Social Recommendations for Facebook Brand Pages

Yu-Ping Chiu

Chinese Culture University, Department of Advertising, Taipei, Taiwan, qyp2@ulive.pccu.edu.tw

Abstract

The objective of this research is to bridge the gap by proposing a content-based framework that is specifically designed to operate in social media environment. This study proposed a recommendation framework that integrated the features of brand pages and information on user behavior on brand pages. Data were obtained from 2,076 official brand pages in Taiwan, and a total of 500,000 interaction data were obtained and processed. Decision trees were used to classify brand pages and detect features that facilitate distinguishing among brand pages. Our method involved a simple training procedure with no restriction to a particular classifier learning algorithm and yields improved results on data sets extracted from a considerable number of Facebook brand pages. The results represented the diversity of social media user criteria in evaluating whether an activity is interesting to users. Moreover, these prevalent features indicate that brand pages distinguish themselves from others through their willingness to engage and interact with users. These findings not only enabled using brand page recommendation models for facilitating the selection of the most suitable brand page, but also useful for researchers seeking to develop a recommendation system for social media.

Keywords: Facebook, Brand page, Recommendation, User behavior, Decision tree

1   Introduction

Social media has emerged as an essential medium for maintaining social connections and sharing information among users [8], [31]. Social media is used not only for entertainment but also as platforms where consumers can discuss brand-related experiences and preferences [30], [43], which has become a part of our everyday lives. Hence, social media has aroused the interest of companies who see the potential to manage customer relationships [1], [15], [23]. Companies can create and moderate brand pages on Facebook. These pages provide users, including individuals, businesses, and organizations, a space for posting information with a specific purpose, audience, and context [28], [30], [44]. Therefore, brand pages are presently the key platform where consumers interact with brands.

Because the threshold for establishing a brand page is quite low, many homogeneous and heterogeneous brand pages continually appear on Facebook. Facebook users presently view a considerable number of brand pages at each visit, hindering the efficient and appropriate selection of brand pages that interest them the most [4], [27]. That is, the overwhelming number of brand pages on Facebook cause difficulties for users to select the right page to join, which is also the research gap. Moreover, the innovation of brand pages has constantly increased data availability, causing users to experience considerable information overload. The type of information can potentially help users solve problems. However, an excessive amount of information on different networks may become an obstacle rather than a benefit [29]. Hence, managing information and network overload on brand pages is crucial and serves as the objective of this study.

In particular, as new forms of technology have emerged, scholarship should revisit existing theoretical constructs to make sense of the evolving pathways of computer-mediated information. According to previous studies, one approach for managing information and network overload is recommending social groups in which users may be interested, enabling them to more effectively focus their attention [34], [48]. Recommendation is a filtering technique that involves suggesting useful information (images, news, music, travel, and web pages) to users on the basis of collected or inferred information [25]. Recommendation systems are important for enabling the personalization of activities in social media where one object is to recommend to people whom they may want to interact [7]. In fact, many social media use social networking analysis for recommending friends to users or as potential partners in life [9], [48]. However, this approach is not effective in managing brand pages. Therefore, this paper focus on recommending brand pages to users to address the following fundamental question: Which pages would be more relevant to users' interests?

In fact, one of the specific features of brand pages is that they enable users to interact and participate with companies, allowing them to share their opinions [37]. That is because users may belong to various brand communities according to their topics of interest, resulting in highly overlapping communities from a structural viewpoint. For example, Facebook users generally interact by selecting and viewing brand pages that are matched on the basis of brand page features and user behavior information. Likewise, users having similar tastes would tend to annotate similar resources that share some topics with one another. When appropriate matches are found, they express their interest by clicking on *like* and leaving a comment [19], [20]. That is, interaction frequency can be considered a success indicator for measuring the popularity of brand pages. Therefore, the objective of this research is to bridge the gap by proposing a content-based framework that is specifically designed to operate in social media environment.

This study proposed a framework for social recommendation and developed a representation for classifying interactions on Facebook brand pages that combines the content from page profiles and features of user behavior. We used the decision tree learning method for determining which feature can be used to examine interaction on brand pages and then classified the most applicable interaction category to recommend to users. Our method involves a simple training procedure with no restriction to classifier learning algorithm and produces the improved results on data sets extracted from the considerable number of Facebook brand pages.

The main contribution of this paper is a new social recommend mechanism. This recommend mechanism suggests items across the different applications based on two of the main characteristics of the brand page- profiles and user behavior, and makes the following contributions toward recommendation systems. First, this study reinforces the importance of revisiting foundational models on social media to effectively incorporate social behavior into recommendation processes in ways that may prove highly beneficial in accentuating users' interests and characterizing communities. Second, this study provides detailed experimental evaluations with real datasets, which are taken from Facebook. The results showed that the decision tree can be applied not only to classify the specific brand pages, but also used to distinguish one brand page from another. Third, we also elucidate the relatively new phenomenon of brand pages on social media and useful implications regarding how to provide businesses with additional insights into consumer attitudes. The ability to keep step with the dynamic features of the brand page is needed to study them in a sincere way.

2   Literature Review

Relevant literature including Facebook brand pages and recommendation mechanism are presented as follow.

2.1   Facebook Brand Pages

The rise and continued growth of Facebook have attracted the interest of companies who see the potential to send their marketing content to the customers [23]. Facebook then provides the brand pages which is a platform for sports teams, films, organizations, and businesses to attract audiences and maintain a long-term relationship with their users or consumers [11]. At this time, companies can establish Facebook brand pages with a specific purpose, audience and topic to market, promote and inform about products and services. These brand pages allow enterprises and celebrities to interact with users and thereby enhance brand visibility to and the loyalty of these people [18], [39]. In recently, more than 1.5 million organizations have Facebook brand pages, and 20 million people *like* Facebook brand pages every day. These brand pages were found to be a successful tool for increasing sales. In addition, they have the potential of improving the relationship between the consumers and the brand, and may influence fans' perceptions and actions [2].

Through news feeds on Facebook, page administrators can send links, advertisements, videos, and information considering a specific purpose, audience, and topic to the market [16]. Once users become fans of brand pages by pressing the *like* button, the new content of the brand page is automatically posted to their news feed. Fans can then post content or comment on brand posts (depending on the communication policy set by the company).

Indicate interest in an existing post by pressing the like button, contact the companies, and interact with other fans [1], [11]. This means that companies set up and moderate brand page to continuously monitoring consumers' activities [46]. That is, brand pages facilitate interactions through the exchange of opinions and information about the brand or products. For example, Coca-Cola has over 36 million fans and actively promotes its activities on its Facebook brand page. In brief, fans tend to be loyal and committed to companies and are receptive to receiving information about the brands [12], [18]. Moreover, Dholakia and Durham [13] and Choi and Scott [10] observed that people who follow brand pages are more willing to visit the stores and spread positive word of mouth on the Internet. As the outcome of this change in the field of marketing, social marketing has emerged.

With the rapid growth of social marketing, researchers have analyzed the features and structures of social media. Distinct features of social media, such as taste, linkage, and subgroup differences, have been studied by Lewis, Kaufman, Gonzalez, Wimmer and Christakis [24] and Mislove, Marcon, Gummadi, Druschel and Bhattacharjee [35], revealing the reality of user behavior and network structures. Ahn, Han, Kwak, Moon and Jeong [3] and Shi, Rui and Whinston [43] analyzed the topology of online social structures and described the popularity of user-generated content. Results showed that it is not a replacement for the traditional marketing techniques but rather as an additional marketing channel that can be integrated with all channels. However, among those studies, little research has systematically examined the problem of Facebook users having to review too many brand pages prior to selecting their preferred page.

Although brand pages contribute toward connecting with companies and users, the presence of too many brand pages causes considerable information and network overload [49]. Moreover, diverse information on brand pages may hinder the clear identification of pages suitable for tracking and interacting by users [38]. Therefore, brand page administrators should develop approaches that improve their exposure and enable existing and potential users to find and track specific brand pages [12]. Considering this issue, the main approach to counter information and network overload is recommending social updates of interest to users, enabling them to more effectively focus their attention [25], [48]. In fact, Facebook users interact with the specific brand pages that are matched on the features and contents. Once the appropriate matches are found, users will express their interest by clicking on like button and leaving a comment [19], [20]. Therefore, it is necessary to find more suitable brand page for users. In this study, a framework for a recommendation system was developed for recommending suitable and attractive brand pages to users.

2.2   Recommendation Mechanism

Recommendation mechanisms attempt to discover user preferences over items by modeling the association between users and items. The mechanism attempts to clarify user preferences by modeling the association between users and items [14], [42], and then prevent information overload by presenting appropriate content to targeted users on the basis of collected or inferred information [29], [45]. Hence, they recommend items that match the preferences of users to assist active users in selecting products from potential options [7]. According to the previous study [7], by applying recommendation techniques it is possible to increase the likelihood of the successful purchase of products by the active user.

Recommendation mechanisms can be categorized as content-based or collaborative filtering (CF) mechanisms. Content-based mechanisms analyze the features of the content in a set and matches them to the preferences or interests of users on the basis of a model developed by analyzing the previous actions of users [48]. That is, the utility $u(c, s)$ of item s for user c is estimated based on the utilities $u(c, s_i)$ assigned by user c to items $s_i \in S$ where s_i is similar to item s. Content-based mechanisms depend on three main processes: content analyzer, profile learner and filtering components. According to previous studies, the content analyzer is used to extract information (such as keywords, concepts) that represent items, and to extract users' reactions towards the items. The profile learner is used to learn users' preferences to construct and update user profiles (from their past reactions towards items). Filtering components match the user profile with items characteristics to accomplish the recommendation [42]. In brief, this mechanism may be preferred where profile information is presented in the form of item descriptions (e.g., for explaining why recommendations are made) [26].

By contrast, collaborative filtering mechanisms recommend items on the basis of aggregated user preferences and the general taste of similar user profiles, irrespective of the availability of item descriptions [17], [42]. The utility $u(c, s)$ of item s for user c is estimated based on the utilities $u(c_j, s)$ assigned to item s by those users $u(c_j, s)$ who are similar to user c. Collaborative filtering mechanisms have two types: user-based and item-based [40]. In user-based collaborative filtering, a user C who is interested in item x will be interested in item y, if there are other users. These users are similar to user C and they are interested in the items x and y. In item-based system, if user C likes item x, and the item x is similar to the item y according to the opinion to other users. Then item y should be recommended to the user C [40]. Thus, in the CF mechanism, user behaviors are used for generating user preferences.

By applying recommendation mechanisms, the likelihood of the successful purchase of products by active users can be increased because services or products are personalized and presented to these users by using information obtained from the purchasing behavior of like-minded users [9], [21]. This mechanism is similar to that applied by Amazon.com. However, traditional content-based and CF mechanisms are not sufficient for Facebook brand page recommendations because they assume that users are independent. That is, the recommendation mechanism needs to consider the interaction between user and user. The interaction relationship was not previously considered. Therefore, as social media expands exponentially, it will become a challenge to distribute the right information and pages to the right person.

This study proposed a framework for social recommendation and integrated user behavior information and profile features into content-based recommendations by using simple and powerful learning and ranking approaches to avoid expensive computation [17], [36], which was vital for the large, real commercial data set used in our experiments. At this time, recommend technologies can play a key role to process the new types of public data and metadata on social media, which can further facilitate creating a list of suitable brand pages, and then ensure users to receive the information they prefer.

3 Materials and Methods

Based on the literature, this study designed Facebook application to collect data and develop the recommendation system. Data collection and construction of decision trees are shown as follow.

3.1 Method Design

To develop an effective recommendation system on social media, this study integrated both factors using probabilistic ranking and machine learning. This method forms the relationship between the features of candidate brand pages as a classifier. The features depended on the selected profile and user behavior information on brand pages; each interaction was labeled with a target value, and regard as the candidates in that pair. In general, any classifier learning method can automatically extract a set of classification rules is applicable in this study. Therefore, to maximize cost-weighted performance on a training set of interactions, this study used a learned model to convert into rule sets containing features of both active and potential users, enabling candidate list generation in a comprehensive data representation [33].

To empirically examine the effectiveness of the aforementioned method, we used Facebook brand pages as the experimental platform. This study used the Facebook API (Application programming interface) for data collection from 2,076 brand pages on December 2017 and then constructed and selected characteristics based on the features and user interaction records of these pages. The corresponding recommended candidates were generated according to the rules derived from the decision tree [47], providing a clear understanding of user preferences and enabling the recommendation of brand pages that may be interesting to users. (see Figure 1)
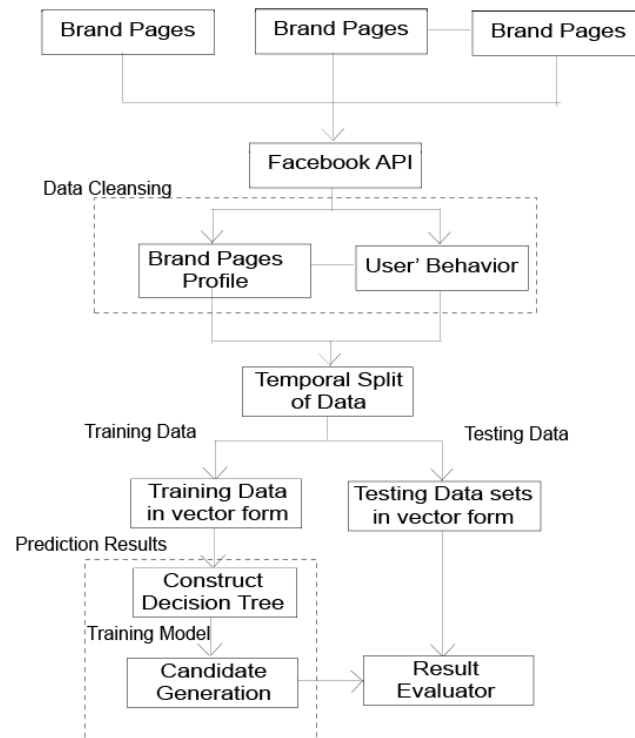
Figure 1: Depicts the process flow of brand page prediction

3.2 Facebook API

Facebook launched its API to the public to attract application developers. The raw data were crawled using the Facebook API and cleaned on the basis of criteria [32]. The first step of the recommendation system involves analyzing and identifying features that capture user trends in interests, social connections, and basic information. We accessed 2,076 brand pages and collected the profile features and user behavior information from each page. And this study does not use any private information.

Brand page profile features. Basic attributes usually described the profile of a brand page, and these attributes are highly domain-specific. This study extracted 19 features, namely category, subcategory, creation time, number of statuses visible to the public, location of members, number of wall posts, number of fans, increase in the number of fans, increase rate of the number of fans, number of discussions on the page, the ratio of discussion, and fans' average age, sex ratio, education level, race, relationship status, political view, affiliations, and lifestyle choices, for characterizing the profile features on Facebook. These data can be easily accessed because they are publicly available, and the service terms of Facebook allow program designers to obtain anonymous usage the data for research purposes [11]. In this study, the above data can be obtained through the APIs offered by Facebook.

User behavior features. This set covered behavior patterns in aggregated interaction summaries over time. Specifically, we extracted 11 features include the cumulative numbers of activities, posts, links, photos, videos, notes, and the total numbers of likes, comments, shares, tags, and check-ins from all members on each brand page were collected. Moreover, because some brand pages have already built for a long time and cumulative many user interaction records, this study also examined user behavior in one month. To avoid members still interacting with the brand pages on December (we used the Facebook API for data collection), this study collected the data three-month ago. That is, the numbers of activities, posts, links, photos, videos, notes, likes, comments, shares, tags, and check-ins from members in September 2017 (only this month) were separately collected (also regard as 11 different features). In addition, we also examined whether these behaviors were those of the administrators or members of brand pages. Hence, we observed 44 features related to the interaction on brand pages.

It can be reasonably assumed that the popularity and interaction frequency of brand pages are associated with their value and attractiveness. Commonly available interaction frequency information included the total number of interactions on brand pages (numbers of likes, comments, and shares) and the total number of fans who interacted on brand pages. To compare the results of these two features, we constructed two decision trees.

The crawled data were cleaned; however, these data cannot be directly analyzed. Hence, processes such as splitting and vectorization were used. We then temporally split the data set to obtain training and testing data sets because the past data (training data, 70%) can be used for training the model, which can subsequently predict the

future data (testing data, 30%). Using this property, we constructed decision trees to enhance brand page recommendation for Facebook users.

3.3 Construction of Decision Trees

The main goal of these splitting algorithms is to find the best split of data with maximum homogeneity on each side. Each recursive iteration purifies data until the algorithm reaches to terminal classes.

Among many supervised techniques used for machine learning, the decision tree classifier is an appropriate choice for the proposed method [5], [21], [47]. According to binary recursive partitioning, a decision tree comprises the parent node t_p and child nodes t_l and t_r. This study adopts an impurity function $i(t)$ to define the maximum homogeneity of child nodes. Therefore, the maximum homogeneity of t_l and t_r is equal to the maximum change in the impurity function $\Delta i(t)$ [Eq. (1)], which indicates that a splitting rule is applied to all variable values to find the optimal split criterion $x_i \leq x_j^R$, and the maximum $\Delta i(t)$ is found. The P_l and P_r indicate the probabilities of the left and right nodes.

$$\Delta\, i(t) = i(t_p) - P_l i(t_l) - P_r i(t_r) \tag{1}$$

Thus, maximum impurity is solved in each recursion step and expressed in Eq. (2). In this equation, x_j is variable j, x_j^R is the most favorable possible variable x_j to split, and M is the number of variables.

$$max_{x_j \leq x_j^R, j=1...M}[i(t) = i(t_p) - P_l\, i(t_l) - P_r i(t_r)] \tag{2}$$

3.4 Candidate Generation

In fact, it is important to generate efficiently candidate lists since the aim of a recommendation system is to present the brand pages with an actual list of relevant items. Existing study often estimates the constructed model on the historical data, but pays little attention to candidate list generation. Specifically, this study used decision trees for constructing a social recommendation system requires a training set for training a recommender, which generates candidate pairs and then used to test on a test data set. To evaluate the performance of the recommendation system, 70% of the data were randomly selected for training and the remaining 30% were used for testing (in total, 2,076 brand pages).

In particular, this study first converted a decision tree into decision rules to produce candidate lists for recommendations. After that, we evaluate all situations in each path from the root to each leaf in the decision tree and then creating a rule for each path. Rules $r \in R$ have the following format:

$$if\ (C_1^a, C_2^a, ... C_n^a) \wedge (C_1^r, C_2^r, ... C_m^r) \wedge (C_1^{ar}, C_2^{ar}, ... C_k^{ar})\ then\ t \tag{3}$$

Where C_i^a is the condition on the active brand page B_a, C_i^r is the condition on the recommended brand page B_r, C_i^{ar} are conditions on the pair (B_a and B_r), and t is the class.

3.5 Experimental Tool

After the data set was collected from 2,076 brand pages, the data were preprocessed in Oracle 10, and decision trees were learned using See5[2] and converted to rules in SQL by using Perl. Furthermore, Waikato Environment for Knowledge Analysis (WEKA) was used for building and classifying the model. Using the feature selection function of WEKA, we observed which features are valuable for measuring the interaction frequency on brand pages. Tree learning requires approximately 10 min per tree. Candidate generation requires approximately 1 hr, whereas ranking requires less than 30 min.

4 Results

The experimental results for recommendations and the decision tree for models are provided as follow.

4.1 Descriptive Statistics

In this study, we evaluated the proposed approach in a realistic setting by using a data set of Facebook brand pages. In total, 2,076 brand pages with 57 common subtype categories were crawled for testing. We cleaned the dataset in order to conduct more meaningful experiments. More than 500,000 interactive data were used in the model. Table 1 shows the number of brand pages, the total number of interactions, and the total number of fans who interacted in the top 10 subtype categories. The results revealed that Facebook users are highly willing to engage and interact in brand pages in *books* and *funny* categories. We then trained two decision trees in the training period.

Table 1: Number of brand pages and interaction of 10 main subtype categories

Ranking	Subtype category	Amount of pages	the total number of interactions	the total number of fans who interacted
1	Books	69	86,085	16,990
2	Funny	143	62,883	20,298
3	Sexuality/Moods	59	50,800	8,290
4	Consume/Lifestyle Information	129	49,924	11,712
5	Other Information	251	25,521	12,242
6	News	28	17,067	7,220
7	Artist/Band	206	15,448	5,450
8	Famous person	116	11,084	4,384
9	Religion and spirituality	41	10,810	5,924
10	Constellation/Divine	38	9,216	3,706

4.2 Experimental Results for Recommendations

In Model 1, we considered *the total number of interactions* as the indicator of interaction frequency on brand pages. This model has twelve rules derived from the decision tree. The results revealed that the suitable category of brand pages is affected by many factors, including *the number of likes received in the current month, the number of links shared in the current month, the number of comments received in the current month, the ratio of discussion* and *the average age*. According to the classification results of decision tree, we converted the rules to the *if…then*… format (see Table 2). For example, if the number of likes received and number of links shared in the current month were >84,684 and >15,195, respectively, then the brand page belonged to Category A.

Table 2: The rules of decision tree in model 1

Rule 1	If the number of likes received in the current month > 84,684 and the number of links shared in the current month > 15,195, then it belonged to Category A
Rule 2	If the number of likes received in the current month > 84,684, 15,195 ≥ the number of links shared in the current month > 2,929, then it belonged to Category A
Rule 3	If the number of likes received in the current month > 84,684 and the number of links shared in the current month ≤ 2,929, then it belonged to Category B
Rule 4	If the number of likes received in the current month ≤ 84,684 and the number of comments received in the current month >1,144, then it belonged to Category B
Rule 5	If the number of likes received in the current month ≤ 84,684, 1144 ≥ the number of comments received in the current month > 326, and the ratio of discussion > 0.075, then it belonged to Category C
Rule 6	If the number of likes received in the current month ≤ 84,684, 1,144 ≥ the number of comments received in the current month > 326, and the ratio of discussion ≤ 0.075, and the average age > 27.75, then it belonged to Category C
Rule 7	If the number of likes received in the current month ≤ 84,684, 1144 ≥ the number of comments received in the current month > 326, and the ratio of discussion ≤ 0.075, and the average age ≤ 27.75, then it belonged to Category D
Rule 8	If the number of likes received in the current month ≤ 84,684, the number of comments received in the current month ≤ 326, and the number of links shared in the current month > 863, then it belonged to Category D
Rule 9	If the number of likes received in the current month ≤ 84,684, 326 ≥ the number of comments received in the current month >187, and the number of links shared in the current month ≤ 863, then it belonged to Category D
Rule 10	If the number of likes received in the current month ≤ 84,684, the number of comments received in the current month ≤ 187, the number of links shared in the current month ≤ 863, the average age >22.25, and the number of likes received in the current month >5,026, then it belonged to Category D
Rule 11	If the number of likes received in the current month ≤ 84,684, the number of comments received in the current month <=326, the number of links shared in the current month ≤ 863, the number of comments received in the current month ≤ 187, the average age >22.25, and the number of likes received in the current month ≤ 5,026, then it belonged to Category D
Rule 12	If the number of likes received in the current month ≤ 84,684, the number of comments received in the current month ≤ 326, the number of links shared in the current month ≤ 863, the number of comments received in the current month ≤ 187, and the average age ≤ 22.25, then it belonged to Category E

It is informative to observe which features are used at the top-level splits made by the decision tree. Figure 2 shows the top 6 splits created by a representative tree. Although the features and splits in the tree varied depending on the sample, the top 6 splits were quite stable, with only minor variations among samples.

At the top level, brand pages that received at least 84,684 *likes* in the current month were highly likely to belong to Class 1. The next split was based on the number of links shared in the current month; if it was >15,195, then the brand page belonged to Category A. If the number of links shared in the current month was <15,195, then whether this number was >2,929 was determined. If the number of links shared in the current month was >2,929 or <2,929, then the brand page belonged to Categories A and B, respectively. That is, brand posts exhibited more online engagement in Category A, and fans were likely to engage with brand pages by clicking like and share button. Moreover, these brand pages were likely to post the form of photo and video to encourage the interaction of fans. On the other hand, fans were likely to make some discussion and leave the comment in Category B. That is, the brand posts enabled users to experience resonance and encouraged them to express their opinion in Category B.

The second class of features, essential for the growth of brand pages (although for reasons that are not intuitively apparent), is the number of comments received in the current month. If a brand page received >326 comments in the current month, it underwent further categorization. If the brand page received >1,144 comments, then it belonged to Category B. However, if <1,144 comments were received, it was categorized further according to the ratio of discussion and average age. The results showed that if the ratio of discussion was >0.075, then the brand page belonged to Category C. Moreover, if the average age was >27.75 or <27.75 years, then the brand page belonged to Categories C and D, respectively (see Figure 2). Therefore, brand fans were older and likely to engage with brand pages in Category C. However, these brand pages were not frequently post the information and even interact with their fans. On the other hand, fans were young and likely to engage with brand pages in Category D. However, the brand posts still cannot encourage fans to express their opinion in Category D. Compare with Category C, not only brand fans were older, but they were also unlikely to engage with brand pages in Category E.

According to the decision tree of Model 1, the brand pages in Category A have the highest interaction frequency among all pages, indicating that the pages are suitable for and attractive to users. Moreover, for evaluating the performance of the decision tree of Model 1, we used the remaining 30% data for testing. The accuracy rate was measured according to the ratio of appropriately clarified brand pages to the total testing data. The results showed an average accuracy of 42.9%. (see Figure 2)
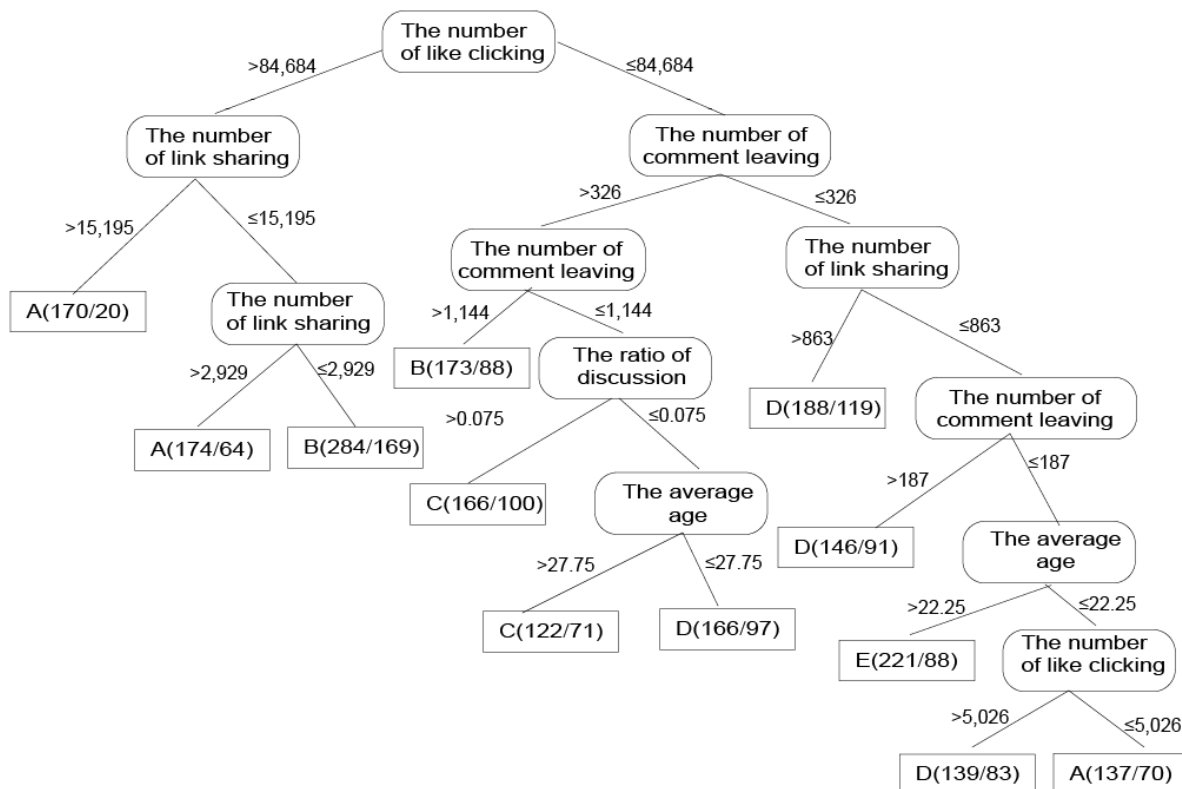

Figure 2: A decision tree for model 1

In Model 2, we adopted *the total number of fans who interacted* on brand pages in the current month as the indicator of interaction frequency. The decision tree produced nine decision rules. For example, if the number of fan

discussions on the page were >12,289, the number of comments received in the current month were >818, and the average age were >23.5 respectively, then the brand page belonged to Category A. (see Table 3)

Table 3: The rules of decision tree in experiment

Rule 1	If the number of fans discussions on the page >12,289, the number of comments received in the current month >818, and the average age >23.5, then it belonged to Category A
Rule 2	If the number of fans discussions on the page >12,289, the number of comments received in the current month >818, and the average age ≤ 23.5, the number of fans discussions on the page >29,583, then it belonged to Category A
Rule 3	If the number of fans discussions on the page >12,289, the number of comments received in the current month >818, the average age ≤ 23.5, and the number of fans discussions on the page ≤ 29,583, then it belonged to Category B
Rule 4	If the number of fans discussions on the page >12,289, and the number of comments received in the current month ≤ 818, then it belonged to Category B
Rule 5	If the number of fans discussions on the page ≤ 12,289, and 1,229 ≥ the number of comments received in the current month >126, then it belonged to Category B
Rule 6	If the number of fans discussions on the page ≤ 12,289, 1,229 ≥ the number of comments received in the current month >126, and the average age >21, then it belonged to Category C
Rule 7	If the number of fans discussions on the page ≤ 12,289, 1,229 ≥ the number of comments received in the current month >126, the average age ≤ 21, and the number of fans discussions on the page >18,513, then it belonged to Category C
Rule 8	If the number of fans discussions on the page ≤ 12,289, 1,229 ≥ the number of comments received in the current month >126, the average age ≤ 21, and the number of fans discussions on the page ≤ 18,513, then it belonged to Category D
Rule 9	If the number of fans discussions on the page ≤ 12,289, the number of comments received in the current month ≤ 126, then it belonged to category E

Figure 3 shows the top 5 splits created by a representative tree. At the top level, we observed that brand pages having at least 12,289 fan discussions on the page in the current month were likely to be in Class 1. The next split was based on the number of comments received in the current month on the brand page. If the brand page received >818 comments, it underwent further categorization according to the average age and number of fan discussions on the page. Moreover, if the average age was >23.5 years, then the brand page belonged to Category A. If the average age was <23.5 years, but the number of fan discussions on the page was >29,583, then the brand page also belonged to Category A. That is, fans were older and likely to discuss the specific issues in these brand pages.

By contrast, if the number of fan discussions on the page was <29,583, then the brand page belonged to Category B. Moreover, if a brand page received >126 comments in the current month, it underwent further classification based on whether the number was >1,229. If the brand page received >1,229 comments, then it belonged to Category B. Therefore, although fans were willing to leave the comment in these pages, they were not likely to interact with other fans.

However, if <1,229 comments were received, then it underwent further categorization according to the average age and number of *likes* received. Moreover, if the average age was >21 years, then the brand page belonged to Category C. If the average age was <21 years, but the number of *likes* received was >18,513, then the brand page also belonged to Category C. The second class of features, essential for the growth of brand pages (although for reasons that are not intuitively apparent), is the number of comments received in the current month. If a brand page received >326 comments in the current month, it underwent further categorization. If the brand page received >1,144 comments, then it belonged to Category B. However, if <1,144 comments were received, it was categorized further according to the ratio of discussion and average age. The results showed that if the ratio of discussion was >0.075, then the brand page belonged to Category C. That is, brand fans were older and likely to engage with brand pages in Category C. However, the brand posts still cannot encourage fans to express their opinion in these pages.

By contrast, if <18,513 *likes* were received, then the page belonged to Category D. That is, brand posts cannot facilitate online engagement in these pages. Figure 3 summarize the decision tree results. To evaluate the performance of the decision tree for Model 2, we used the remaining 30% of the data for testing. According to the feature and model selection results, a success rate of 49.9% was obtained for the testing data set.
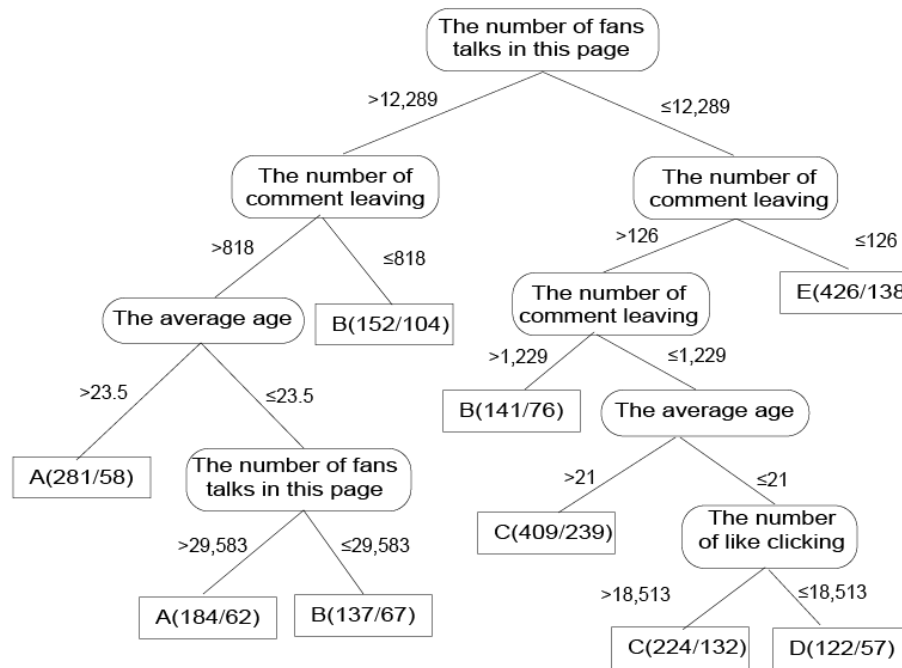
Figure 3: A decision tree for model 2

In conclusion, Model 1 revealed that the number of *likes* received in the current month was the most critical feature in prediction. The number of link sharing instances in the current month, number of comments received in the current month, ratio of discussion, and average age are also useful because they produce the optimal machine learning algorithm performance. Model 2 revealed that the number of fan discussions on the page, number of comments received in the current month, average age, and the number of l*ikes* received in the current month contributed in examining the interaction frequency on brand pages.

5 Conclusion

This section describes the discussion, theoretical and practical implications. Then, the limitation and directions for future studies are discussed.

5.1 Discussion

Social media has changed the creation, distribution, and consumption of brand-related content, transferring the power to shape brand images from advertisers to online consumer [50]. The principal objective of this study was to provide accurate methods that will be readily applicable to Facebook. Therefore, this study developed a recommendation system, which is based on two of the core elements of Facebook brand pages-profile features and user behavior features, to improve the quality of service for users.

The experimental results revealed several valuable findings. First, social media platforms merely recommended the same communities to all members based on the community similarities. This study extended this concept and evaluated the recommendation system through an extensive user study. We proposed a machine learning framework for using both profile information and user behavior information to recommend brand pages by integrating feature collection, decision tree construction, and candidate generation. The evaluation of brand page data confirmed the satisfactory performance of personalized recommendations based on social data, consistent with the finding of past research [41]. That is, these results signified that when considering the large-scale user data of brand pages, automatic analysis can be used for identifying users with certain interaction behavior. This factor can be scaled up far more easily than manual human assessment can. Second, similar with the previous study [6] that extracted several features on social media, this study extracted 63 features for characterizing Facebook brand pages, and then transformed the real data input features with boosted decision trees to increase the prediction accuracy of probabilistic linear classifiers. These prevalent features clearly indicate that brand pages distinguish themselves from others through their willingness to engage and interact with users [19]. In addition, each recommended feature is accompanied by an explanation that includes the features that led to its recommendation, as well as their relationships with the brand pages and users. Some features involve partially connected some meanings with other features.

80

Third, the study presented by Heimbach and Hinz [16] differs importantly from our work. The results revealed that both the total number of interactions and a total number of fans who interacted on brand pages were vital for classifying suitable brand pages. However, these two models had diverse crucial features and rules for classifying the brand pages, and the categorization results differed. For example, in Model 1, the number of *likes* received in the current month was the most crucial feature in the prediction, whereas the number of fan discussions on the page was the most critical feature in Model 2. These differences represent the diversity of social media user criteria in evaluating whether an activity is interesting to users. Such brand pages require more resources than do brand pages that only post specific content. Fourth, this study provides some interesting insights regarding the problems inherent to brand pages recommendation, which may ultimately result in the development of better solutions for this problem. That is, results were not only consistent with previous studies [21], [47] that using decision trees to give a convenient way of doing feature selection by means of feature importance, but also presented the tradeoff between the number of boosted decision trees and accuracy. That is, it is advantageous to keep the number of trees small to keep computation and memory contained. This study is relatively flexible and easily applicable to social services working with brand pages.

In sum, based on the extensive evaluation, this study suggests a particular machine learning method ad a set of features for learning user preferences that can provide the best performance on the personalized recommendation. Our method is simple to train, is not restricted to a particular classifier learning algorithm, and gives improved results on datasets extracted from Facebook brand pages. Our work can also successfully recommend brand pages to users even when they have not previously belonged to any brand pages. Therefore, this is useful for other researchers seeking to develop recommendation systems on social media.

5.2 Theoretical Implications

The results have several theoretical implications. First, this study adds to the emerging literature on recommendation system on social media and situates this phenomenon to existing models of information seeking. In particular, previous studies have specifically focused on recommended friends and groups [22]. Our method extended previous studies and involves a simple training procedure with no restriction to classifier learning algorithm and produce the improved results on data sets extracted from the considerable number of Facebook brand pages. This model is also useful for future researchers seeking to develop a recommendation system for social media. Second, the availability of complex data sets on social media communities and these communities' gradual evolution leads to a search for more refined theoretical model. That is, 19 profile features and 11 user behavior feature from brand pages used in this study are rarely included in previous studies. This study aimed to integrate user behavior information into content-based recommendations by using simple and powerful learning and ranking approaches to avoid expensive computation [17], [36], which was vital for the large, real commercial data set used in our experiments. Such a combination may be viewed as social recommendation variation of a traditional hybrid recommender. In other words, we effectively incorporate social behavior into recommendation processes in ways that may prove highly beneficial in accentuating users' interests and characterizing communities.

Third, decision tree has seldom been used and discussed on social media [28], this study reinforces the importance of social recommendation and provides detailed experimental evaluations with real datasets, which are taken from Facebook. It also suggested that decision tree can be applied not only to classify the specific brand pages, but also used to find the value of features. Fourth, the method proposed in this study has various applications, including information distribution systems based on the features of user profiles. A methodology for structuring information provided to users from their friends, groups, and networks must be developed. By applying recommendation mechanisms, the likelihood of the successful purchase of products by active users can be increased by using information obtained from the purchasing behavior of like-minded users [9], [21], so the mechanism we used was similar to that applied by Amazon.com. As social media exponentially expand, it will be challenging to distribute the appropriate information to concerned people. If the brand pages followed by users can be identified, then users can be provided with information in which they are interested.

5.3 Practical Implications

The results of our study have several practical implications. First, the increasing diversification of the modern media market splits media consumers into atomized niches. For users, it is difficult to recognize which brand pages are popular and interactive. This study developed recommendation system to find more suitable brand page to join by matching users' profiles with pages' identity. Evaluation of real data confirms that users generally like the proposed features, and machine learning can be effectively applied to provide a personalized recommendation of brand pages and help users deal with cognitive overload. In particular, Facebook platform can classify all brand pages to generate list of most favorable pages for each user. Moreover, once the most suitable brand pages are found, other nodes in the same sub-tree or neighboring tree share similarity with the most suitable brand pages.

Second, this study showed that the decision tree can be applied not only to classifying brand pages, but also used to find the value of features that distinguish one page from another. Therefore, brand page managers can develop their strategies using the key rules of the decision tree to improve operational performance. To strengthen user engagement levels, brand page administrators generally must craft engagement strategies that specifically

encourage users to click the *like* button or leave a comment. Thus, the proposed method can increase the likelihood of a brand page to be followed instead of only being published and waiting to be found. In particular, we suggest that brand page managers need to build their strategies to attract these lurk fans so that they will post likes and comments to improve service quality for Facebook users. These managers could affect the friends of these members and increase the exposure and discussion of their brand pages. For example, when the targeted audience is informational fans, managers should pay attention to the information within posts and provide abundant content that relates to the information.

5.4   Limitation and Directions for Future Studies

Certain limitations must be considered when interpreting the results of our study. First, the limitations of our study include a relatively small data set extracted from Facebook brand pages. To improve service quality for Facebook users, this study developed recommendation system to find the most suitable brand page to join by matching users' profiles with pages' identity. In this preliminary study, although 63 externally features were collected and used for building the decision trees, this study did not examine the influence of the dimensions of comments and other behavior to avoid the issue of privacy, thereby affecting the recommendation quality. As we have seen in the current recommendation mechanism on the market, this study can only have the 49.9% success rate. Future studies should consider the length, content, and quality of comments for collecting data without privacy issues in re-examining the influence of the conditions to obtain a more precise and complete understanding of social interaction in this area. Moreover, it is imperative to deeply explore the relative importance of different features of social network activities, which will contribute to further improvement of the performance of personalized activity recommendation. Second, other notable features include the textual content of activities and the targeted friends of friends in activities need to be discussed. That is, future studies may consider using feature selection for reducing the feature space size, and it is also interested in exploring more deeply the relative importance of different features of social activities. These directions may further improve the performance of personalized recommendation of activities.

Third, this study used the Facebook API for data collection from 2,076 brand pages by author, future studies can try to cooperate with Facebook to gain more specific data which we cannot download. Moreover, Open Graph enables users to connect with people and topic that interest them, would facilitate recommending activities. When new social data is retrieved, users will be provided with recommendations about whether the new data is interesting to them. In fact, considering the varying features and purposes of social media, future studies should focus on other major social media, such as Twitter and LinkedIn, to gain a broader understanding of the critical features for predicting popular brand pages. Four, we also elucidate the relatively new phenomenon of brand pages on social media and useful implications regarding how to provide businesses with additional insights into consumer attitudes. Many advertising techniques have already been implemented in providing target-based advertising to social media users. For instance, Amazon.com provides advertisements on the basis of user behaviors and search keywords, whereas Google AdSense provides them on the basis of the context around its banner. Future studies can replicate the concept of this study to examine the performance of recommendation mechanisms, and the ability to keep step with the dynamic features of the brand page is needed to study them in a sincere way.

References

[1]   P. Adamopoulos, A. Ghose and V. Todri. The impact of user personality traits on word of mouth: Text-mining social media platforms, Information Systems Research, vol. 29, no. 3, pp. 612-640, 2018.
[2]   M. T. Adjei, S. M. Noble and C. H. Noble, The influence of C2C communications in online brand communities on customer purchase behavior, Journal of the Academy of Marketing Science, vol. 38, no. 5, pp. 634-653, 2010.
[3]   Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, Analysis of topological characteristics of huge online social networking services, in Proceedings of the 16th International Conference on World Wide Web, ACM, Banff Alberta Canada, 2007, pp. 835-844.
[4]   T. Akcura, K. Altinkemer and H. Chen, Noninfluentials and information dissemination in the microblogging community, Information Technology and Management, vol. 19, no. 2, pp. 89-106, 2018.
[5]   T. T. Allen, Z. Sui and N. L. Parker, Timely decision analysis enabled by efficient social media modeling, Decision Analysis, vol. 14, no. 4, pp. 250-260, 2017.
[6]   E.-A. Baatarjav, S. Phithakkitnukoon and R. Dantu, Group recommendation system for facebook, in Proceedings OTM Confederated International Conferences On the Move to Meaningful Internet Systems, Berlin, Heidelberg: Springer, 2008, pp. 211-219.
[7]   X. Cai, M. Bain, A. Krzywicki, W. Wobcke, Y. S. Kim, P. Compton, and A. Mahidadia, Collaborative filtering for people to people recommendation in social networks, in Proceedings Australasian Joint Conference on Artificial Intelligence. Berlin, Heidelberg: Springer, 2010. pp. 476-485.
[8]   D. Castilho, P. O. V. de Melo and F. Benevenuto. The strength of the work ties, Information Sciences, vol. 375, pp. 155-170, 2017.
[9]   A. Cutler, and B. Kulis, Inferring human traits from Facebook statuses, in Proceedings International Conference on Social Informatics, Saint-Petersburg, Russia: Springer, 2018, pp. 167-195.

[10] J. H. Choi and J. E. Scott, Electronic word of mouth and knowledge sharing on social network sites: A social capital perspective, Journal of Theoretical and Applied Electronic Commerce Research, vol. 8, no. 1, pp. 69-82, 2013.

[11] I. P. Cvijikj and F. Michahelles, Online engagement factors on Facebook brand pages, Social Network Analysis and Mining,. vol. 3, no. 4, pp. 843-861, 2013.

[12] L. De Vries, S. Gensler and P. S. Leeflang, Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing, Journal of Interactive Marketing, vol. 26, no. 2, pp. 83-91, 2012.

[13] U. M. Dholakia and E. Durham, One café chain's Facebook experiment, Harvard Business Review, vol. 88, no. 3, pp. 26, 2010.

[14] I. Guy, People recommendation on social media, In: Brusilovsky P., He D. (eds) Social Information Access. Lecture Notes in Computer Science book series (LNCS, volume 10100), Cham: Springer, 2018, pp. 570-623.

[15] W. He, F.-K. Wang, Y. Chen, and S. Zha, An exploratory investigation of social media adoption by small businesses, Information Technology and Management, vol. 18, no. 2, pp. 149-160, 2017.

[16] I. Heimbach and O. Hinz, The impact of sharing mechanism design on content sharing in online social networks, Information Systems Research, vol. 29, no. 3, pp. 592-611, 2018.

[17] T. Hofmann, Collaborative filtering via gaussian probabilistic latent semantic analysis, in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, ACM, Toronto Canada, 2003, pp. 259-266.

[18] L.-C. Hsu, How to make loyal fans in virtual communities: Evidence from the facebook fan pages of smartphones, Journal of Theoretical and Applied Electronic Commerce Research, vol. 14, no. 3, pp. 168-189, 2019.

[19] Y.-L. Hwong, C. Oliver, M. Van Kranendonk, C. Sammut, and Y. Seroussi, What makes you tick? The psychology of social media engagement in space science communication, Computers in Human Behavior, vol. 68, pp. 480-492, 2017.

[20] Y. Jiao, M.-S. Jo and E. Sarigöllü, Social value and content value in social media: Two paths to psychological well-being, Journal of Organizational Computing and Electronic Commerce, vol. 27, no. 1, pp. 3-24, 2017.

[21] P. Karhade, M. J. Shaw and R. Subramanyam, Patterns in information sustems portfolio prioritization: Evidence from decision tree induction, MIS Quarterly. vol. 39, no. 2, pp. 413-434, 2015.

[22] M. L. Kasavana, K. Nusair and K. Teodosic, Online social networking: Redefining the human web, Journal of Hospitality and Tourism Technology, vol. 1, no. 1, pp. 68-82, 2010.

[23] F. Kujur and S. Singh, Visual communication and consumer-brand relationship on social networking aites-uses & gratifications theory perspective, Journal of Theoretical and Applied Electronic Commerce Research, vol. 15, no. 1, pp. 30-47, 2020.

[24] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, Tastes, ties, and time: A new social network dataset using Facebook.com, Social networks, vol. 30, no. 4, pp. 330-342, 2008.

[25] Y. Li, L. Lu and L. Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in e-commerce, Expert Systems with Applications, vol. 28, no. 1, pp. 67-77, 2005.

[26] Y.-M. Li, H.-W. Hsiao and Y.-L. Lee, Recommending social network applications via social filtering mechanisms, Information Sciences, vol. 239, no. 2, pp. 18-30, 2013.

[27] C. Lipizzi, L. Iandoli and J. E. R. Marquez, Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies, Technological Forecasting and Social Change, vol. 109, pp. 35-49, 2016.

[28] Y. Liu, J. Wang, Y. Jiang, J. Sun, and J. Shang, Identifying impact of intrinsic factors on topic preferences in online social media: A nonparametric hierarchical Bayesian approach, Information Sciences, vol. 423, pp. 219-234, 2018.

[29] P. Luarn and Y.-P. Chiu. Key variables to predict tie strength on social network sites, Internet Research, vol. 25, no. 2, pp. 218-238, 2015.

[30] P. Luarn, Y.-F. Lin and Y.-P. Chiu, Influence of Facebook brand-page posts on online engagement, Online Information Review, vol. 39, no. 4, pp. 505-519, 2015.

[31] P. Luarn, J.-C. Yang and Y.-P. Chiu, The network effect on information dissemination on social network sites, Computers in Human Behavior, vol. 37, pp. 1-8, 2014.

[32] A. Majumdar and I. Bose, Detection of financial rumors using big data analytics: the case of the bombay stock exchange, Journal of Organizational Computing and Electronic Commerce, vol. 28, no. 2, pp. 79-97, 2018.

[33] C. J. Mantas and J. Abellán, Credal-C4. 5: Decision tree based on imprecise probabilities to classify noisy data, Expert Systems with Applications, vol. 41, no. 10, pp. 4625-4637, 2014.

[34] P. Mikalef, M. Giannakos and A. Pateli, Shopping and word-of-mouth intentions on social media, Journal of Theoretical and Applied Electronic Commerce Research, vol. 8, no. 1, pp. 17-34, 2013.

[35] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, Measurement and analysis of online social networks, in Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, ACAM, San Diego California USA, 2007, pp. 29-42.

[36] S. Moro, P. Rita and B. Vala, Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach, Journal of Business Research, vol. 69, no. 9, pp. 3341-3351, 2016.

[37] M. Nair, Understanding and measuring the value of social media, Journal of Corporate Accounting & Finance, vol. 22, no. 3, pp. 45-51, 2011.

[38] I. Pentina and M. Tarafdar. From information to knowing: Exploring the role of social media in contemporary news consumption. Computers in Human Behavior. vol. 35, no. 2, pp. 211-223, 2014.

[39] C. Ruiz-Mafe, J. Martí-Parreño and S. Sanz-Blas, Key drivers of consumer loyalty to Facebook fan pages, Online Information Review, vol. 38, no. 3, pp. 362-380, 2014.

[40] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, Item-based collaborative filtering recommendation algorithms, In Proceedings of the 10th international conference on World Wide Web, pp. 285-295, 2001.

[41] V. Scuotto, M. Del Giudice, M. R. Della Peruta, and S. Tarba, The performance implications of leveraging internal innovation through social media networks: An empirical verification of the smart fashion industry, Technological Forecasting and Social Change, vol. 120, pp. 184-194, 2017.

[42] K. Sellami, M. Ahmed-Nacer and P. Tiako, From social network to semantic social network in recommender system, Social and Information Networks, vol. 9, no. 4, pp. 1-5, 2014.

[43] Z. Shi, H. Rui and A. B. Whinston, Content sharing in a social broadcasting environment: Evidence from twitter, MIS quarterly, vol. 38, no. 1, pp. 123-142, 2014.

[44] B. Shipps, Social networks, interactivity and satisfaction: assessing socio-technical behavioral factors as an extension to technology acceptance, Journal of Theoretical and Applied Electronic Commerce Research, vol. 8, no. 1, pp. 35-52, 2013.

[45] Y. Tan, H. Zhao, Y. Wang, and M. Qiu, Probability matrix decomposition based collaborative filtering recommendation algorithm, Informatica, vol. 42, no. 2, pp. 265-271, 2018.

[46] S. Walczak and D. G. Gregg, Factors influencing corporate online identity: A new paradigm, Journal of Theoretical and Applied Electronic Commerce Research, vol. 4, no. 3, pp. 17-29, 2009.

[47] L. Yang, D. Jiang, L. He, E. Pei, M. C. Oveneke, and H. Sahli, Decision tree based depression classification from audio video and language information, in Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, Amsterdam the Netherlands, 2016, pp. 89-96.

[48] J. Zhang, Y. Wang and J. Vassileva, SocConnect: A personalized social network aggregator and recommender, Information Processing & Management, vol. 49, no. 3, pp. 721-737, 2013.

[49] Y.-C. Zhang, M. Medo, J. Ren, T. Zhou, T. Li, and F. Yang, Recommendation model based on opinion diffusion, EPL (Europhysics Letters), vol. 80, no. 6, pp. 68003, 2007.

[50] J. Zhao, H. Gao, Y. Li, and J. Liu, Which factors affect the duration of hot topics on social media platforms? Quality & Quantity, vol. 51, no. 5, pp. 2395-2407, 2017.