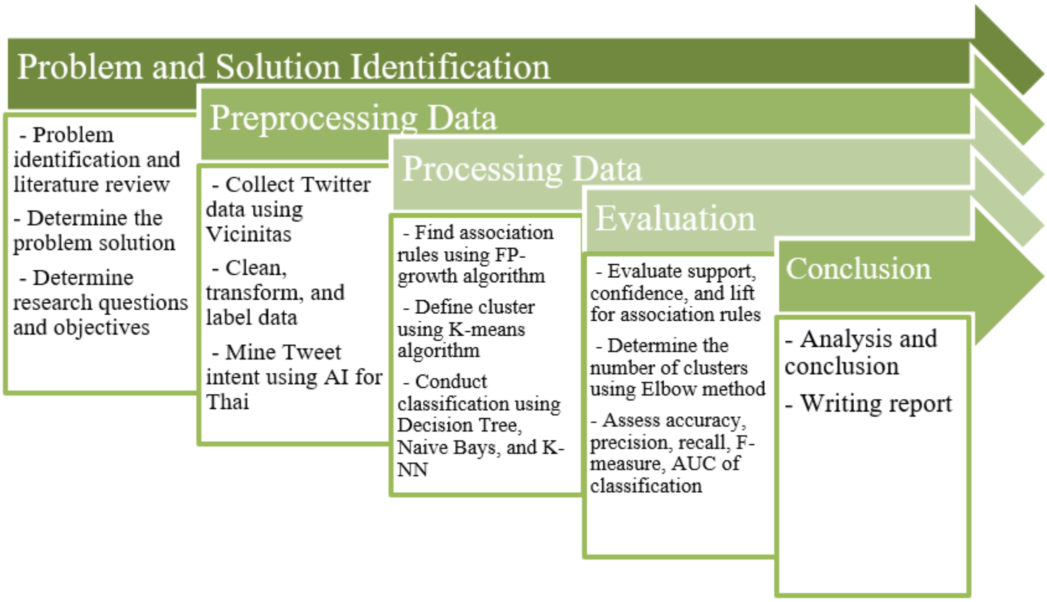
4. METHODOLOGY

SM data from millions of users are important for businesses to know about their customers' needs and preferences and how to satisfy them (Khder et al., 2021). Therefore, this study is a sub-project of a project titled "The Analysis of Twitter Usage in Thai Banks" to examine SM data for businesses. Figure 1 presents the overall research processes applied from the study by Fitri et al. (2019), which is in line with the study by Asali (2021). The details of each process are as follows:

4.1 Problem and Solution Identification

At this stage, I identified problems that were the low follower engagement and the right characteristics of MGC in terms of Tweets from Thai commercial banks. I then formed research questions and objectives as described in the Introduction section. The scope of this study was limited to nine commercial banks (private banks) in Thailand that have mobile banking, and five of them were the best performers in 2019 and 2020. Only banks having official Twitter accounts were chosen. I carried out literature studies to determine possible solutions (mining techniques) and to review past research. SM analytics is a way to gather data from SM and analyze that data, normally in the text-based form, for insights. In businesses, it is typically used to understand brand popularity and performance (Asali, 2021). Text mining was also an important step in discovering knowledge and profitable information in which resources are unstructured data like SM data (Ghobakhloo & Ghobakhloo, 2022). I planned

Figure 1.
Research procedure



to apply RapidMiner for academics because it is one of the leading software platforms for machine learning, text mining, predictive analysis, and business analytics. Besides, it is placed in the leading quadrant of Gartner's Magic Quadrant for Advanced Analytics in 2014 and received one of the strongest satisfaction ratings from the Rexer Analytics Data Miner Survey in 2011 (Dwivedi et al., 2016).

4.2 Preprocessing Data

To automate the data collection process, I applied the Vicinitas Twitter analysis tool (https://vicinitas.io) for data collection from nine banks' Twitter accounts (searching by User Tweets). At first, Twitter data of each bank were sorted descending by the number of FAV to ensure a high engagement level. I used quota sampling to get a balanced dataset of 10,000 tweets (5,000 tweets from five high-performing banks and 5,000 tweets from four low-performing banks). Data from Vicinitas consisted of Tweet ID, Text, Name, Screen Name, UTC, Created At, Favorites, Retweets, Language, Client, Tweet Type, URLs, Hashtag, Mentions, Media Type, and Media URLs.

In the cleaning process, there were missing values in some fields. For example, Media Type consisted of "animated_gif," "photo," and "video" only, so "text" was filled into the empty cells. Data (Text) in a Microsoft Excel file were also cleaned—for instance, replacing the % sign with "percentage" (in Thai) and removing newlines, before using it as input for AI for Thai (www.aiforthai.in.th). A researcher had to write a Python code to call the API of AI for Thai in the developer mode and process some steps after retrieving output back from the API to convert the output into the Microsoft Excel format.

AI for Thai is an AI service platform for the Thai language developed by the National Electronics and Computer Technology Center (NECTEC), Thailand. It contains several API natural language processing (NLP) services for researchers and practitioners such as sentiment analysis (S-Sense) to apply NLP in real cases (Khruahong et al., 2020; Tapsai et al., 2019). S-Sense analyzes the message purpose (intent) and gives results in terms of the percentage of confidence to be the sentiment, announcement, request, or question. Intent classification is a branch of text classification focusing on analyzing intents from texts (Pandey et al., 2018). In this study, intent analysis was done at the level of tweets. Images, videos, and animated GIFs were not analyzed because this study focused on NLP only. A tweet's characteristics from Vicinitas and the intent mining results from AI for Thai were later transformed or coded to be input for the next procedure. Features were extracted and grouped as marketer-controllable features and the fundamental features, which marketers cannot change during the communication (tweet) process, as shown in Table 1. Because favorites and retweets are dependent variables, they were coded into zero or one, which means that the tweet was retweeted or not and favorited or not, as same as the study by Alboqami et al. (2015).

4.3 Processing Data

Data-mining techniques for big data consist of classification, clustering, association rules, and prediction (More & Moily, 2021; Nanayakkara et al., 2021). Banks could apply these techniques—for example, classification to detect fraud, clustering to identify customer service classes, association rules

Table 1.
A tweet's characteristics and their codes in this study

Tweet Data	Source	Categorized As/Coded As
Tweet ID	**Vicinitas**	**Running ID**
Marketer-Controllable Characteristics:		
Tweet Type	Vicinitas	Tweet, Retweet, Reply
Media Type	Vicinitas	photo, animated_gif, video, text
Day	Vicinitas	Mon – Sun: Extracted from 'Created At' field
Time_C	Vicinitas	Hr: Extracted from 'Created At' field Morning: Hr >= 6 and < 12 Afternoon: Hr >= 12 and < 18 Evening: Hr >= 18 and < 22 Night: Hr >= 22 or < 6
Hashtag_C	Vicinitas	True: Hashtags > 0 False: Hashtag = 0
Mention_C	Vicinitas	True: Mentions > 0 False: Mentions = 0
Intent_request	AI for Thai	0-100 confidence percentage
Intent_sentiment	AI for Thai	0-100 confidence percentage
Intent_question	AI for Thai	0-100 confidence percentage
Intent_announcement	AI for Thai	0-100 confidence percentage
Intent Cluster (ClusterNo)	Result of Clustering	Cluster_0 – cluster 5: Extracted from intent_request, intent_ sentiment, intent_question, and intent_announcement
Fundamental Characteristics:		
BankHiRev	Classified by researcher	High: 5 high-performing banks Low: 4 lower performing banks
BankCode	Classified by researcher	1–9: Represent tweets from nine banks
Follow_C	Vicinitas	Hi: Followers more than the median value Low: Followers less than or equal the median value
Fav_C	Vicinitas	True: Favorites > 0 False: Favorites = 0
RT_C	Vicinitas	True: Retweets > 0 False: Retweets = 0

to uncover relationships in data, and prediction to predict fraud (More & Moily, 2021). Association mining and clustering are unsupervised approaches, whereas classification is a kind of supervised algorithm (Nanayakkara et al., 2021; Soofi & Awan, 2017). For classification, Decision Tree (DT) and K-nearest neighbor (K-NN) have been applied in the previously published literature (Nanayakkara et al., 2021). Other previous studies have shown that SM data analysis for marketing purposes using clustering has increased significantly (Benslama & Jallouli, 2020). Benslama and Jallouli (2020) indicated that banking is one of the fields that benefited from using clustering tools. According to Voican (2020), in the financial sector, k-means were applied to cluster catalog sales and credit scoring, while decision trees were used to predict banking credit and insurance. Twitter and Facebook are the most studied platforms for clustering SM data, whereas K-means are the most used technique among clustering techniques that are mainly used in sentiment, content, text, and algorithmic analyses (Benslama & Jallouli, 2020; Nanayakkara et al., 2021). It is a method that is easy to understand and has a time complexity lower than others (Mehta & Dang, 2011).

In this process, association rules using the frequent pattern growth (FP-growth) algorithm are first applied to uncover interesting relationships between features or intents (intent_request, intent_sentiment, intent_question, and intent_announcement) of the posts being favorited and retweeted. There are 4,535 FAV tweets and 2,984 RT tweets. Association rules consist of two steps: the measurement of frequency feature and rule generation (Pong-Inwong & Songpan, 2019). Association rules are generated by analyzing data for frequent if/ then patterns and evaluating the criteria support and confidence to reveal the most important relationships. Confidence shows the number of times the if/ then statements are true, whereas support indicates how frequently the items appear in the data (RapidMiner GmbH, 2021). The FP-Growth algorithm in RapidMiner, compared with other frequent itemset mining algorithms (e.g., the Apriori algorithm), uses only two data scans and thus is applicable even on large datasets RapidMiner GmbH, 2018).

The k-means clustering algorithm is employed to group tweets as clusters by their intents. According to Nanayakkara et al. (2021), it is one of the most frequently used data-mining techniques for social media data. The k-means clustering algorithm defines the collection of objects in one group that are similar between them and dissimilar to objects in other groups. It is the most widely used unsupervised learning algorithm for unlabeled data, which aims at finding the given clusters in data. Each data point is assigned to a cluster based on its (smallest) distance between the centroid of that cluster and the data point (Garg & Rani, 2017; Li & Liu, 2010; Patil et al., 2014).

Lastly, Decision Tree (DT), Naïve Bayes (NB), and K-nearest neighbor (K-NN) are applied to classify two classes (favorited vs. non-favorited messages), and another two classes (retweeted vs non-retweeted posts). According to Gong (2022), top machine learning algorithms for classification problems are logistic regression, decision trees, random forest, support vector machine, k-nearest neighbor, and naive Bayes. DT, NB, and K-NN have been widely used in previous studies on SM and text mining (Ashari et al., 2013; Bayhaqy et al., 2018; Okazaki et al., 2015; Rasjid & Setiawan, 2017; Zulfikar et al., 2017). All features/characteristics except Tweet ID and Intent percentage (intent_request, intent_sentiment, intent_question, and intent_announcement) are applied as the marketer-controllable feature set and the fundamental feature set. DT algorithms are the most commonly used technique in classification (Soofi & Awan, 2017). DT linearly divides data using limits in attributes and generates a decision tree.

The division is chosen based on a metric such as data entropy (Okazaki et al., 2015). DT descriptions are most widely used with logic methods. A decision tree is a flowchart-like structure, where each node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes show class distributions (Ashari et al., 2013; Bayhaqy et al., 2018; Jadhav & Channe, 2016). The Naïve bays classifier is considered to be the fastest classifier, highly scalable, and can handle several types of data (Pedamkar, 2022). It is an intuitive technique based on Bayes' probability laws. Each feature contributes to the model information (Ashari et al., 2013; Okazaki et al., 2015; Patil et al., 2014; Zulfikar et al., 2017. NB considers the presence or absence of a particular feature of a class independently from the presence or absence of other features when the class variable is given (Jadhav & Channe, 2016; Rasjid & Setiawan, 2017). K-NN is effective for large training data and robust to noises (Soofi & Awan, 2017). It classifies an element according to its neighbors using lazy-learning algorithms. Depending on the K value, it considers the K-NN and evaluates the value of the data instance, which is not classified. K-NN works as follows:

1. Initialize the value of K
2. Calculate distance between input and training samples
3. Sort the distances
4. Take top K-nearest neighbors
5. Apply the simple majority
6. Predict the class label with more neighbors for input samples (Bayhaqy et al., 2018; Jadhav & Channe, 2016; Okazaki et al., 2015; Rasjid & Setiawan, 2017).

The classification phases consist of finding a model from a labeled dataset and applying the model to a new or unseen dataset (Ashari et al., 2013). The classification results from the training data are applied to the testing data in the next procedure.

5. EVALUATION

Support, confidence, and lift are calculated for association rules. A support value is the statistical probability of the co-occurrence of items in a transaction. Association rules, in which their supports are greater than a user-specified value, are concluded to have minimum support (Pong-Inwong & Songpan, 2019). A confidence value is the probability of seeing the rule's consequent under the condition that the transactions also contain the antecedent. The lift measures how many times one intent and another intent occur together more often than expected if they are statistically independent (Sheikh, Tanveer, & Hamdani, 2004).

For the K-means clustering algorithm, the number of clusters is predefined and selected based on the Elbow method, which runs the K-means clustering algorithm on the dataset for a range of cluster numbers and calculates the sum of squared error (SSE) for each value. The number of clusters is chosen based on the smallest value with a low SSE (Garg & Rani, 2017). The result of the elbow method becomes K in the K-Means clustering algorithm (Bholowalia & Kumar, 2014).

K-fold cross-validation is applied to evaluate the models (Bayhaqy et al., 2018). Ten is the common number of K in the field of applied machine learning (James et al., 2013). The larger K could decrease the difference in size between the training set and the resampling subsets, resulting in a smaller bias (Kuhn & Johnson, 2013). The classifications are assessed using accuracy, precision, recall, F-measure, and area under the receiver operating characteristic curve values (AUC). True positive (TP) is the number of instances correctly classified into their categories. False positive (FP) is the number of instances incorrectly classified into their categories. True negative (TN) is the number of instances correctly classified as not a part of another category. False negative (FN) is the number of instances incorrectly classified as not a part of another category. Accuracy = (TP + TN)/(TP + FP + FN + TN). It shows how much success the algorithm is. Precision = TP/(TP + FP). It measures the situation when an instance, not belonging to a category, is classified as a part of the category. Recall = TP/(TP + FN). It is used to evaluate the situation when an instance is correctly classified into its category. F-measure/ F-score = 2 x ([Precision x Recall]/[Precision + Recall]). It balances two measures, showing the best in terms of precision and recall (Ashari et al., 2013; Bayhaqy et al., 2018; Bilal et al., 2016; Fitri et al., 2019; Ghobakhloo & Ghobakhloo, 2022; Okazaki et al., 2015. ROC (receiver operating characteristic) curve is a graph showing the performance of a classification model at all classification thresholds, whereas AUC measures the area underneath the entire ROC curve telling how much a model is capable to distinguish between classes. AUC values are ranging from 0.0 to 1.0 (Google Inc., n.d.). The higher the AUC, the better the model is at predicting 0s and 1s correctly.

6. RESULTS AND DISCUSSION

In this step, the analysis and synthesis of the research results are conducted. Discussion and implications for both researchers and marketers are summarized and written.

6.1 General Descriptive Statistics

There is a total of 10,000 Bank Twitter official posts from high-performing and low- performing banks equally. The average followers of high-performing banks are 239,114, whereas the average followers of low-performing banks are 2,767. As shown in Table 2, superior banks have an average FAV and RT per tweet more than others. Both groups use a close number of hashtags on average. High profitable banks slightly use mentions more than low profitable banks. In terms of tweet types, both groups mainly employ tweets to communicate with their followers. High-performing banks generally use texts, while low-performing banks incorporate photos in their posts. Both groups normally tweet on Friday at nighttime and highly apply sentiments as a communication strategy.

Table 3 shows tweets' characteristics and engagement in terms of FAV and RT. MGC in the form of tweets receives the highest FAV and RT per message. Although most banks normally tweet messages with texts or photos on Friday at nighttime, tweets with videos on Sunday morning received the highest FAV and RT on average.

Table 2.
Descriptive statistics regarding the communication strategy in high-performing and low-performing banks

Bank Group/Avg.	Avg. Favorites	Avg. Retweets	Avg. Hashtags	Avg. Mentions
High performing	*27*	*22*	*1*	*1*
Low performing	5	2	*1*	0
Bank Group\ Tweet Type	Reply	Retweet	Tweet	Total Tweets
High performing	1,946 (38.92%)	143 (2.86%)	*2,911 (58.22%)*	5,000 (100%)
Low performing	511 (10.22%)	6 (0.12%)	*4,483 (89.66%)*	5,000 (100%)
Bank Group\ Media Type	animated_gif	photo	text	video
High performing	5 (0.1%)	1,435 (28.7%)	*3,433 (68.66%)*	127 (2.54%)
Low performing	2 (0.04%)	*2,594 (51.88%)*	2,335 (46.7%)	69 (1.38%)
Bank Group\ Time	Morning	Afternoon	Evening	Night
High performing	1,533 (30.66%)	1,280 (25.6%)	342 (6.84%)	*1,845 (36.9%)*
Low performing	1,816 (36.32%)	745 (14.9%)	2 (0.04%)	*2,437 (48.74%)*
Bank Group\ Day	Sun	Mon	Tue	Wed
High performing	497 (9.94%)	606 (12.12%)	776 (15.52%)	783 (15.66%)
Low performing	389 (7.78%)	875 (17.5%)	785 (15.7%)	831 (16.62%)
Bank Group\Day	Thu	Fri	Sat	
High performing	785 (15.7%)	*933 (18.66%)*	620 (12.4%)	
Low performing	852 (17.04%)	*883 (17.66%)*	385 (7.7%)	
Bank Group\Intent	Request	Sentiment	Question	Announcement
High performing	1,567 (31.34%)	*3,580 (71.6%)*	1,172 (23.44%)	714 (14.28%)
Low performing	672 (13.44%)	*3,474 (69.48%)*	1,157 (23.14%)	880 (17.6%)

Table 3.
Tweets' characteristics and engagement

Tweet Type/Avg.	Avg. Favorites	Avg. Retweets
Reply	1	1
Retweet	0	0
Tweet	*21*	*16*
Time\ Avg.	Avg. Favorites	Avg. Retweets
Morning	*20*	*18*
Afternoon	16	10
Evening	1	1
Night	14	9
Day\ Avg.	Avg. Favorites	Avg. Retweets
Sun	*23*	*33*
Mon	14	10
Tue	18	9
Wed	14	10
Thu	15	11
Fri	17	11
Sat	14	6
Media Type\ Avg.	Avg. Favorites	Avg. Retweets
animated_gif	10	1
photo	27	16
text	4	3
video	*136*	*180*

6.2 Data Processing

6.2.1 Intent and Association Analyses

For intent mining, all tweets are read from an Excel file and input to the S-Sense API of AI for Thai using Python. The results are first saved in a text file, which is later converted into an Excel format. Outputs from intent analysis using S-Sense consist of intent_request, intent_sentiment, intent_question, and intent_announcement in the form of confidence percentage. Tweet data are classified into one or more intent categories if these percentages are higher than 50. Intent categories are not mutually exclusive. The overview of the intended strategy of each bank is shown in Table 4. Sentiments are still the main intent as one of the communication strategies of all banks.

For the next phase, raw data in terms of the confidence percentage of intents are used as input for association analysis to discover patterns or co-occurrences between intent types of 4,535 favorited messages and 2,984 retweeted messages. They are converted into binomial data before mining association rules. Association rules guide which ones (intents) should go together for effective e-WOM (Weerawatnodom et al., 2017). To identify meaningful frequent itemsets, rules with lift values of more than 1, minimum confidence equal to 0.9, and minimum support equal to 0.01 are applied, as same as the past research (Weerawatnodom et al., 2017). The summary of association rules for FAV and RT is presented in Table 5. All rules are associated with the sentiment. According

Table 4.
Tweet intent types classified by banks

Bank No./Intent	Request	Sentiment	Question	Announcement	Total Tweets
1	546 (54.6%)	*794 (79.4%)*	373 (37.3%)	91 (9.1%)	1,000 (100%)
2	371 (37.1%)	*676 (67.6%)*	129 (12.9%)	105 (10.5%)	1,000 (100%)
3	42 (4.2%)	*377 (37.7%)*	33 (3.3%)	107 (10.7%)	1,000 (100%)
4	292 (29.2%)	*887 (88.7%)*	291 (29.1%)	187 (18.7%)	1,000 (100%)
5	316 (31.6%)	*846 (84.6%)*	346 (34.6%)	224 (22.4%)	1,000 (100%)
6	182 (9.6%)	*1,227 (64.9%)*	557 (29.4%)	387 (20.5%)	1,892 (100%)
7	205 (10.8%)	*1,264 (66.8%)*	224 (11.8%)	257 (13.6%)	1,892 (100%)
8	198 (25.4%)	*625 (80%)*	225 (28.8%)	184 (23.6%)	781 (100%)
9	87 (20%)	*358 (82.3%)*	151 (34.7%)	52 (12%)	435 (100%)

Table 5.
Discovered association rules for FAV and RT

Association Rule for FAV	Support	Confidence	Lift
Intent_question, intent_request, intent_announcement => intent_sentiment	0.019	0.966	1.250
Intent_request, intent_announcement => intent_sentiment	0.035	0.935	1.210
Intent_question, intent_announcement => intent_sentiment	0.052	0.929	1.201
Association Rule for RT	Support	Confidence	Lift
Intent_question, intent_request, intent_announcement => intent_sentiment	0.027	0.976	1.213
Intent_request, intent_announcement => intent_sentiment	0.048	0.947	1.177
Intent_question, intent_announcement => intent_sentiment	0.072	0.935	1.163
Intent_announcement => intent_sentiment	0.197	0.902	1.121
Intent_question, intent_request => intent_sentiment	0.116	0.901	1.121

to the FAV model, if banks post tweets with a question, request, or announcement, they are likely to add a sentiment. FAV tweets with requests or questions and announcements also tend to include sentiments. For the RT model, the first three rules are the same as the rules in the FAV model. In addition, RT tweets with announcements frequently have sentiments and those tweets with questions and requests comprise sentiments. Most confidence measures are high, showing the strength of the associations (Mosley, 2012).

6.2.2 Clustering Social Media Strategy by Intent

The confidence percentage of intents also becomes raw data for the k-means clustering algorithm. The number of clusters is selected using the Elbow method. Eighteen iterations are conducted from k = 2 to k = 20 to find the optimal number of k for the k-means. Figure 2 shows the scatter plot of k on the x axis and the SSE or average within centroid distance on the y axis. Seven is the value at the elbow of the arm, which has the best cluster distance performance (Garg & Rani, 2017). The result of clustering 10,000 tweets by intent is represented in Table 6. Cluster 0 to cluster 6 represent tweet patterns with different confidence percentages of each intent type, which are tweets with high request and sentiment (C0: HiReq_HiSenti); tweets with high sentiment and announcement (C1:

Figure 2.
Finding the number of clusters with the elbow method

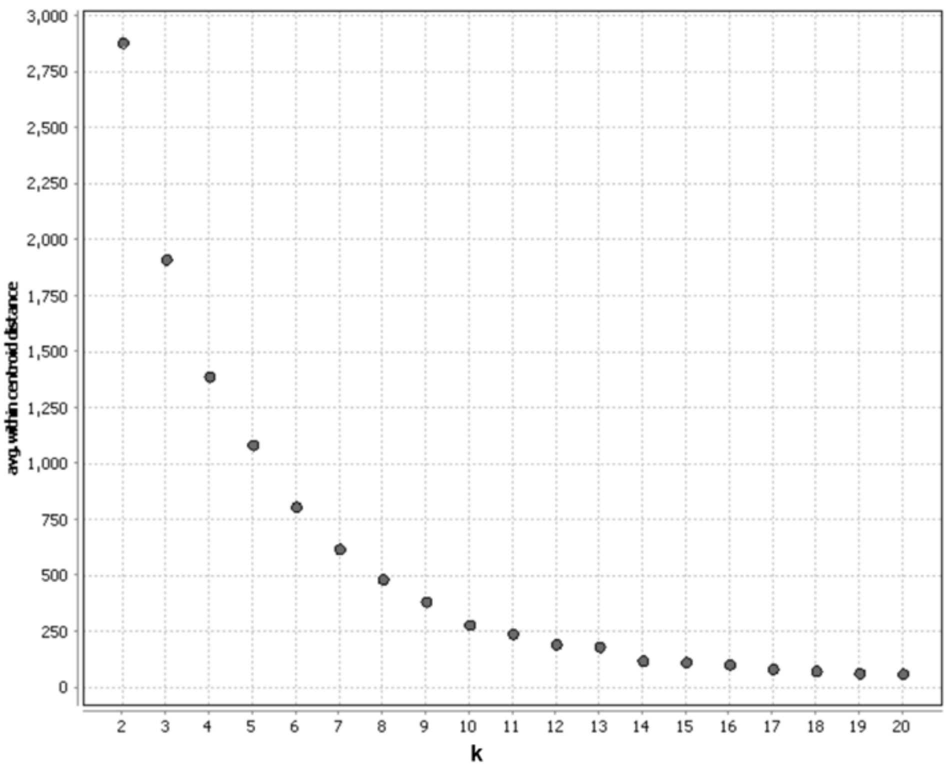


Table 6.
The centroid table of clusters

Cluster/Intent	intent_request	intent_sentiment	intent_question	intent_announcement
cluster_0	75.240	78.672	0.000	5.615
cluster_1	1.180	80.500	0.407	69.222
cluster_2	73.993	83.353	83.053	11.779
cluster_3	7.660	0.000	0.000	6.945
cluster_4	18.931	0.000	81.158	8.924
cluster_5	0.000	82.529	81.938	13.840
cluster_6	0.000	77.830	0.000	0.000

HiSenti_HiAnn); tweets with high request, sentiment, and question (C2: HiReq_HiSenti_HiQues); tweets with low request, sentiment, question, and announce (C3: LoAll); tweets with high question (C4: HiQues); tweets with high sentiment and question (C5: HiSenti_HiQues); and tweets with high sentiment (C6: HiSenti), respectively. C6, C3, and C0 are the top three content strategies used by Thai commercial banks.

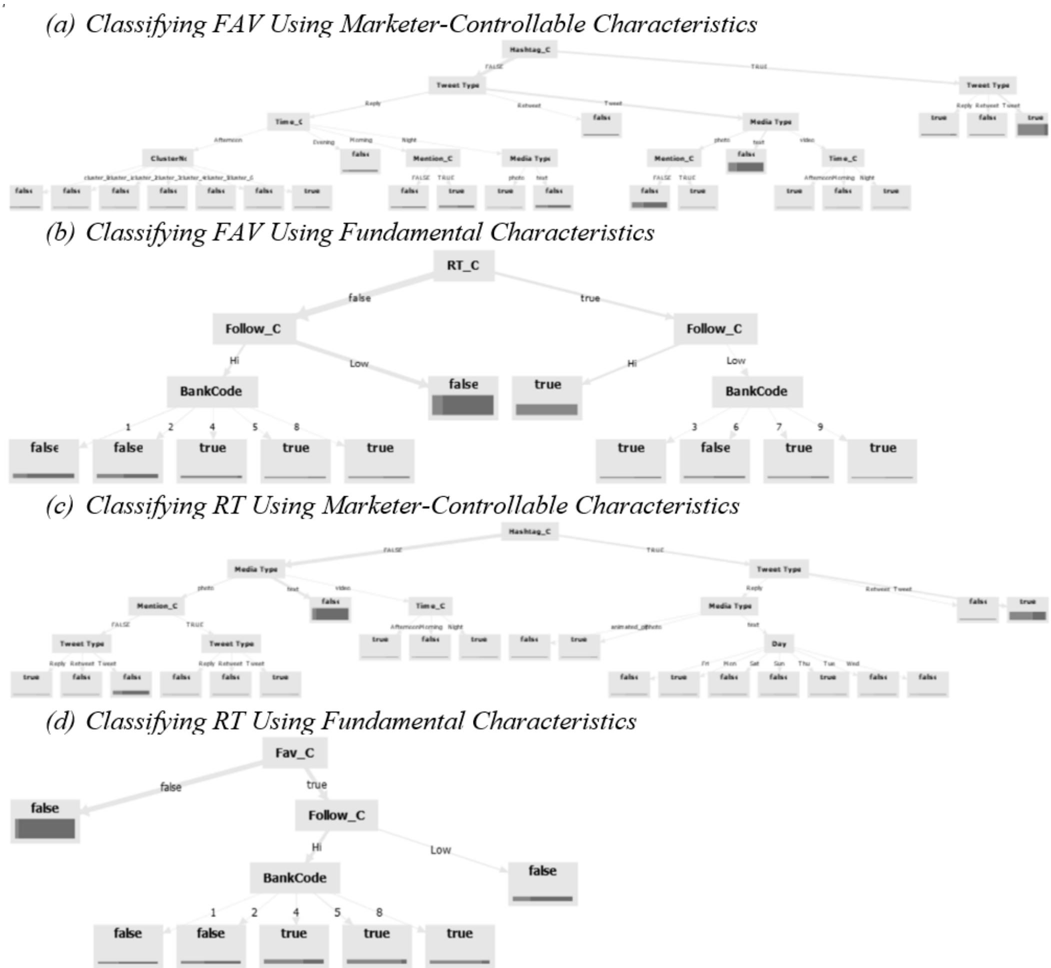
Table 7.
Comparing the model performance of each mining techniques

Labeled Attribute: FAV Prediction Attributes: Marketer-Controllable Characteristics	Accuracy	Precision	Recall		AUC
*Decision Tree	*0.719*	0.719	0.798	*0.756*	*0.749*
Naïve Bayes	0.713	0.707	*0.811*	0.755	0.718
K-Nearest Neighbor	0.624	*0.757*	0.460	0.572	0.696
Random Forest	0.719	0.705	0.837	0.765	0.779
*Gradient Boosted Trees	*0.724*	0.701	*0.865*	*0.774*	*0.787*
Labeled attribute: FAV Prediction attributes: Fundamental characteristics	Accuracy	Precision	Recall	F-Measure	AUC
**Decision Tree	*0.827*	0.784	*0.943*	*0.856*	*0.863*
Naïve Bayes	0.766	*0.800*	0.762	0.781	0.845
K-Nearest Neighbor	0.454	N/A	0	N/A	0.500
Random Forest	0.825	*0.791*	0.926	0.853	*0.879*
Gradient Boosted Trees	0.823	0.774	*0.956*	0.855	*0.879*
Labeled attribute: RT Prediction attributes: Marketer-controllable characteristics	Accuracy	Precision	Recall	F-Measure	AUC
*Decision Tree	*0.785*	*0.868*	0.818	*0.842*	*0.797*
Naïve Bayes	0.774	0.850	*0.824*	0.837	0.794
K-Nearest Neighbor	0.744	0.866	0.752	0.805	0.765
*Random Forest	*0.789*	0.860	0.836	0.848	*0.823*
Gradient Boosted Trees	0.781	0.801	*0.917*	*0.855*	0.829
Labeled attribute: RT Prediction attributes: Fundamental characteristics	Accuracy	Precision	Recall	F-Measure	AUC
**Decision Tree	*0.858*	0.878	*0.926*	*0.901*	*0.891*
Naïve Bayes	0.832	0.911	0.843	0.876	0.880
K-Nearest Neighbor	0.788	*0.950*	0.736	0.829	0.837
*Random Forest	*0.858*	0.878	*0.926*	*0.901*	0.900
Gradient Boosted Trees	0.856	*0.886*	0.912	0.899	*0.906*

6.2.3 Classification of Customer Engagement Using Twitter Content Strategies

There are 4,535 favorited tweets and 5,465 non-favorite tweets. Of all tweets, 2,984 tweets are retweeted, but 7,016 tweets are not. Fundamental characteristics such as the number of followers possibly strongly affect engagement more than marketer-controllable characteristics. Thus, this study classifies labeled attributes or groups (i.e., FAV and RT) using two sets of prediction attributes (marketer-controllable and fundamental characteristics) to explicitly explore the effect of manageable factors. Marketer-controllable characteristics consist of Tweet Type, Media Type, Day, Time_C, Hashtag_C, Mention_C, and ClusterNo, whereas fundamental factors are composed of BankHiRev, BankCode, Follow_C, and Fav_C or RT_C. Table 7 shows the model performances. Three algorithms (DT, NB, and K-NN) are the focus of this study. The accuracy and F-score of models in this study are comparable with the study by Asali (2021), which used machine learning algorithms to classify tweets' sentiment, and the study by Ligiarta and Ruldeviyani (2022), which analyzed the sentiment of the Indonesia Commuter Line (KRL) using machine learning approaches. After finding that DT is the best algorithm in terms of accuracy for all cases, I also employed two more algorithms relating to DT, Random Forest and Gradient Boosted Trees, to guide future model development for the best performance. Random forest is a collection of DT; it is better than DT in generalization, but less interpretable. It is one of the top six machine-learning algorithms for classification problems (Gong, 2022). Gradient Boosted DT is a strong machine-learning model composed of multiple weak models such as DT (Google.com, n.d.). Compared with DT, Gradient Boosted Trees are better than DT in classifying FAV using marketer-controllable characteristics. Random Forest models are better than DT models in classifying RT using manageable features and are slightly better in classifying RT using fundamental features. However, the DT is an eager learning algorithm that is simple to understand and interpret (Ashari et al., 2013; Zulfikar et al., 2017). It also outperforms in terms of accuracy, F-measure, and AUC in all cases over NB and K-NN; therefore, findings are interpreted based on DT results. According to a guide of AUC for classifying the accuracy of a diagnostic test (Vidya et al., 2015), AUCs of all DT models also have fair or good classifications.

Figure 3.
The decision trees for FAV and RT



As shown in Figures 3(a) and 3(b), Hashtag_C, Tweet Type, Time_C & Media Type, ClusterNo, Mention_C, Media Type, and Time_C are manageable features as the splitting criteria for the FAV tree. Posts that tend to gain favorites include those with no hashtags in reply mode posted in the afternoon using intent cluster 6 (HiSenti); posts with no hashtags in reply mode posted in the morning with mentions, or posts with no hashtags in reply mode posted in the nighttime using photos; posts with no hashtags in tweet mode using photos having mentions; and posts with no hashtags in tweet mode using video posting in the afternoon or night. Posts with hashtags in reply or tweet modes also possibly receive favorites. In terms of fundamental characteristics, posts with no RT of accounts with high followers of banks no. 4, no. 5, and no. 8; posts with RT of accounts with high followers; and posts with RT of accounts with low followers of banks no. 3, no. 7, and no. 9 are more likely to receive favorites.

As shown in Figures 3(c) and 3(d), Hashtag_C, Media Type & Tweet Type, Mention_C, Time_C, Media Type, and Tweet Type & Day are manageable features as the splitting criteria for the RT tree. Posts with no hashtags using photos with no mentions in reply mode; posts with no hashtags using photos with mentions in tweet mode; posts with no hashtags using video posting in the afternoon; posts with hashtags in reply mode using photos; posts with hashtags in reply mode using texts posting on Monday or Thursday; and posts with hashtags in tweet mode tend to gain retweets. With regard to fundamental features, posts having favorites of accounts with high followers of banks no. 4, no. 5, and no. 8 possibly gain more retweets.

6.3 Discussion

Senadheera et al. (2011) stated that small banks are less active on Twitter. Jaman et al. (2020) indicated that SM may not be suitable for all businesses. My work showed that high-performing banks have both a higher number of followers, average FAVs per tweet, and average RTs per tweet more than low-performing ones. Surucu-Balci et al. (2020) specified that the content type and the existence of call-to-action of container lines' Tweets significantly influence their stakeholders' engagement rate. Users showing sentiment on tweets will increase their social influence receiving partial support (Lahuerta-Otero & Cordero-Gutiérrez, 2016). This study showed the various combination of content types (intent types, particularly sentiment) from association rules in favorited or retweeted messages. The clustering of tweets by intent is also one of the splitting criteria in the FAV model. Similar to the study by Weerawatnodom et al. (2017) in which hashtags and pictures are a part of the association rules for the RT and FAV models, this study explored them as features for the classification algorithms.

For classification, compared with NB and K-NN, DT yields the best results in terms of accuracy, just as it did in previous studies (Bayhaqy et al., 2018; Karim & Rahman, 2013; Soni, Ansari, Sharma, & Soni, 2011). Nevertheless, findings show the contrast results from the study by Jadhav and Channe (2016), indicating that K-NN is better than NB and DT in both the segment challenge case with the medium-sized dataset and the supermarket case with a large dataset. Bilal et al. (2016) showed NB performs best in classifying Roman Urdu opinions, and Ashari et al. (2013) revealed that NB outperforms DT and K-NN. Weerawatnodom et al. (2017) indicated that videos, mentions, discount or promotion information, social news update, and event news significantly contributes to the RT model, whereas discount or promotion information, interaction with customers, and event news are significant predictors of the FAV model. The study by Alboqami et al. (2015) also pointed out that pictures, mentions, product or service information, and direct answers to customers significantly predict RTs, while pictures, hashtags, product or service information, interactions with customers, and direct answers to customers significantly predict FAVs. The use of hashtags and mentions on tweets significantly increases social users' influence (Lahuerta-Otero & Cordero-Gutiérrez, 2016). In this study, hashtags, tweet types, intent clusters, and mentions relate to FAVs or RTs as well. Hashtags are the main split criterion in both FAV and RT models, conforming to the study by Lahuerta-Otero & Cordero-Gutiérrez (2016) suggesting that influencers generally on average use more hashtags and mentions.

7. THEORETICAL CONTRIBUTION AND PRACTICAL IMPLICATIONS

For theoretical contribution, according to the complexity of the Thai language and the importance of automated analysis mentioned in Mosley (2012), this study showed the application of AI for Thai (S-Sense) to identify intent types. Four intent types and intent clusters could be used in further research as features or independent variables to classify or predict the engagement of SM fans. The present study also shows novel approaches to discovering MGC and exploring fans' reactions to brand communication. Firms are strongly encouraged to apply data-mining techniques, for instance, to identify prosumers who are energetic endorsers of positive feedback in online environments (Okazaki et al., 2015). These data-mining exercises could be a guideline for other studies to analyze MGC for companies in other sectors.

For practical implications, the knowledge from this study could aid banks in effectively developing their communication strategies on Twitter to increase positive engagement. Findings showed that high-performing banks receive more FAV and RT than low-performing banks. Popular messages of both high-performing and low-performing banks generally are tweets, replies, and retweets, respectively. Media types of popular posts are texts/photos, videos, and animated_gif, respectively. Popular tweets of both groups appear in the night, morning, afternoon, and evening. Both groups frequently post on Friday and rarely post on Sunday and use the sentiment as the main intent in their posts. In sum, beloved posts of both high-performing and low-performing banks have quite similar characteristics. The feature "BankHiRev" is also not a significant classifier in either FAV or RT models.

In addition, findings revealed that sentiment is the most prevalent intent type employed by all banks. Sentiment with other intent types—for example, "question + request + announcement + sentiment," "request + announcement + sentiment," and "question + announcement + sentiment"—is a communication strategy, drawing both FAVs and RTs. These intents could be also grouped tweets into seven clusters

- **cluster_0:** Tweets with request and sentiment
- **cluster_1:** Tweets with sentiment and announcement
- **cluster_2:** Tweets with request, sentiment, and question
- **cluster_3:** Tweets with no intents
- **cluster_4:** Tweets with question
- **cluster_5:** Tweets with sentiment and question
- **cluster_6:** Tweets with sentiment.

According to the association rules, announcement and sentiment (cluster_1) and question, request, and sentiment (cluster_2) generate retweets as well. Using these clusters together with other features in tweet messages (both marketer-controllable and marketer-uncontrollable features) shows that hashtag, tweet type, media type, time, cluster, and mention can be classified between FAV and non-FAV messages, whereas hashtag, media type, tweet type, mention, time, and day are good classifiers for RT and non-RT messages. The number of followers and bank code could indicate FAV and RT posts as well. The number of RT and FAV also influence each other.

Based on the DT model, both bank types could employ reply posts with sentiment posted in the afternoon; reply posts with mentions posted in the morning; reply posts with photos posted in the nighttime; tweet posts with photos and mentions; tweet posts with videos posted in the afternoon or night; or reply or tweet posts with hashtags as strategies to receive favorites. Banks no. 4, no. 5, and no. 8 with high followers and banks no. 3, no. 7, no. 9 with low followers could gain favorites. To enhance the chance of retweets, bank marketers should use replies or tweets with photos; tweets with photos and mentions; posts with videos posted in the afternoon; reply tweets with photos and hashtags; reply tweets with text and hashtags posted on Monday or Thursday; or tweets with hashtags. Bank no. 4, no. 5, and no. 8 with high followers and their messages received favorites tend to gain retweets. For using the models to predict future engagement, banks could apply other algorithms such as Gradient Boosted Trees or Random Forests and compare them with the DT results to generate models with better performance.

8. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This study contributed to the literature of e-WOM on SM as follows. First, this study shared the example application of two tools—a Twitter analysis tool (Vicinitas) and a sentiment analysis tool (S-Sense by AI for Thai)—to automatically collect and process Twitter data. Second, this work applied novel methods: data mining, to find the associations among message intents, to automatically group messages by intents, and to identify common characteristics of retweeted and favorited MGC tweets. Findings showed the effectiveness of both marketer-controllable and fundamental features in representing e-WOM. Third, although user intents are often explicitly expressed in tweets, few studies extracted intent categories of tweets in the context of commerce marketing (Wang et al., 2015). To our knowledge, this is the first large-scale study to explore the insights regarding MGC intents in the financial sector, to combine them with other tweet characteristics to investigate the SM strategies, and to highlight the features crucial for Thai commercial banks. Last, this paper guides future research about automated and generalizable processes to develop and evaluate models. These processes could be applied to various industries or academics through their Twitter official accounts to determine the effective MGC.

This research has some limitations that could be a direction for future work. First, this study has imbalanced RT cases. The non-RT class dominates the datasets, which may influence the classification model, particularly in NB (Zulfikar et al., 2017). Future work should collect data from both classes equally. Second, only Twitter data from Thai commercial banks is collected. Although this study covered most private commercial banks in Thailand, future research should examine banks in other Southeast Asian countries to increase the generalizability of findings and to compare strategies of banks in the same region. Third, only two forms of e-WOM, FAV and RT, were considered. Other forms of engagement such as replies/comments or shares should be captured as well. Fourth, although using an automated tool to analyze tweet content decreases human errors and time-consuming problems, future research should use content analysis to extract interesting aspects from MGC tweets and use this data to later train classification models. Last, this work focused on only MGC; future work should expand the scope to understand the effectiveness of both tweets generated by marketers and followers. Other tweet characteristics and data-mining algorithms should be also applied to improve the analytical performance.

COMPETING INTERESTS

All authors of this article declare there are no competing interest.

FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Funding for this research was covered by the author of the article.

REFERENCES

Abbas, M. J., Khalil, L. S., Haikal, A., Dash, M. E., Dongmo, G., & Okoroha, K. R. (2021). Eliciting emotion and action increases social media engagement: An analysis of influential orthopaedic surgeons. *Arthroscopy, Sports Medicine, and Rehabilitation*, *3*(5), e1301–e1308. doi:10.1016/j.asmr.2021.05.011 PMID:34712967

Abunadi, I. (2015). Characteristics of electronic integrated system and trust in the provider of service. *International Journal of Computers and Applications*, *132*(4), 23–31. doi:10.5120/ijca2015907410

Ahmad, S. R., & Khan, M. N. (2021). Efficiency measurement of Indian banking industry: An empirical comparative analysis. *International Journal of Financial Research*, *12*(4), 135–145. doi:10.5430/ijfr.v12n4p135

Ajina, A. S. (2019). The role of social media engagement in influencing customer loyalty in Saudi banking industry. *International Review of Management and Marketing*, *9*(3), 87–92. doi:10.32479/irmm.8060

Akhlaq, A., Ali, W., & Gul, K. (2021). The impact of using social networking sites at work on organizational knowledge. *International Journal of Asian Business and Information Management*, *12*(3), 347–365. doi:10.4018/IJABIM.20210701.oa21

Alamsyah, A., & Indraswari, A. A. (2017). Social network and sentiment analysis for social customer relationship management in Indonesia banking sector. *Advanced Science Letters*, *23*(4), 3808–3812. doi:10.1166/asl.2017.9279

Alboqami, H., Al-Karaghouli, W., Baeshen, Y., Erkan, I., Evans, C., & Ghoneim, A. (2015). Electronic word of mouth in social media: The common characteristics of retweeted and favourited marketer-generated content posted on Twitter. *International Journal of Internet Marketing and Advertising*, *9*(4), 338–358. doi:10.1504/IJIMA.2015.072886

Alotaibi, M. S. (2013). The impact of twitter on Saudi banking sectors in the presence of social media: An evaluative study. *International Research: Journal of Library and Information Science*, *3*(4), 618–630.

Anshari, M., Almunawar, M. N., & Masri, M. (2020). Financial technology and disruptive innovation in business: Concept and application. *International Journal of Asian Business and Information Management*, *11*(4), 29–43. doi:10.4018/IJABIM.2020100103

Asali, A. G. (2021). Social media analysis for investigating consumer sentiment on mobile banking. *Journal of International Conference Proceedings*, *4*(2), 241–253. doi:10.32535/jicp.v4i2.1247

Ashari, A., Paryudi, I., & Tjoa, A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications*, *4*(11). Advance online publication. doi:10.14569/IJACSA.2013.041105

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). *Sentiment analysis about e-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes* [Paper presentation]. *2018 International Conference On Orange Technologies (ICOT)*. Nusa Dua, Bali, Indonesia. doi:10.1109/ICOT.2018.8705796

Benslama, T., & Jallouli, R. (2020). *Clustering of social media data and marketing decisions* [Paper presentation]. Digital Economy. Emerging Technologies and Business Innovation: 5th International Conference on Digital Economy. ICDEc 2020, Bucharest, Romania. https://link.springer.com/book/10.1007/978-3-030-64642-4

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computers and Applications*, *105*(9), 17–24.

Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University-Computer and Information Sciences*, *28*(3), 330–344. doi:10.1016/j.jksuci.2015.11.003

Bohlin, E., Shaikh, A. A., & Hanafizadeh, P. (2018). Social network banking: A case study of 100 leading global banks. *International Journal of E-Business Research*, *14*(2), 1–13. doi:10.4018/IJEBR.2018040101

Botchway, R. K., Jibril, A. B., Oplatková, Z. K., & Chovancová, M. (2020). Deductions from a Sub-Saharan African bank's tweets: A sentiment analysis approach. *Cogent Economics & Finance*, *8*(1), 1776006. doi:10.1080/23322039.2020.1776006

Boyd, D., Golder, S., & Lotan, G. (2010, January 5–8). *Tweet, tweet, retweet: Conversational aspects of retweeting on twitter* [Paper presentation]. *43rd Hawaii International Conference on System Sciences*, Honolulu, HI, USA.

Cahyono, E. F., Rani, L. N., & Kassim, S. (2020). Perceptions of the 7P marketing mix of Islamic banks in Indonesia: What do Twitter users say about it? *International Journal of Innovation*, *Creativity and Change*, *11*(11), 300–319.

Chahine, S., & Malhotra, N. K. (2018). Impact of social media strategies on stock price: The case of Twitter. *European Journal of Marketing*, *52*(7/8), 1526–1549. doi:10.1108/EJM-10-2017-0718

Chikandiwa, S. T., Contogiannis, E., & Jembere, E. (2013). The adoption of social media marketing in South African banks. *European Business Review*, *25*(4), 365–381. doi:10.1108/EBR-02-2013-0013

Chu, S.-C., & Sung, Y. (2015). Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter. *Electronic Commerce Research and Applications*, *14*(4), 251–260. doi:10.1016/j.elerap.2015.04.002

Cosimato, S., & Troisi, O. (2015). Stakeholder engagement and social media communication in banking industry: Monte dei Paschi di Siena case study. *Journal of Business and Economics*, *6*(7), 249–258. doi:10.15341/jbe(2155-7950)/07.06.2015/005

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from Twitter social networks. *Marketing Science*, *35*(3), 343–362. doi:10.1287/mksc.2015.0968

Dwivedi, S., Kasliwal, P., & Soni, S. (2016). Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). In *Proceedings of the 2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. IEEE. doi:10.1109/CDAN.2016.7570894

Ehrmann, M., & Wabitsch, A. (2022, April). Central bank communication with non-experts–A road to nowhere? *Journal of Monetary Economics*, *127*, 69–85. doi:10.1016/j.jmoneco.2022.02.003

Farina, V., Gabbi, G., & Previati, D. (2014). Good news, bad news: A proposal to measure banks' reputation using Twitter. In T. Lindblom, S. Sjögren, & M. Willesson (Eds.), *Governance, regulation and bank stability* (pp. 242–259). Palgrave Macmillan. doi:10.1057/9781137413543_11

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of anti-LGBT campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest algorithm. *Procedia Computer Science*, *161*, 765–772. doi:10.1016/j.procs.2019.11.181

Garg, N., & Rani, R. (2017). Analysis and visualization of Twitter data using k-means clustering. In *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE. doi:10.1109/ICCONS.2017.8250547

Ghobakhloo, M., & Ghobakhloo, M. (2022). Design of a personalized recommender system using sentiment analysis in social media (case study: Banking system). *Social Network Analysis and Mining*, *12*(1), 84. doi:10.1007/s13278-022-00900-0

Gong, D. (2022, February 23). *Top 6 machine learning algorithms for classification: How to build a machine learning model pipeline in Python*. Retrieved from https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501

Google.com. (n.d.). *Gradient Boosted Decision Trees*. Retrieved from https://developers.google.com/machine-learning/decision-forests/intro-to-gbdt

Google Inc. (n.d.). *Classification: ROC Curve and AUC*. Retrieved from https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

Grover, P., & Kar, A. K. (2020). User engagement for mobile payment service providers–Introducing the social media engagement model. *Journal of Retailing and Consumer Services*, *53*, 101718. doi:10.1016/j.jretconser.2018.12.002

Gunarathne, P., Rui, H., & Seidmann, A. (2015). Customer service on social media: The effect of customer popularity and sentiment on airline response. In *Proceedings of the 2015 48th Hawaii International Conference on System Sciences*. IEEE. doi:10.1109/HICSS.2015.397

Gunarathne, P., Rui, H., & Seidmann, A. (2018). When social media delivers customer service: Differential customer treatment in the airline industry. *Management Information Systems Quarterly*, *42*(2), 489–520. doi:10.25300/MISQ/2018/14290

Hafez, M. (2021). The role of social media marketing on overall brand equity in the telecommunication sector in Bangladesh: A moderated mediation model of brand love and value co-creation. *International Journal of Asian Business and Information Management*, *12*(3), 1–15. doi:10.4018/IJABIM.294102

Hamroun, M., & Gouider, M. S. (2020). A survey on intention analysis: Successful approaches and open challenges. *Journal of Intelligent Information Systems*, *55*(3), 423–443. doi:10.1007/s10844-020-00604-x

Hamzah, A., & Hidayatullah, A. F. (2018). Clustering on Twitter: Case study Twitter account of higher education institution in Indonesia. *MATEC Web of Conferences, 154*. doi:10.1051/matecconf/201815403010

Hoffman, D. L., & Fodor, M. (2010, October 1). Can you measure the ROI of your social media marketing? *MIT Sloan Management Review*, *52*, 41.

Hollerit, B., Kröll, M., & Strohmaier, M. (2013, May 13–17). Towards linking buyers and sellers: Detecting commercial intent on Twitter. In *WWW '13: Proceedings of the 22nd International Conference on World Wide Web*. Association for Computing Machinery. doi:10.1145/2487788.2488009

Hougaard, S. A. (2017). *Tweeting for a cause: A content analysis of successful charitable nonprofits' publishing strategies on Twitter* (28104218) [Master of Arts Thesis]. Brigham Young University. Retrieved from https://scholarsarchive.byu.edu/etd/6274

Icha, O., & Edwin, A. (2015). Effectiveness of social media networks as a strategic tool for organizational marketing management. *Journal of Internet Banking and Commerce*, *S2*(006), 1–9. doi:10.4172/1204-5357.S2-006

Illia, L., Colleoni, E., & Meggiorin, K. (2021). How infomediaries on Twitter influence business outcomes of a bank. *International Journal of Bank Marketing*, *39*(5), 709–724. doi:10.1108/IJBM-08-2020-0414

Jadhav, S. D., & Channe, H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Scientific Research*, *5*(1), 1842–1845.

Jaman, S. F. I. H., Damit, N. J. H., Ishak, N. A., Ason, M. L. A., Tamin, M. R., Tangphadungrutch, K., & Almunawar, M. N. (2020). The adoption of social media as marketing tools: Case small and medium enterprises in Brunei Darussalam. *International Journal of Asian Business and Information Management*, *11*(2), 28–50. doi:10.4018/IJABIM.2020040103

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer. doi:10.1007/978-1-4614-7138-7

Karim, M., & Rahman, R. M. (2013). Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing. *Journal of Software Engineering and Applications*, *6*(4), 196–206. Advance online publication. doi:10.4236/jsea.2013.64025

Kaur, N., & Kiran, R. (2015). E-banking service quality and customer loyalty: Changing dynamics of public, private and foreign bank consumers in India. *Global Business and Management Research*, *7*(1), 74–92.

Kaura, V. (2013). Antecedents of customer satisfaction: A study of Indian public and private sector banks. *International Journal of Bank Marketing*, *31*(3), 167–186. doi:10.1108/02652321311315285

Khajeheian, D., & Mirahmadi, F. (2015). Social media, traditional media and marketing communication of public relations: A study of banking industry. *American Journal of Marketing Research*, *1*(2), 79–87.

Khanum, M. A., Nagrami, S. A., & Trivedi, M. C. (2016, March 12). *Use of social media to drive business advantage in banking* [Paper presentation]. Advancement in Computer Engineering and Information Technology (ACEIT-2016), Lucknow, India.

Khder, M. A., Abu-Alsondos, I. A., & Bahar, A. Y. (2021). The impact of implementing data mining in business intelligence. *International Journal of Entrepreneurship*, *25*(4S), 1–7.

Khruahong, S., Asawasakulson, A., & Krom, W. N. (2020, Oct. 25–28, 2020). Social media analytics in comments of multiple vehicle brands on social networking sites in Thailand. In Y. Luo, (Ed.). *Cooperative design, visualization and engineering (CDVE2020)*. Springer. doi:10.1007/978-3-030-60816-3_39

Kim, E., Sung, Y., & Kang, H. (2014). Brand followers' retweeting behavior on Twitter: How brand relationships influence brand electronic word-of-mouth. *Computers in Human Behavior*, *37*, 18–25. doi:10.1016/j.chb.2014.04.020

Kirakosyan, K. (2014). Managerial perspective on social media implementation in banking industry. Comparative study on Romanian and Mexican Banks. *Revista de Management Comparat Internaţional*, *15*(3), 297–311.

Kordzadeh, N., & Young, D. K. (2022). How social media analytics can inform content strategies. *Journal of Computer Information Systems*, *62*(1), 128–140. doi:10.1080/08874417.2020.1736691

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer., doi:10.1007/978-1-4614-6849-3

Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on Twitter. *Computers in Human Behavior*, *64*, 575–583. doi:10.1016/j.chb.2016.07.035

Lahuerta-Otero, E., Cordero-Gutiérrez, R., & De la Prieta-Pintado, F. (2018). Retweet or like? That is the question. *Online Information Review*, *42*(5), 562–578. doi:10.1108/OIR-04-2017-0135

Li, G., & Liu, F. (2010). A clustering-based approach on sentiment analysis. In *Proceedings of the 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*. IEEE. doi:10.1109/ISKE.2010.5680859

Ligiarta, M. A., & Ruldeviyani, Y. (2022). Customer satisfaction analysis of mobile banking application based on Twitter data. In *Proceedings of the 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*. IEEE. doi:10.1109/ICE3IS56585.2022.10010221

McShane, L., Pancer, E., Poole, M., & Deng, Q. (2021). Emoji, playfulness, and brand engagement on Twitter. *Journal of Interactive Marketing*, *53*(1), 96–110. doi:10.1016/j.intmar.2020.06.002

Mehta, N., & Dang, S. (2011). A review of clustering techiques in various applications for effective data mining. *International Journal of Research in IT & Management*, *1*(2), 50–66.

Menon, R. G. V., Sigurdsson, V., Larsen, N. M., Fagerstrøm, A., Sørensen, H., Marteinsdottir, H. G., & Foxall, G. R. (2019). How to grow brand post engagement on Facebook and Twitter for airlines? An empirical investigation of design and content factors. *Journal of Air Transport Management*, *79*, 101678. doi:10.1016/j.jairtraman.2019.05.002

Mishael, Q., & Ayesh, A. (2020). *Investigating classification techniques with feature selection for intention mining from Twitter feed*. 10.48550/arXiv.2001.10380

Mishra, U. S., Sahoo, K. K., & Patra, S. K. (2010). Service quality assessment in banking industry of India: A comparative study between public and private sectors. *European Journal of Soil Science*, *16*(4), 653–669.

Mitic, M., & Kapoulas, A. (2012). Understanding the role of social media in bank marketing. *Marketing Intelligence & Planning*, *30*(7), 668–686. Advance online publication. doi:10.1108/02634501211273797

More, R., & Moily, Y. (2021). Big data analysis in banking sector. *International Journal of Engineering Research and Applications*, *11*(4), 1–5.

Mosley, R. C., Jr. (2012). *Social media analytics: Data mining applied to insurance Twitter posts* [Paper presentation]. Casualty Actuarial Society E-Forum. https://www.casact.org/sites/default/files/2021-02/pubs_forum_12wforumpt2_mosley.pdf

Mucan, B., & Özeltürkay, E. Y. (2014). Social media creates competitive advantages: How Turkish banks use this power? A content analysis of Turkish banks through their webpages. *Procedia: Social and Behavioral Sciences*, *148*, 137–145. Advance online publication. doi:10.1016/j.sbspro.2014.07.027

Nanath, K., & Joy, G. (2021). Leveraging Twitter data to analyze the virality of COVID-19 tweets: A text mining approach. *Behaviour & Information Technology*, *42*(2), 196–214. doi:10.1080/0144929X.2021.1941259

Nanayakkara, A., Kumara, B. T. G. S., & Rathnayaka, R. M. K. T. (2021). A survey of finding trends in data mining techniques for social media analysis. *Sri Lanka Journal of Social Sciences and Humanities*, *1*(2), 37–50. doi:10.4038/sljssh.v1i2.36

Ocampo, L., Besabella, O., Fallore, M., Guinandal, A. R., Merabueno, A., Himang, C. M., & Yamagishi, K. (2021). An integrated AHP-TOPSIS for evaluating online marketing strategies for the hospitality industry. *International Journal of Asian Business and Information Management*, *12*(3), 1–28. doi:10.4018/IJABIM.20210701.oa11

Okazaki, S., Díaz-Martín, A. M., Rozano, M., & Menéndez-Benito, H. D. (2015). Using Twitter to engage with customers: A data mining approach. *Internet Research*, *25*(3), 416–434. doi:10.1108/IntR-11-2013-0249

Olaleye, S. A., Adeegbe, J. M., Dada, O. A., & Bounab, Y. (2020, April 20–24). *Insight from Nigerian banking customers discussions: A study of contextual semantic search and twitter sentiment analysis* [Paper presentation]. *26th Conference of Open Innovations Association (FRUCT)*, Yaroslavl, Russia.

Pandey, R., Purohit, H., Stabile, B., & Grant, A. (2018, December 3–6). Distributional semantics approach to detect intent in Twitter conversations on sexual assaults. In *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE. doi:10.1109/WI.2018.00-80

Park, Y.-E., & Javed, Y. (2020). Insights discovery through hidden sentiment in big data: Evidence from Saudi Arabia's financial sector. *The Journal of Asian Finance*, *Economics and Business*, *7*(6), 457–464. doi:10.13106/jafeb.2020.vol7.no6.457

Patil, P. H., Thube, S., Ratnaparkhi, B., & Rajeswari, K. (2014). Analysis of different data mining tools using classification, clustering and association rule mining. *International Journal of Computers and Applications*, *93*(8), 35–39. doi:10.5120/16238-5766

Pedamkar, P. (2022). *Classification algorithms*. Retrieved from https://www.educba.com/classification-algorithms/

Pedrood, B., & Purohit, H. (2018). Mining help intent on twitter during disasters via transfer learning with sparse coding. In R. Thomson, C. Dancy, A. Hyder, & H. Bisgin (Eds.), Lecture Notes in Computer Science: Vol. 10899. *Social, Cultural, and Behavioral Modeling. BP-BRiMS 2018*. Springer. doi:10.1007/978-3-319-93372-6_16

Pong-Inwong, C., & Songpan, W. (2019). Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining. *International Journal of Machine Learning and Cybernetics*, *10*(8), 2177–2186. doi:10.1007/s13042-018-0800-2

Purohit, H., Dong, G., Shalin, V., Thirunarayan, K., & Sheth, A. (2015, December 19–21). Intent classification of short-text on social media. In *Proceedings of the 2015 IEEE International Conference on Smart City/Socialcom/Sustaincom (SmartCity)*. IEEE. doi:10.1109/SmartCity.2015.75

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, *29*(6), 2438–2447. doi:10.1016/j.chb.2013.05.007

Rantanen, A., Salminen, J., Ginter, F., & Jansen, B. J. (2019). Classifying online corporate reputation with machine learning: A study in the banking domain. *Internet Research*, *30*(1), 45–66. doi:10.1108/INTR-07-2018-0318

RapidMiner GmbH. (2018). *FP-Growth (RapidMiner Studio Core).* Retrieved from https://docs.rapidminer.com/8.0/studio/operators/modeling/associations/fp_growth.html

RapidMiner GmbH. (2021). *Create Association Rules (RapidMiner Studio Core).* Retrieved from https://docs.rapidminer.com/latest/studio/operators/modeling/associations/create_association_rules.html#:~:text=lift%3A%20The%20lift%20of%20a,independence%20are%20X%20and%20Y

Rasjid, Z. E., & Setiawan, R. (2017). Performance comparison and optimization of text document classification using k-NN and Naïve Bayes classification techniques. *Procedia Computer Science*, *116*, 107–112. doi:10.1016/j.procs.2017.10.017

Reyes-Menendez, A., Saura, J. R., & Thomas, S. B. (2020). Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *International Journal of Information Management*, *54*, 1–11. doi:10.1016/j.ijinfomgt.2020.102129

Sahu, S., Rout, S. K., & Mohanty, D. (2015, December 21–23). Twitter sentiment analysis—A more enhanced way of classification and scoring. In *Proceedings of the 2015 IEEE International Symposium on Nanoelectronic and Information Systems*. IEEE. doi:10.1109/iNIS.2015.40

Saura, J. R., Reyes-Menendez, A., & Palos-Sanchez, P. (2019). Are black Friday deals worth it? Mining Twitter users' sentiment and behavior response. *Journal of Open Innovation*, *5*(3), 58. doi:10.3390/joitmc5030058

Senadheera, V., Warren, M., & Leitch, S. (2011, July 7–11). A study into how Australian banks use social media. In *Proceedings of the 15th Pacific Asia Conference on Information Systems (PACIS2011)*. University of Queensland.

Shabbir, M. S. (2020). Attributes ensuring positive consumer evaluation in brand extension of Pakistan. *International Journal of Asian Business and Information Management*, *11*(4), 71–84. doi:10.4018/IJABIM.2020100106

Shabbir, M. S., & Zeb, A. (2020). Nexus and perception of customers toward conventional banking systems: Does the Islamic banking system exist as a competitor? *International Journal of Asian Business and Information Management*, *11*(4), 54–70. doi:10.4018/IJABIM.2020100105

Shaheen, Z. (2022). WeChat's effect on online purchase intention of fast moving consumer goods. *International Journal of Asian Business and Information Management*, *13*(1), 1–25. doi:10.4018/IJABIM.302249

Shakeel, M., Barsaiyan, S., & Sijoria, C. (2020). Twitter as a customer service management platform: A study on Indian banks. *Journal of Content*. *Community & Communication*, *11*(6), 84–104. doi:10.31620/JCCC.06.20/07

Sheikh, L. M., Tanveer, B., & Hamdani, M. A. (2004). Interesting measures for mining association rules. In *Proceedings of the 8th International Multitopic Conference*. IEEE. doi:10.1109/INMIC.2004.1492964

Singh, D. (2013). Service quality and customer satisfaction: A comparative study of an Indian public vs private bank. *Malaysian Management Journal, 17*, 59–75. https://e-journal.uum.edu.my/index.php/mmj/article/view/8994

Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computers and Applications*, *17*(8), 43–48. doi:10.5120/2237-2860

Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: Applications and issues. *Journal of Basic and Applied Sciences*, *13*, 459–465. doi:10.6000/1927-5129.2017.13.76

Sridevi, P., Niduthavolu, S., & Vedanthachari, L. N. (2020). Analysis of content strategies of selected brand tweets and its influence on information diffusion. *Journal of Advances in Management Research*, *18*(2), 227–249. doi:10.1108/JAMR-06-2020-0107

StatCounter. (2021). *Social media stats in Thailand*. Retrieved from https://gs.statcounter.com/social-media-stats/all/thailand

Surucu-Balci, E., Balci, G., & Yuen, K. F. (2020). Social media engagement of stakeholders: A decision tree approach in container shipping. *Computers in Industry*, *115*, 103152. doi:10.1016/j.compind.2019.103152

Tafesse, W., & Wien, A. (2018). Using message strategy to drive consumer behavioral engagement on social media. *Journal of Consumer Marketing*, *35*(3), 241–253. doi:10.1108/JCM-08-2016-1905

Tapsai, C., Meesad, P., & Unger, H. (2019). An overview on the development of Thai natural language processing. *Information Technology Journal*, *15*(2), 45–52.

Vidya, N. A., Fanany, M. I., & Budi, I. (2015). Twitter sentiment to analyze net brand reputation of mobile phone providers. *Procedia Computer Science*, *72*, 519–526. doi:10.1016/j.procs.2015.12.159

Voican, O. (2020). Using data mining methods to solve classification problems in financial-banking institutions. *Economic Computation and Economic Cybernetics Studies and Research*, *54*(1), 159–176. doi:10.24818/18423264/54.1.20.11

Wang, J., Cong, G., Zhao, W. X., & Li, X. (2015, January 25–30). Mining user intents in Twitter: A semi-supervised approach to inferring intent categories for tweets. In *IAAAI'15: Proceedings of the AAAI Conference on Artificial Intelligence*. Association for Computing Machinery. doi:10.1609/aaai.v29i1.9196

Wang, P., & McCarthy, B. (2021). What do people "like" on Facebook? Content marketing strategies used by retail bank brands in Australia and Singapore. *Australasian Marketing Journal*, *29*(2), 155–176. doi:10.1016/j.ausmj.2020.04.008

Wang, W., Cheng, X., Liu, Z., Lin, Y., Shen, Y., Hu, B., Zhang, Z., Zeng, X., Zhou, J., Gu, J., & Luo, M. (2022, May 9–12). Intent mining: A social and semantic enhanced topic model for operation-friendly digital marketing. In *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE. doi:10.1109/ICDE53745.2022.00308

We Are Social Inc. & Hootsuite. (2020). *Digital 2020 Thailand*. Retrieved from https://wearesocial.com/digital-2020

Weerawatnodom, N., Watanapa, N., & Watanapa, B. (2017, November 22–24). *Features of marketer-generated content tweets for electronic word of mouth in banking context* [Paper presentation]. The 9th International Conference on Advances in Information Technology (IAIT-2017), University of Technology Thonburi, Thailand. doi:10.18502/kss.v3i1.1398

Zhang, Y., Dong, C., & Cheng, Y. (2022). How do nonprofit organizations (NPOs) effectively engage with the public on social media? Examining the effects of interactivity and emotion on Twitter. *Internet Research*. 10.1108/INTR-05-2021-0290

Zulfikar, W. B., Irfan, M., Alam, C. N., & Indra, M. (2017). The comparation of text mining with Naive Bayes classifier, nearest neighbor, and decision tree to detect Indonesian swear words on Twitter. In *Proceedings of the 2017 5th International Conference on Cyber and IT Service Management (CITSM)*. IEEE. doi:10.1109/CITSM.2017.8089231

Mathupayas Thongmak is an associate professor of Information Systems in the MIS department of Thammasat Business School, Thammasat University in Bangkok, Thailand. She holds a Ph.D. in Computer Engineering from the Faculty of Engineering, Chulalongkorn University. Her research interests include adoption and diffusion of IS/IT, education/ learning relating to IS/IT, social media & marketing, software quality, digital piracy, and green ICT. Her research has appeared in various journals, including Journal of Computer Information Systems, International Journal of Electronic Commerce Studies, Online Journal of Applied Knowledge Management, Electronic Green Journal, Knowledge Management & E-Learning, Journal of E-Learning and Higher Education, International Journal of Information Technologies and Systems Approach, International Journal of Software Engineering and Knowledge Engineering, and Education for Information. She has reviewed multiple journals, including Sage Open, Engineering Journal, Children and Youth Services Review, International Journal of Electronic Commerce Studies, Knowledge Management & E-Learning, Recent Patents on Computer Science, and Telematics and Informatics.