Analysis of the Influence of Social Media Communication on Consumers' Emotional Tendency in Online Celebrity

Liu Meng, Wan Anita binti Wan Abas

Analysis of the Influence of Social Media Communication on Consumers' Emotional Tendency in Online Celebrity

Liu Meng, Wan Anita binti Wan Abas

Faculty of modern languages and communication,communication department,Universiti Putra Malaysia 43400 UPM Serdang,Selangor Darul Ehsan, Malaysia, School of culture and media , Huanghuai University, Zhumadian, Henan 463000, China
Email: gs62213@student.upm.edu.my
Corresponding author email: anita@upm.edu.my

Abstract
This paper studies the characteristics and composition of consumers' perceived value, puts forward the factors that affect consumers' perceived value on the internet, and defines the driving factors that affect consumers' perceived value on the internet. At the same time, the basic user behavior data is counted through different dimensions, and statistical features, time interval features and computational features are generated. At present, the network economy is developing rapidly, among which online shopping has become an important channel for consumers to buy goods. Based on the behavioral data of social media, this paper analyzes and models the data, and mainly studies the influence of SNS (Social networks) communication on consumers' emotional tendency in online celebrity. In order to verify the correctness of the hypothesis model, this paper designs a questionnaire and tests the algorithm in all aspects. Through the continuous optimization and experiment of features and models, the prediction effect of the model established in this study is stable and good prediction results have been achieved. The verification shows that the precision of this model can reach 0.967; The recall rate can reach 0.941; The value of F1 can reach 0.952. This study is helpful to further improve the economic and social benefits of e-business.
Keywords: Social Media, Spread, Consumers, Emotional Tendency

Introduction
The market size of China's Internet celebrity economy industry is growing year by year and is expected to reach about 200 billion yuan by 2025. In the market segment, short video platform and live streaming platform are the main growth points. China's full-year short video market was worth 100 billion yuan in 2022, according to market research firm Schamp Consulting Group. The market size of live streaming platforms is 60 billion yuan. In terms of policies, the Chinese government has strengthened the supervision of the Internet celebrity economy industry, requiring Internet celebrities to be more standardized in content review, account management and other aspects. At the same time, the Chinese government is also promoting the construction of laws and regulations in the Internet celebrity economy industry to regulate the development of the industry. This paper studies the behavioral data of social

media and the emotional tendency and emotional dependence of social network communication on Internet celebrity consumers. At present, there is no consistent definition of what is online celebrity. This paper believes that online celebrity not only refers to people who have gained high popularity and wide influence in the public by virtue of various Internet media, but also includes abstract concepts, objects, etc. In the online media environment, the comprehensive role of the community of interests such as online celebrities, marketing teams, old media and the psychological needs of the audience has promoted the emergence of online celebrities (Hudson et al., 2016). The online celebrity economy in real business is developing rapidly, and different types of online celebrities appear in public view (Dessart et al., 2015). Under the background of verticalization and communization of social applications, many new online communities based on interest have emerged. Through combing the relevant literature of Internet celebrities, it is found that its development has obvious stage characteristics, and every leap-forward development is inextricably linked with the media form at that time (Benetoli et al., 2017). social media is based on user relations, and its main purpose is to facilitate the completion of social behavior between users. SNS provides an interactive place for business users and consumer users. SNS is based on user relationships, and its main purpose is to facilitate the completion of social behaviors among users. Immersive SNS, as a refined product of SNS in the era of immersion communication, can be regarded as a part of SNS adapting to the development of the times (Oberoi et al., 2017). SNS is also gradually introducing innovative services. Based on the original SNS information dissemination, it introduces a new social media marketing (SMM) concept. The innovation of Internet and mobile technology has spawned SNS, which has ushered in a new era of marketing (Zhao et al., 2018). Compared with the simplistic old media, the Internet has the characteristics of openness and freedom, and everyone can become a communicator in the Internet context. The fiery economy of online celebrity has driven the birth of a new business model. Internet users on the e-business platform gain profits by opening online stores and carrying out product promotion commissions (Zhu et al., 2016). "online celebrity+E-commerce" will become an important way for online celebrity to seek cash, and also a means for merchants to promote their marketing. Online celebrity's choice of live webcast marketing mode to improve the popularity and influence of platforms, brands or individuals, and realize the transaction through the flow, thus achieving its marketing purpose, has great potential and realizing ability (Rahman et al., 2018). In the process of making consumption decisions, consumers have the ability to perceive goods, evaluate goods and make purchasing decisions. People can participate in interest communities on the Internet according to their own interest points, and get social fun through the exchange of experience information. The easy manipulation of socialized media also encourages people to produce online content, and the threshold of professional media content production is gradually broken.

Marketing in social media platforms is more socialized (Kumar et al., 2013). Users choose websites for information browsing with independent selectivity, and there is no employment relationship between SNS and users. Therefore, when users choose SNS and browse advertising information, they can choose to watch product marketing advertisements according to their preferences (Shao et al., 2012). When the key factors that affect consumers' decision-making during watching live broadcast are made clear, the marketing team can know more about consumers (Staccini, & Fernandez-Luque, 2016). Affective analysis is a branch of natural language processing. It is a method of automatic text classification based on potency. It mainly studies the positive or negative emotions expressed or implied in ideas. In the wave of mobile Internet, users' behaviors in the network environment account for an increasing

proportion. Users' click, browse, purchase, comment and other behaviors on SNS have become very important reference data for service providers (Leung et al., 2013). In order to achieve accurate prediction of user behavior. This paper takes the behavior data of SNS as the starting point, analyzes and models the data, and mainly studies the impact of online social media communication on consumers' emotional tendencies. Its innovations are as follows:

① This paper systematically reveals the influence of brand communication based on SNS on consumers' emotions. On this basis, the influencing factors of consumers' purchasing interest under SMM are deeply discussed, and the methods of constructing SMM strategy by analyzing consumers' purchasing intention are sorted out by combining with the characteristics of SNS platforms. It provides necessary decision-making basis and support for the development of SMM-based e-business for reference of enterprises.
② In this paper, the decision tree idea is used, and the feature selection method based on gradient descent tree is used to obtain the feature importance. At the same time, the basic characteristics of users and commodities are extracted by encoding the basic data. The basic user behavior data is counted through different dimensions, and statistical features, time interval features and computational features are generated.

This paper mainly includes five sections, namely, introduction, definition of related concepts and basic theory, model construction, model evaluation and so on. The specific contents are as follows: The first section is an introduction. The introduction briefly introduces the research background, significance and innovation of this paper. The second section summarizes the related research status, and briefly describes the research content and research methods of this paper. The third section is the focus of the whole research work. This section first summarizes the related content of SNS in online celebrity; Then it discusses the influence of brand communication of SNS on consumers. Finally, the analysis of consumers' purchase intention is given, and the corresponding model is constructed. The fourth section introduces the sampling ideas and data processing methods of this research data model. At the same time, cross-validation is carried out and training parameters are optimized. Compare the prediction results and evaluate and optimize the model. The fifth section is the research conclusion and prospect. This section first summarizes the conclusions of the study, and then points out the limitations and future research directions of this study.

Related Work
The background of the rise of network stars with the same attributes as "online celebrity" from many angles, and believed that their fame was mainly generated from bottom to top in an anti-mainstream way through events hype, song singing, graphic communication and other ways (Keegan & Rowley, 2017; Wang et al., 2014). Research by Blaseg et al (2020) showed that the form and content characteristics of corporate social networking site accounts such as interaction, interactivity, and self-exposure level can indirectly affect users' willingness to participate through mediating factors such as users' social presence and perception of information usefulness.Tha (2013) has pointed out that in addition to the impact of SNS on individuals, it also has a positive impact on the development of social capital. Users' in-depth use of SNS applications and self-presentation on the platform can generate substantial social capital. For advertisers, advertising on SNS has become a new way of marketing. Advertising in SNS achieves intimate interaction with the audience through precise positioning, enhances the communication effect of advertising, and presents fission-type communication (Qutteina et al., 2019). From the perspective of Ng (2013) semiotics, this paper deeply analyzes the formation reasons, transmission ways and social influence of "online celebrity". Social

presence can improve consumers' trust, safety and risk perception, which in turn affects final purchase intentions and behaviors. It is believed that social presence is a key factor for consumers to shop online. Chen et al (2015) think that the essence of "online celebrity" economy is a type of fan economy subdivision, and it is a form of transforming fan trust into purchasing power . The study by Dehghani and Tumer (2015) is based on the TAM model, but introduces more variables. Such as the user's access time to the network, the user's preference for risk, the user's perception of the entertainment of the network, etc. When analyzing and processing the collected data, factor analysis and multiple regression analysis were selected in this study. The results of the data analysis support the main hypotheses put forward by the study. Taking the rapid rise of e-business in the Internet age as the background, Gottlieb et al (2011) put forward several major problems that must be faced in the development of the Internet celebrity economy, including the short lifespan of "Internet celebrities", the stable market of the real economy and the industrial model of the Internet celebrity economy. Uncertainty, etc. Duffett (2015) pointed out that the number of SNS users is huge and continues to grow, and they generate countless content. How to quickly and accurately identify brands from big data is difficult and requires high speed and accuracy. Hsu et al (2017) proposed that in the network environment, consumers buy goods online, which is based on the recognition of the storefront design of online shopping malls. The pros and cons of virtual store interface design directly affect customers' purchasing desire. If the design of the virtual store interface conforms to the aesthetic psychology of customers, it is convenient for customers to search for information and purchase goods, which will help customers to build a sense of trust in the company and increase their willingness to buy. Research by Tiwary et al (2021) believes that the level of brand self-disclosure and the social connectivity of social networking site platforms affect users' attitudes towards the brand community of social networking sites. Simon and Tossan (2018) pointed out that SNS also changes the social communication mode of college students, promotes their active participation in social activities, and promotes the formation of personal values, but it also causes college students to rely on mobile phones excessively. Chang et al (2015) believe that SNS is an interactive virtual space. Scholz et al (2015) pointed out that convenience is the advantage of hidden promotion in SMM. SMM simplifies shopping, saving time and effort. Using SNS shopping, consumers not only hope to complete the shopping, but also hope to get various consumption pleasures while shopping.

Based on this, this paper takes the behavior data of SNS as the starting point, analyzes and models the data, and mainly studies the influence of SNS communication on consumers' emotional tendency in online celebrity. This paper analyzes the characteristics of online consumers' purchase intention, and expounds that external factors and internal perception factors are the main factors that affect consumers' purchase intention in SMM situation. At the same time, the basic characteristics of users and commodities are extracted by encoding the basic data. The basic user behavior data is counted through different dimensions, and statistical features, time interval features and computational features are generated. Through the continuous optimization and experiment of features and models, the prediction effect of the model established in this study is stable and good prediction results have been achieved. It provides necessary decision-making basis and support for the development of SMM-based e-business for reference of enterprises.

Methodology

Research on SNS in online celebrity

"Socialization" refers to people's interaction in social life, that is, people transmit and exchange information and ideas in a certain way in order to achieve a certain purpose (Duan & Dholakia, 2015). However, it is still an important feature of SNS to give users great creativity and communication ability. SNS makes use of the convenience of the network, and combines social communication with media information dissemination, so that people can enjoy social entertainment while browsing and interacting with information. Social network can be regarded as a cyberspace. SNS provides an interactive place for business users and consumer users. SNS is based on user relationships, and its main purpose is to facilitate the completion of social behaviors among users. Immersive SNS, as a refined product of SNS in the era of immersion communication, can be regarded as a part of SNS adapting to the development of the times. They are the relationship between inclusion and inclusion, which have both similarities and some differences. Immersive SNS not only has the general attributes of SNS, but also emphasizes its unique immersion characteristics. From the perspective of enterprises, it is necessary to study the influence of social media communication on consumers' emotional tendency, so as to analyze consumers' purchasing intention and provide necessary decision-making basis and support for the development of e-commerce.

There are various SMM modes. It contains not only words, pictures, audio and video, but also the penetration of marketing information through SNS interaction, which is a widely used marketing mode in the Internet eral (Oh et al., 2017). Based on this, social networks can be roughly divided into the following four types: ① Interest needs. ② Social needs. ③ Transaction demand. ④ Fantasy needs. Nowadays, SNS gives consumers great autonomy, and at the same time, consumers pay more attention to active search and sharing. The characteristics of socialization, such as participation, communication and community, have completely changed the way people participate in the network. Get rid of the geographical and time constraints, people's social life can be carried out through SNS. In the process of making consumption decisions, consumers have the ability to perceive goods, evaluate goods and make purchasing decisions. People can participate in interest communities on the Internet according to their own interest points, and get social fun through the exchange of experience information. The easy manipulation of socialized media also encourages people to produce online content, and the threshold of professional media content production is gradually broken. The motivation of consumers' demand for goods is an important internal reason to promote consumers' purchasing decisions. Any communication activity can only be realized with the help of media. Online celebrity was born on the Internet, but it is also popular with the help of network media. Online celebrity expresses himself through the media, attracts fans' attention through the media, and even makes a profit through the media. Its development and dissemination have typical media brand. SNS is user-centered and focuses on user experience. SNS features refer to the features of content and design on SNS that can meet users' needs. The characteristics of SNS are as follows: ① It is user-centered. ② Its essence is a platform for providing comprehensive services. ③ It exists and develops to encourage communication and interaction to be more efficient. ④ It makes the traditional face-to-face offline communication become a reality. ⑤ It is the channel for users to build their personal image in the network.

Analysis of the influence of brand communication of SNS on consumers' emotion

In the modern environment, everyone can be a disseminator of information. The main body of information dissemination is no longer concentrated in the hands of people or institutions with professional knowledge, and information dissemination is no longer linear, but radiates outward with countless human nodes. Everyone is spreading information. Trust is an essential and important condition in the operation of modern commercial activities. Especially in e-business, the quality of goods purchased online can't be tested in time, and the goods purchased online lack the necessary bill guarantee, so trust becomes more important. Strengthening the trust between enterprises and consumers is helpful to improve consumers' desire to buy. Brand assets are brand assets or liabilities associated with a brand name or logo, which can increase or decrease the value of the company that provides this product or service and the customers who buy this product or service. SMM model is not a traditional sales to the public, but a personalized and innovative marketing activity through SNS platforms. Brand marketing activities based on SNS can influence consumer's brand attitude and then consumer-based brand equity. In addition, participation gives consumers greater control. Participation is regarded as an active behavior. Consumers can control the content and frequency of browsing and searching. Participation gives consumers a stronger sense of control. Figure 1 shows the influence mechanism model of brand communication on consumers based on SNS.
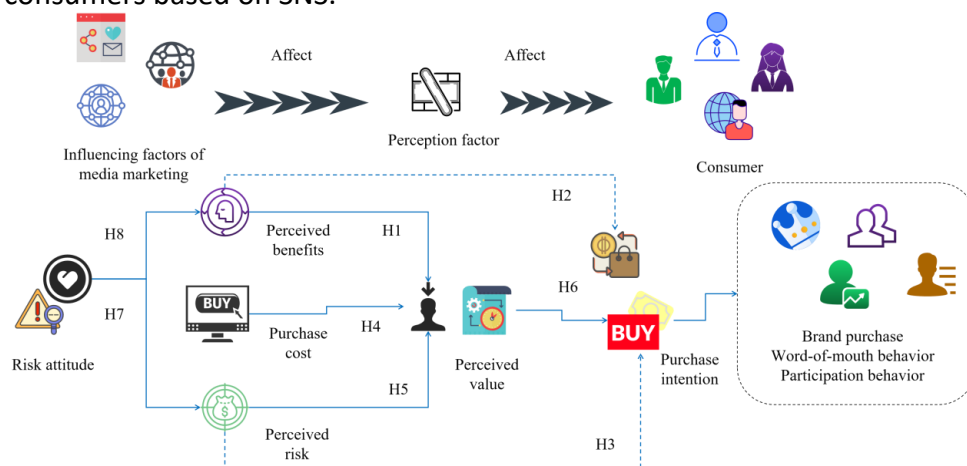

Figure 1. Model of the influence mechanism of brand communication on consumers based on SNS.

In the development of immersive SNS, the communication subject is not only limited to the specific media platform, but also called a part of the communication subject. The communication tools, communication technologies and communication relationships involved in the communication platform that provide users with a strong sense of immersion in the communication process enrich the connotation of the communication subject. According to the theory of rational behavior, consumers' behavior intention comes from consumers' subjective standards, that is, values and attitudes towards commodities. Consumer's subjective standard is the standard for consumers to judge objective things, and consumer's attitude towards goods is consumer's judgment on the value of goods based on subjective standard. Therefore, both subjective standard and attitude towards goods are consumers' subjective cognition and opinions. However, there is a kind of users in SNS in online celebrity, who have a large number of fans, and their comments have a wide influence, commonly known as opinion leaders. Opinion plays an important role in the diffusion and

adoption of products. E-business generally lacks social warmth, and SNS can improve consumers' purchasing level by providing them with social presence experience. For enterprises, only when consumers' desirable behavior changes can their brand communication activities based on SNS create realistic economic value.

Perceived value is a concept based on customers' subjective perception, which reflects consumers' attitude and evaluation towards products or services. Perceived value is highly subjective, and at the same time, perceived value is practical and dynamic. SNS based on mobile Internet has also entered the vertical stage after "barbaric growth" to gather a huge number of users. The development of SNS has been integrated into all aspects of people's lives. The spread of "online celebrity" is closely related to the development of SNS. SNS has become an important means for individuals to build network relationships. Community is a new relationship group based on social network. Under the background of verticalization and communization of social applications, many new online communities based on interest have emerged. Through combing the relevant literature of Internet celebrities, it is found that its development has obvious stage characteristics, and every leap-forward development is inextricably linked with the media form at that time. Every development of media forms will lead to the advancement of online celebrity. Due to the enhancement of openness and interactivity of various vertical SNS, the same user may belong to many different communities, and different communities often form different degrees of association among different communities because of the relationship among community members. Word-of-mouth information on SNS has higher credibility because it is more spontaneous and basically won't be interfered by enterprises. Therefore, brand communication based on SNS is essentially a brand co-creation practice involving both enterprises and consumers. The influencing factors of SMM influence consumers' purchase intention through intermediary factors, which to some extent reflects the influence trend of SMM on consumers' purchase intention.

Analysis of consumers' purchase intention and model building

This paper takes the behavior data of social media as the starting point to analyze and model the data. Whether the model is established and how much influence each variable has need to be estimated and verified by using sample data. In the hypothetical model, there are six potential variables. Dirty data refers to some data in the original system that is beyond the given range, or meaningless for research purposes, or the data format is illegal and the coding is not standardized. Data cleaning refers to washing away these "dirty data" in the source data. Data cleaning mainly includes abnormal value processing, de duplication processing and noise processing. After a comprehensive analysis of the data, the statistical features to be extracted include the following: ① basic information features. ② Statistical characteristics of six behaviors in different time slices. ③ Statistical characteristics of the number of goods operated by users in different time slices. ④ Statistical characteristics of the number of brands operated by users in different time slices.

The process of data standardization is the process of scaling the data, and the standardized data will fall within a small specific interval. The most typical standardization method is the Z-score method, which can make the eigenvalues obey the standard positive distribution after processing. The transformation formula is:

$$x' = (x - \mu) / \sigma \qquad (1)$$

Where x and x' represent the features before and after normalization, respectively. μ is the sample mean; δ is the sample standard deviation, and these two values can be calculated from the sample. When there are enough samples, the z-score normalization method

performs well and is suitable for processing modern data-rich and noisy scenes. Combined with z-score normalization, the average and standard deviation of the attribute A are normalized. The value v of the normalized A becomes v', which is calculated by:

$$v' = \frac{v - \overline{A}}{\sigma_A} \qquad (2)$$

Where the mean of A is:

$$\overline{A} = \frac{\sum A}{n} \qquad (3)$$

The standard deviation of A is:

$$\sigma_A = \sqrt{\frac{\sum (A - \overline{A})^2}{n - 1}} \qquad (4)$$

After normalization, all variables have a mean value of 0 and a variance of 1. The interval scaling method uses the boundary value of feature data to make the feature value in a specific range. This paper uses two maxima to scale the feature data. Its transformation function is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (5)$$

Among them, $\min(x)$ and $\max(x)$ represent the minimum and maximum values of the sample data, respectively. This method can make the data fall between [0,1]. Binarization is to set a threshold T. Values greater than T are assigned the value 1, and values less than T are assigned the value 0. The formula is expressed as follows:

$$x' = \begin{cases} 1, & x > threshold \\ 0, & x \le threshold \end{cases} \qquad (6)$$

Firstly, according to the existing theories and research results, this paper establishes the initial model of hypothesis, and describes the assumed causal relationship between variables through the path diagram. This can not only express the model in a direct and clear way, but also transform the model into a modeling equation through the road map, so that the model can be built on a correct theoretical basis. In this model, the influencing factors of SMM include product placement, activity marketing, experience marketing and interaction. The influencing factors of SMM influence consumers' purchase intention through intermediary factors, which to some extent reflects the influence trend of SMM on consumers' purchase intention.

Attenuation is characterized by the summation of the attenuation of statistical behaviors over time, as shown in the following formula:

$$F(x, y) = \sum f(x, y) = \sum \left[y * \log\left(1 + \frac{1}{x+1}\right) \right] \qquad (7)$$

Among them, $f = \log\left(1 + \frac{1}{x+1}\right)$ is the decay function; y represents the number of actions of the user in the current day/hour; x represents the time interval from the current day/hour to the day before the prediction day. The decay function is shown in Figure 2. The horizontal axis represents the ratio of the number of actions to the time interval, and the logarithmic function is gradient descent.
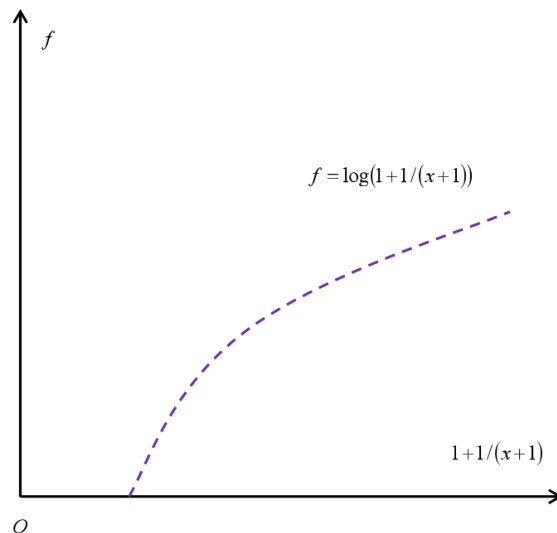
Figure 2. Attenuation function diagram.

According to the original data, some attribute values can't exceed the specified range. If they exceed this range, they will be wrong data. If the wrong data is input into the database, it will seriously affect the data mining effect, so it is necessary to preprocess these data. In practical projects, data missing is a common problem. For missing values, the simplest way is to delete them directly. In the database, data are organized together in the form of tables. For missing values, just delete the records with missing values. In this paper, the missing value is filled with the most possible value. In this study, the concept of "time slicing" was introduced in the feature extraction stage: some studies also used "sliding window" to expand the feature dimension. Due to the short time limit of data, the sliding window technique is not adopted in this paper. Figure 3 shows the data set partitioning rules.
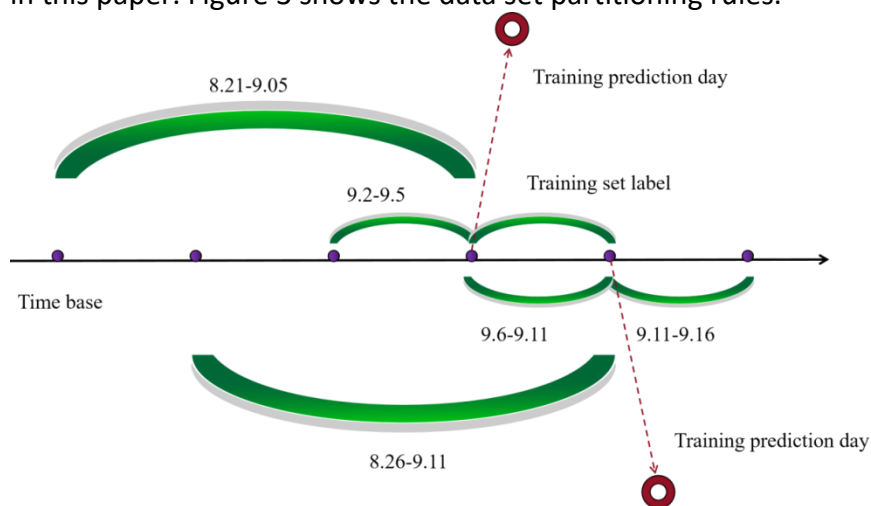

Figure 3. Data set division rule diagram.

Model identification mainly considers whether each free parameter in the model can get the unique solution from the observed data. Therefore, it is necessary to check whether the established model belongs to just-recognized structural model, under-recognized structural model or over-recognized structural model. Identification is carried out through the parameters of the linear equation. The necessary but insufficient condition for model identification is that the number of parameters of the model is not more than the number of observed variance and covariance. Therefore, the model should be tested for identifiability

according to this necessary condition. In this paper, the decision tree idea is used, and the feature selection method based on gradient descent tree is used to obtain the feature importance. In the process of interaction, everyone will perform in various ways, intentionally or unintentionally, in order to maintain, strengthen or change others' impression of themselves. In the Internet, due to the anonymity and virtuality of cyberspace, netizens can manage their impressions more easily, play any role, and reshape their image without any restrictions. "Information entropy" is the average amount of information after eliminating redundancy, which indicates the degree of information confusion. The formula for calculating information entropy is:

$$I(S_1, S_2, S_3, \ldots, S_m) = -\sum_{i=1}^{m} P_i \log_2 (P_i) \qquad (8)$$

Among them, S is the sample set, which defines m different classes $C_i (i = 1,2,3,\ldots,m)$; S_i is the number of samples in the C_i class, and P_i is the probability that any sample belongs to C_i: S_i / S. Using information gain to select sample classification attributes, the information gain obtained after branch A is as follows:

$$Gain(A) = I(S_1, S_2, S_3, \ldots, S_m) - E(A) \qquad (9)$$

Among them, $I(S_1, S_2, S_3, \ldots, S_m)$ is the expected information entropy of a given sample S, and the calculation formula of $E(A)$ is as follows:

$$E(A) = -\sum_{j=1}^{v} \frac{S_{1j} + S_{2j} + S_{3j} + \ldots + S_{mj}}{s} I(S_{1j} + S_{2j} + S_{3j} + \ldots + S_{mj}) \qquad (10)$$

Information gain is the deviation of variable values when the data set is divided into small subsets. To reduce this bias, use the information gain rate formula:

$$SplitInfo\ (S, v) = \sum_{i=1}^{m} \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|} \qquad (11)$$

Thus, the gain ratio can be obtained:

$$GrainRatio\ = \frac{Grain(S, v)}{SplitInfo\ (S, v)} \qquad (12)$$

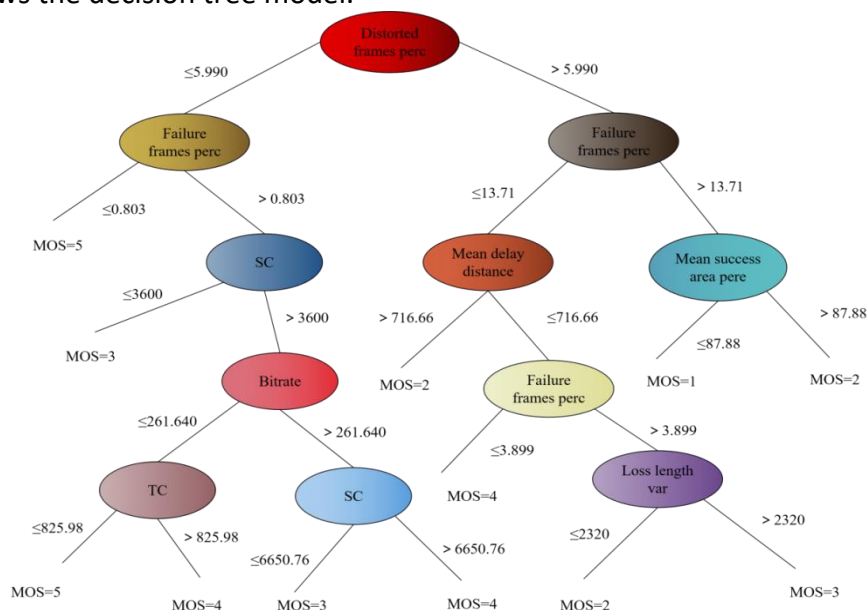Figure 4 shows the decision tree model.

Figure 4. Decision tree model.

In this paper, the data set is downsampled at first, and the training set obtained by this sampling is used to train a classifier. Instead of putting back the mass sample with correct classification, the mass sample with smaller scale is downsampled again to train the second classifier … and so on. Finally, the results of multiple classifiers are combined to get the final result. Part of the database is relatively large in scale, data analysis, mining and other links are complicated, which takes a long time, and the mining results are prone to deviation. In view of this situation, we can compress the data set by data reduction technology. The compressed data set is smaller than the source data set, but the original data must remain intact. After the data set specification, the mining will be effective and the analysis results will be consistent.

Result Analysis and Discussion

As an external marketing stimulus, enterprise brand communication based on SNS can affect consumers' internal cognitive, emotional and attitude psychological reactions, and then affect consumers' brand or brand community related behaviors. This paper selects different groups of people as the survey object. Through 500 questionnaires distributed to consumers using the SMM platform, 489 questionnaires were collected, including 243 college students and 246 urban working people. After review, 8 of them have never used SNS shopping, and 3 of them have the same answer to the questionnaire, so they are excluded. The final valid questionnaires were 478. The sample size of this questionnaire meets the requirements of structural equation model analysis. The experimental environment of this paper is configured with 8g memory; Hard disk 50g; Cpu:i7-3630 or above. The system environment is centos7.0 and the software environment is python. Cronbach'sα reliability coefficient is adopted in this paper. Table 1 shows the results of reliability and validity analysis.

Table 1
Reliability and validity analysis

Index		α	KMO
External influence factors	Implantable advertising	0.809	0.765
	Marketing activities	0.871	0.892
	Experiential marketing	0.806	0.687
	Interactivity	0.810	0.826
Intrinsic perception	Perceived value	0.889	0.734
	Perceived risk	0.879	0.782
Purchase intention	Intention	0.863	0.768

As can be seen from Table 1, Cronbach'sAlpha values of the whole questionnaire and each sub-scale are all greater than 0.5, and the internal consistency of the questionnaire is good, which indicates that the questionnaire design meets the requirements and the data have good correlation, so the sample data can be analyzed by factor analysis.

Before fitting the sample data with the model, the sample data must satisfy some assumptions of the structural equation model. These assumptions include: linear relationship, multivariate normal distribution, no peak value and skewness value, and data should be measured at a fixed distance or scale. In addition, in order to avoid the over-fitting of the learning model, the samples are usually divided into three types: training set, verification set

and test set. In this paper, the samples are divided into the following proportions: training set is 60%, cross-validation set is 20%, and test set is 20%. At the same time, the proportion of positive and negative samples in each set should be as consistent as possible. Firstly, this paper sorts out the noise lexicon of denoised scripts from the sampled samples. Secondly, according to the denoising script, the whole amount of data is screened, and the data with noise words are deleted. Figure 5 shows the discussion trend of a topic in SNS. The abscissa is time, and the ordinate is the amount of discussion.
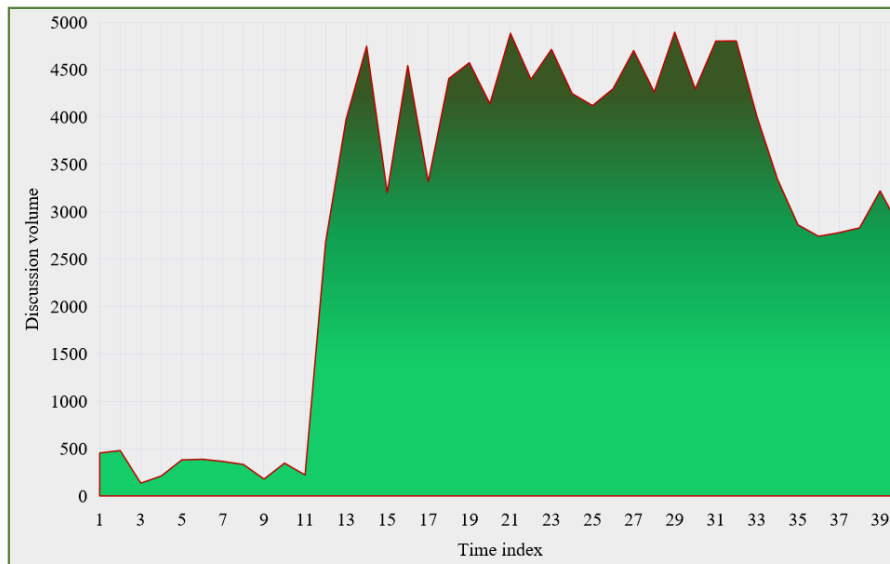

Figure 5. Trend of discussion of a topic on SNS.

It can be seen that, driven by key opinion leaders, consumers have launched a heated discussion on this topic. During the period of data analysis and interception, the amount of discussion increased greatly.

In this paper, the user behavior data of 100 consecutive days are selected for feature analysis, and five categories of features are extracted according to the results of feature analysis. In the data preprocessing stage, the null values in the features are processed, and the one-hot coding method is adopted for individual category features. After dimension reduction, the 640-dimensional features are finally determined. In this paper, AMOS software is used for correlation and regression analysis of the hypothetical model, and the data analysis results are shown in Table 2. Generally, the absolute value of the free parameter t should be greater than 1.9, which means that the significance level reaches 0.05.

Table 2
Model Validation Structural Analysis

Hypoth esis	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Coeffici ent	0.921	-0.420	0.979	-0.522	-0.748	-0.736	-0.639	-0.657	-0.882	-0.756
T value	4.782	-2.897	4.859	-3.882	4.693	-2.108	4.637	-3.112	3.784	4.559
Conclus ion	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort	Supp ort

The following is the training and evaluation of the model. The evaluation indicators of the model used in this study include: ① Recall rate: the ratio of positive cases reflecting the predicted pair to the real positive cases. ② Accuracy: refers to the accuracy between the obtained value and the real value; The ratio of correctly predicted positive cases to the total predicted positive cases. ③F value: F-score is the weighted harmonic average of Precision and Recall. Figure 6 shows the experimental results of recall rate of different algorithms. Figure 7 shows the experimental results of precision of different algorithms. Figure 8 shows the experimental results of F values of different algorithms.
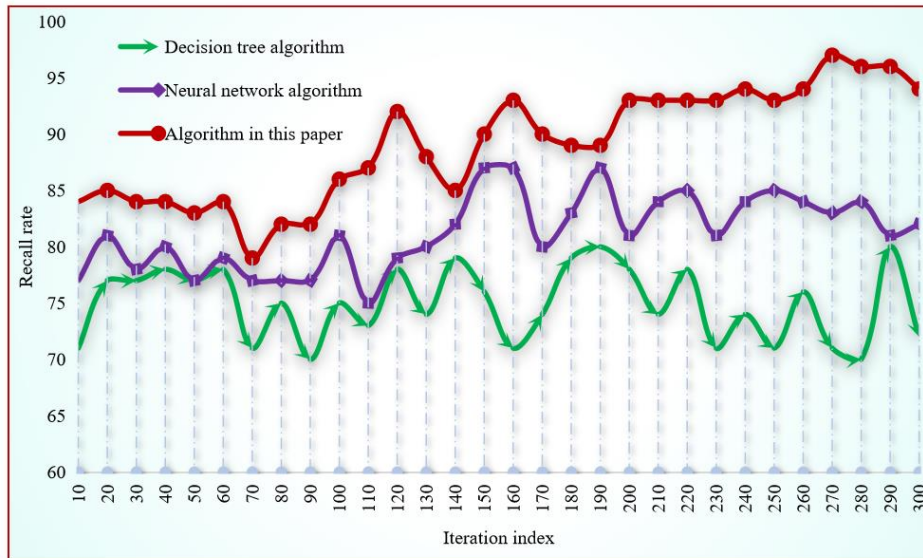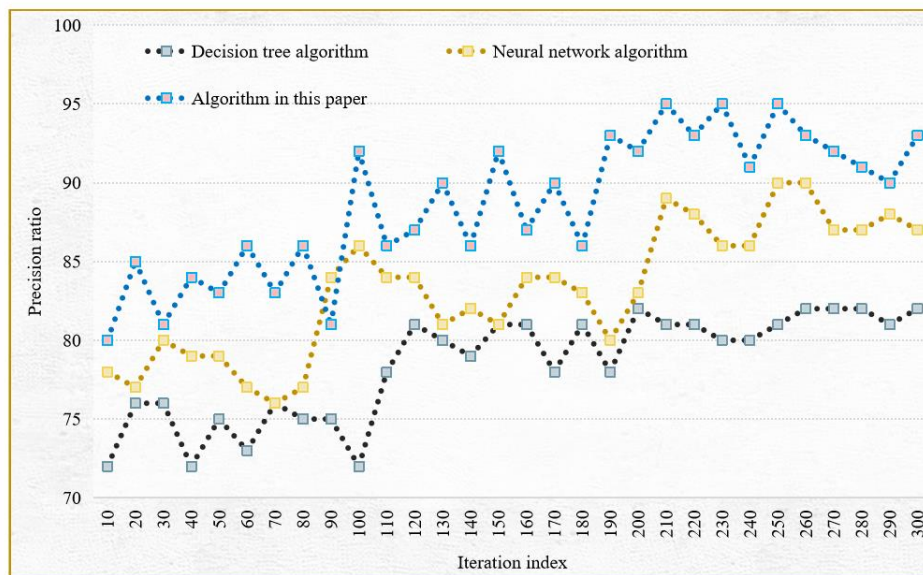

Figure 6. Recall experiment results.

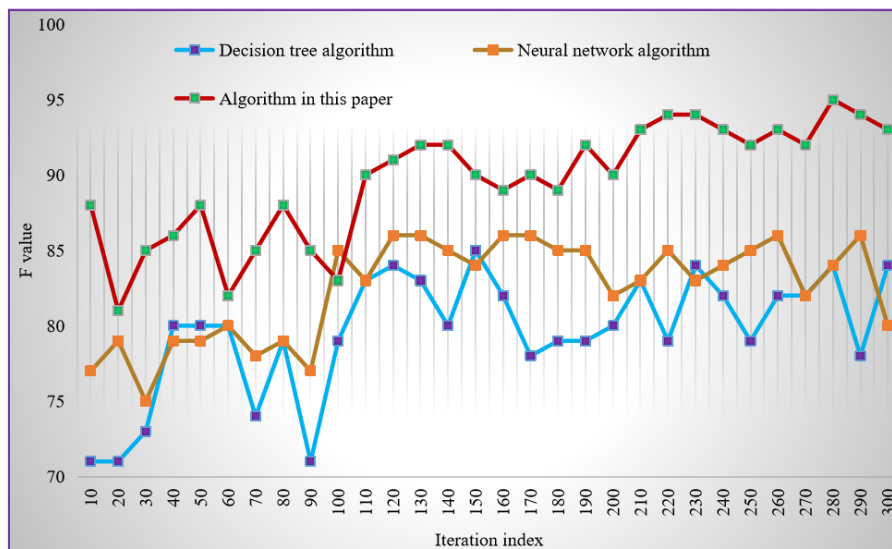Figure 7. Precision experimental results.

Figure 8. F value experimental results.

Through the analysis of the prediction results, the score of this algorithm is quite good, and the precision, recall and F value all reach the expected values. Data mining mainly analyzes different dimensions of acquired data, and constructs dimensions that can be quantitatively analyzed according to the "subject-feature-emotion" triad. Based on the data information of consumers' gender, location, publishing platform, publishing time, etc., multi-level portraits of users are depicted. The effectiveness of the prediction model in the actual application scenario should be considered. The training and prediction of feature extraction model should be completed in one day, and then the prediction results should be pushed out. The task running time of the whole process should be controlled within four hours. Therefore, after counting and analyzing the relationship between feature dimension and model training prediction time, this paper continues to reduce dimension.

In order to verify the accuracy and consistency of evaluation results of different models, this paper calculates and compares three evaluation indexes, including Pearson correlation coefficient (PCC), Spearman rank correlation coefficient (SCC) and outlier ratio (OR). Table 3 shows the experimental results of evaluation indexes of different models.

Table 3
Experimental results of evaluation indexes of different models

Algorithm	PCC	SCC	OR
K-means model	0.497	0.551	0
Neural network model	0.529	0.638	0
Model of this paper	0.857	0.849	0

The comparison between the predicted results in this paper and the actual results is shown in Figure 9.
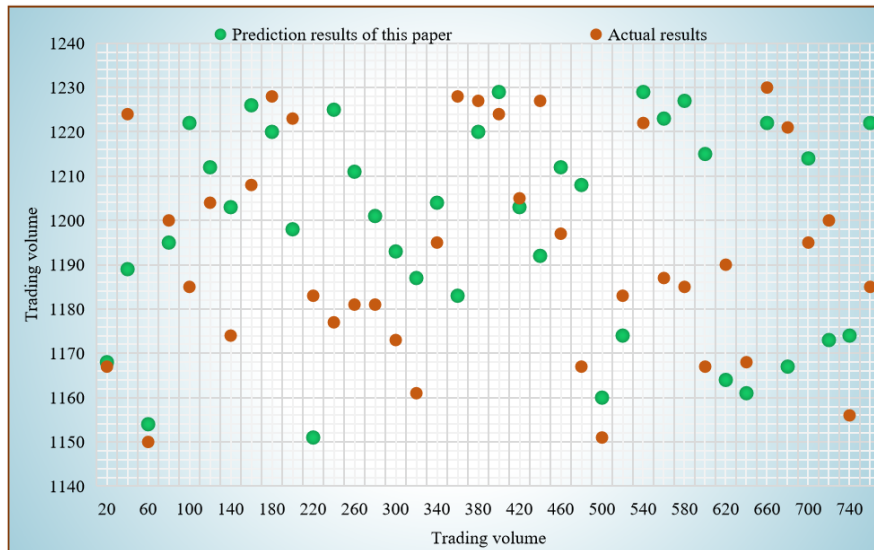
Figure 9. Comparison of the predicted results in this paper and the actual results.

After testing the realization of the default parameters of the model, we can optimize the parameters to see if the accuracy of the model can be improved. In this paper, GridSearchCV is used to find the optimal parameters. This method determines the best parameters by traversing the combination of various parameters of the model and cross-verifying, which is more suitable for small data sets. The optimal parameter results of GridSearchCV are shown in Table 4.

Table 4
Parameter result table

Parameter name	Optimal value	Optimal score
n_estimators	140	0.877
max_depth	'max_depth': 5	0.879
min_samples_split	'min_samples_split': 50	
min_samples_split	'min_samples_leaf': 50, 'min_samples_split': 80	0.886
min_samples_leaf	45	0.874
max_features		

Data training and prediction are carried out by using the optimized parameters. The training results are shown in Table 5.

Table 5
Training prediction results table

Index	Numerical value
Precision ratio	0.967
Recall rate	0.942
F1	0.957
AUC	0.91

From the prediction results of the two scenarios, the prediction accuracy of both known data and unknown data has decreased to some extent, but the deviation of prediction accuracy between unknown data and known data has become smaller, which proves that the generalization ability of the model has become stronger. Meanwhile, the prediction accuracy is improved. Through the continuous optimization and experiment of features and models, the prediction effect of the model established in this study is stable and good prediction results have been achieved. The verification shows that the precision of this model can reach 0.967; The recall rate can reach 0.941; The value of F1 can reach 0.952. The results show that this method has certain reliability.

According to the research, to enhance the amount of SMM advertising information, we can quickly design through the analysis results obtained by data mining, and realize the automatic push of marketing advertisements through data mining, so that the whole marketing promotion process will be more closely linked and the promotion quality will be improved. According to the cultivation theory of communication, the "symbolic reality" prompted by mass media in modern society has a great influence on people's understanding of the real world. At the same time, the basic characteristics of users and commodities are extracted by encoding the basic data. The basic user behavior data is counted through different dimensions, and statistical features, time interval features and computational features are generated.

Conclusions

From the perspective of enterprises, it is necessary to study the influence of social media communication on consumers' emotional inclination, so as to analyze consumers' purchase intention and provide necessary decision-making basis and support for the development of e-business. Brand communication based on corporate social media is a new thing. Therefore, only by comprehensively using various methods can researchers more profoundly reveal the internal laws of its influence mechanism and effect on consumers. This paper systematically reveals the influence of corporate brand communication based on SNS on consumers' emotions. On this basis, the influencing factors of consumers' purchasing interest under SMM are deeply discussed, and the methods of constructing SMM strategy by analyzing consumers' purchasing intention are sorted out by combining with the characteristics of SNS platforms. At the same time, the basic characteristics of users and commodities are extracted by encoding the basic data. The basic user behavior data is counted through different dimensions, and statistical features, time interval features and computational features are generated. Through the continuous optimization and experiment of features and models, the prediction effect of the model established in this study is stable and good prediction results have been achieved. The verification shows that the precision of this model can reach 0.967; The recall rate can reach 0.941; The value of F1 can reach 0.952. The results show that this method has certain reliability. In the network marketing, consumers can be better stimulated to buy again by recommending according to their past consumption habits. This is of great help to accurately capture consumer demand, and has a good guiding role in the marketing strategy decision of the operation team. In addition, the features extracted in this study are all combinations of basic attributes, and in the future work, we will continue to optimize them from the perspectives of features and models.

References

Dessart, L., Veloutsou, C., & Morgan-Thomas, A. (2015). Consumer engagement in online brand communities: A Social Media Perspective. Journal of Product & Brand Management, 24(1), 28-42. doi:10.1108/jpbm-06-2014-0635

Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer–brand relationships: A three-country study of Brand Perceptions and marketing behaviors. *International Journal of Research in Marketing, 33*(1), 27-41. doi:10.1016/j.ijresmar.2015.06.004

Benetoli, A., Chen, T. F., & Aslani, P. (2017). Consumer health-related activities on social media: Exploratory study. *Journal of Medical Internet Research*, *19*(10). https://doi.org/10.2196/jmir.7656

Oberoi, P., Patel, C., & Haon, C. (2017). Technology sourcing for website personalization and social media marketing: A study of e-retailing industry. *Journal of Business Research, 80*, 10-23. doi:10.1016/j.jbusres.2017.06.005

Zhao, R., Geng, Y., Liu, Y., Tao, X., & Xue, B. (2018). Consumers' perception, purchase intention, and willingness to pay for carbon-labeled products: A case study of chengdu in China. *Journal of Cleaner Production, 171*, 1664-1671. doi:10.1016/j.jclepro.2017.10.143

Zhu, Z., Wang, J., Wang, X., & Wan, X. (2016). Exploring factors of user's peer-influence behavior in social media on purchase intention: Evidence from QQ. *Computers in Human Behavior, 63*, 980-987. doi:10.1016/j.chb.2016.05.037

Rahman, Z., Moghavvemmi, S., Suberamanaian, K., Zanuddin, H., & Bin Md Nasir, H. N. (2018). Mediating impact of fan-page engagement on social media connectedness and followers purchase intention. *Online Information Review, 42*(7), 1082-1105. doi:10.1108/oir-05-2017-0150

Kumar, V., Bhaskaran, V., Mirchandani, R., & Shah, M. (2013). Practice prize winner—creating a measurable social media marketing strategy: Increasing the value and roi of intangibles and Tangibles for Hokey Pokey. *Marketing Science, 32*(2), 194-212. doi:10.1287/mksc.1120.0768

Shao, J., Li, X., & Li, Z. (2012). Micro-film marketing of Chinese tourism destinations via social media: The case of Shaoxing. *2012 IEEE 14th International Conference on Commerce and Enterprise Computing*. doi:10.1109/cec.2012.36

Staccini, P., & Fernandez-Luque, L. (2016). All that glitters is not gold: Consumer health informatics and education in the era of social media and Health Apps. *Yearbook of Medical Informatics, 25*(01), 188-193. doi:10.15265/iy-2016-045

Leung, X. Y., Bai, B., & Stahura, K. A. (2013). The marketing effectiveness of social media in the hotel industry. *Journal of Hospitality & Tourism Research, 39*(2), 147-169. doi:10.1177/1096348012471381

Keegan, B. J., & Rowley, J. (2017). Evaluation and decision making in social media marketing. *Management Decision, 55*(1), 15-31. doi:10.1108/md-10-2015-0450

Wang, F., Zheng, P., Freeman, B., & Chapman, S. (2014). Chinese tobacco companies' Social Media Marketing Strategies. *Tobacco Control, 24*(4), 408-409. doi:10.1136/tobaccocontrol-2014-051714

Blaseg, D., Schulze, C., & Skiera, B. (2020). Consumer Protection on kickstarter. *Marketing Science, 39*(1), 211-233. doi:10.1287/mksc.2019.1203

Tham, J. (2013). Book review: *the zen of social media marketing: An easier way to build credibility, generate buzz, and increase revenue*, by Shama Hyder Kabani. *Journalism & Mass Communication Quarterly, 90*(3), 590-591. doi:10.1177/1077699013496037

Qutteina, Y., Hallez, L., Mennes, N., De Backer, C., & Smits, T. (2019). What do adolescents see on social media? A diary study of Food Marketing Images on social media. *Frontiers in Psychology, 10*. doi:10.3389/fpsyg.2019.02637

Ng, C. S. (2013). Intention to purchase on social commerce websites across cultures: A cross-regional study. *Information & Management, 50*(8), 609-620. doi:10.1016/j.im.2013.08.002

Chen, M., Yin, S., Xu, Y., & Wang, Z. (2015). Consumers' willingness to pay for tomatoes carrying different organic labels. *British Food Journal, 117*(11), 2814-2830. doi:10.1108/bfj-12-2014-0415

Dehghani, M., & Tumer, M. (2015). A research on effectiveness of Facebook advertising on enhancing purchase intention of consumers. *Computers in Human Behavior, 49*, 597-600. doi:10.1016/j.chb.2015.03.051

Gottlieb, U. R., Brown, M. R., & Drennan, J. (2011). The influence of service quality and Trade Show Effectiveness on post-show purchase intention. *European Journal of Marketing, 45*(11/12), 1642-1659. doi:10.1108/03090561111167324

Duffett, R. G. (2015). Facebook advertising's influence on intention-to-purchase and purchase amongst millennials. *Internet Research, 25*(4), 498-526. doi:10.1108/intr-01-2014-0020

Hsu, C., Yu, L., & Chang, K. (2017). Exploring the effects of online customer reviews, regulatory focus, and product type on purchase intention: Perceived justice as a moderator. *Computers in Human Behavior, 69*, 335-346. doi:10.1016/j.chb.2016.12.056

Tiwary, N. K., Kumar, R. K., Sarraf, S., Kumar, P., & Rana, N. P. (2021). Impact assessment of social media usage in B2B marketing: A review of the literature and a way forward. *Journal of Business Research, 131*, 121-139. doi:10.1016/j.jbusres.2021.03.028

Simon, F., & Tossan, V. (2018). Does brand-consumer social sharing matter? A relational framework of customer engagement to brand-hosted Social Media. *Journal of Business Research, 85*, 175-184. doi:10.1016/j.jbusres.2017.12.050

Chang, Y., Yu, H., & Lu, H. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research, 68*(4), 777-782. doi:10.1016/j.jbusres.2014.11.027

Scholz, M., Dorner, V., Franz, M., & Hinz, O. (2015). Measuring consumers' willingness to pay with utility-based recommendation systems. *Decision Support Systems, 72*, 60-71. doi:10.1016/j.dss.2015.02.006

Duan, J., & Dholakia, N. (2015). The reshaping of Chinese consumer values in the social media era. *Qualitative Market Research: An International Journal, 18*(4), 409-426. doi:10.1108/qmr-07-2014-0058

Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in Social Media. *Information & Management, 54*(1), 25-37. doi:10.1016/j.im.2016.03.004