



HAL
open science

Exploitation of Social Network Data for Forecasting Garment Sales

Chandadevi Giri, Sebastien Thomassey, Xianyi Zeng

► **To cite this version:**

Chandadevi Giri, Sebastien Thomassey, Xianyi Zeng. Exploitation of Social Network Data for Forecasting Garment Sales. International Journal of Computational Intelligence Systems, 2019, International Journal of Computational Intelligence Systems, 12 (2), pp.1423-1435. 10.2991/ij-cis.d.191109.001 . hal-04509164

HAL Id: hal-04509164

<https://hal.univ-lille.fr/hal-04509164v1>

Submitted on 18 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Special Issue

Exploitation of Social Network Data for Forecasting Garment Sales

Chandadevi Giri^{1,2,3,4,*}, Sebastien Thomassey¹, Xianyi Zeng¹¹Laboratoire de Génie et Matériaux Textiles (GEMTEX), ENSAIT, F-59000 Lille, France²The Swedish School of Textiles, University of Borås, S-50190 Borås, Sweden³College of Textile and Clothing Engineering, Soochow University, Suzhou 215168, China⁴Automatique, Génie informatique, Traitement du Signal et des Images, Université Lille Nord de France, F-59000 Lille, France

ARTICLE INFO

Article History

Received 24 Apr 2019

Accepted 26 Sep 2019

Keywords

Social Media Data

Forecasting

Naïve Bayes

Sentiment analysis

Fuzzy forecasting model

ABSTRACT

Growing use of social media such as Twitter, Instagram, Facebook, etc., by consumers leads to the vast repository of consumer generated data. Collecting and exploiting these data has been a great challenge for clothing industry. This paper aims to study the impact of Twitter on garment sales. In this direction, we have collected tweets and sales data for one of the popular apparel brands for 6 months from April 2018 – September 2018. Lexicon Approach was used to classify Tweets by sentence using Naïve Bayes model applying enhanced version of Lexicon dictionary. Sentiments were extracted from consumer tweets, which was used to map the uncertainty in forecasting model. The results from this study indicate that there is a correlation between the apparel sales and consumer tweets for an apparel brand. “Social Media Based Forecasting (SMBF)” is designed which is a fuzzy time series forecasting model to forecast sales using historical sales data and social media data. SMBF was evaluated and its performance was compared with Exponential Forecasting (EF) model. SMBF model outperforms the EF model. The result from this study demonstrated that social media data helps to improve the forecasting of garment sales and this model could be easily integrated to any time series forecasting model.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Due to the advent of social media, people, in general, have, like never before, been able to express and share their opinions about their various shopping experiences. In a consumer and market research framework, knowing about consumer experiences is crucial for the companies to improve their products and services they provide. Companies are now in a fierce competition to collect information about the experiences and opinions of their customers from social media as this information provides them with the valuable insights, which can be used to improve the satisfaction of their customers. Traditionally, companies relied on sales forecasting models to estimate their future product sales, which in turn help them in the strategic and effective management of their business operations. However, embedding consumer opinions into sales forecast models and studying their influence on the sales is increasingly becoming a new research trend [1]. There is a shift from the study of effects of marketing and advertisement on product sales to the study of social media influence. Social media is considered to be the digital word-of-mouth platform, which significantly influences consumers shopping behavior and thus, the product sales [2]. Many studies have explored the relationship between consumer sentiments on social media and the product sales [3]. Fashion industry is

grappling with numerous challenges of understanding their customers' opinions and adjusting their business models accordingly [4]. The recent trend of fast fashion is evident to the fact that consumers are rapidly evolving in terms of their changing preferences for new products, designs and features, and also the short product life cycle. Many industries such as television, telecommunication, banks, etc., tap social media to extract customer opinions, and adjust their operations based on the insights derived from social media [5]. However, there is a dearth of research on models in the fashion industry that study the factors such as social media and its effects on product sales. Fashion industry faces operational challenges emanating from the uncertainty in product demand and consumer choices. In this context, traditional forecast models that take into account the effects of marketing campaigns and advertisements on product sales fail to capture the effects of social media. Given the need for real time sales forecast in order to be able to manage various logistics related problems such as inventory, fashion industry needs to consider studying the effects of social media on the sales.

Motivated by the aforementioned problem facing garment industry, this research, aims to forecast garment sales by exploiting twitter data and investigating its influence of twitter as a social media on garment sales. Focusing into line, this study presents a hybrid sales forecast model “Social Media Based Forecasting (SMBF)” which is a fuzzy time series forecasting model that maps

* Corresponding author. Email: chandadevi.giri@ensait.fr

historical information of the sales and the “Impact of sentiments from social media” (Fuzzy Sentiment Impact on Sales [FSIS] model) on garment sales. This research uses real twitter data and historical sales data for an Italian fashion brand, and applied developed model to examine its forecast performance. We believe that this paper is the first endeavor to use sales forecast model on social media data in a fashion industry. In this study, we propose a sales forecast model that is combined with sentiment analysis of social media platform, and we investigate if the performance of garment sales forecast can be improved using this model. The results from this study contributes to addressing overarching research question of forecasting fashion product sales by exploiting social media data, and it can provide decision makers in fashion industry with significant practical insights to optimize their business strategies. Moreover, we show that given the high performance of the model presented in this paper, it is possible to enable sustainable garment production as it would prevent the risk of over production of garments.

Detailed discussion of the existing literature is explained in Section 2. Experimental design and methodology is discussed in Section 3. This is followed by experimental results in Section 4, illustrating the classification results, correlation results and the model performance. Sections 5 and 6 explain the results, discussion and conclusion.

2. LITERATURE REVIEW

This section discusses existing literature related to sales forecasting in a fashion industry, in general, and the use of social media data based sales forecasting for various products. Maria E. Nenni, in paper [6], presents an excellent review of different existing sales forecasting methods used in fashion industry, and explores their suitability to forecast demands and sales with different characteristics. It is highlighted that inaccuracy or low performance of forecast models could be the result of unavailability of historical data, short selling and high level of uncertainty related to consumer preferences and satisfaction. Limitations of traditional sales forecast models such as auto-regressive (AR) model, extreme learning machine (ELM) model and artificial neural network (ANN) model are highlighted in the work [7] by studying the performance of newly developed ELM model with adaptive metrics of input and comparing its accuracy with the traditional ones. In the study carried out by [8], more accurate models based on fuzzy logic, data mining and neural network are used to improve the precision of sales forecasting in a clothing industry. The first study of consumer opinions in terms of online ratings given by consumers is studied [9], wherein Bass model is developed to forecast movie sales on box office. This study was the first approach to investigate the relationship between online reviews as a proxy for spread of consumers’ word-of-mouth and its influence on sales forecast. In a study done by [10], a linear regression model is used in combination with sentiment classifier to forecast box office sales, and the study (concluded) concludes that sentiments from Twitter (highly influence) have high influence on movie sales on box office. To forecast fluctuations in Amazon’s daily products sales, natural language processing is used in [11] to extract sentiments from online product reviews, and the relationship between sentiments and the product sales (have) has been studied. In a similar study done by [12], AR sentiment aware model has been applied to forecast rev-

enue on box office by extracting sentiments from online blogs and their influence on sales was found significant. Another domain, such as earthquake forecasting, where twitter data is used to predict damages caused by earthquakes is studied in [13]. In this study, machine learning models such as naïve Bayes (NB) and support vector machine (SVM) are used to classify tweets related to earthquake incidents and further embedded with spatial smoothing and regression models to estimate the loss due to earthquake. Very interesting application of social media analytics is presented in the study by [14], where authors used deep learning methods to forecast disease outbreak from twitter data with high accuracy. Popularity prediction model based on deep learning is developed in [15], where they extracted social media data to predict overall public interest and social trend. Another interesting industry example of social media based prediction is presented in [16], where machine learning models are applied to derive sentiments from news and social media and used to predict stock prices. Study done by [17] presents a graph theory based convolutional neural network model to forecast propaganda in a political system and also to predict election outcomes in South Africa and Kenya by extracting sentiments from tweets. Due to the growing trend of fast fashion, product life cycle is shrinking rapidly, and therefore, traditional long-term sales forecast models are irrelevant and are not very useful. Short-term demand and replenishment forecasting in a fashion industry using deep learning methods based on real historical industry data is conducted in [18]. Moreover, the influence of consumer sentiments, derived from a big data stream of online customer reviews generated in an e-commerce industry, has been studied in [19] by applying efficient methods to visualize short term demand forecast. This study has emphasized that online consumer sentiments are the promising factor to forecast short-term demand at a product level. Consumer sentiments derived from social media platform such as twitter, are studied in [20] to gauge the current fashion market trend and to predict the consumer perception toward brands. Study in [21] provides a comprehensive discussion on the usefulness of social media data for the operational activities of the industries. It is observed that social media has proved to be a significant factor for understanding of (various social outcomes) consumer behavior and it has a high influence on today’s competitive digital market. In this respect, we aim to study the consumer sentiments from their tweets about fashion brand and to explore their influence on garment sales, which is the first study of its kind in a fashion industry domain.

3. EXPERIMENTAL DESIGN AND METHODS

This section presents the schema of all the steps involved in the SMBF model implementation. It is divided into sub-sections briefing different stages involved in the development of model.

Subsequently, research framework (adopted in) this study is explained in Section 3.1. In the following sections, the topics covered are as follows: data collection steps in Section 3.2, data pre-processing in Section 3.3, sentiment extraction from fashion brand tweets in Section 3.4, Exponential Forecast model explanation in Section 3.5, modeling uncertainty of tweet sentiments using fuzzy inference in Section 3.6, stepwise implementation of SMBF model in Section 3.7, and finally, forecasting performance measuring indicator is explained in Section 3.8.

3.1. Research Framework

Research outline is shown in Figure 1, which includes sales prediction exploiting social media data and sales data in the following steps:

- I. Data Collection—To accomplish the task, Twitter and sales data spanning six months were collected from Italian fashion brand.

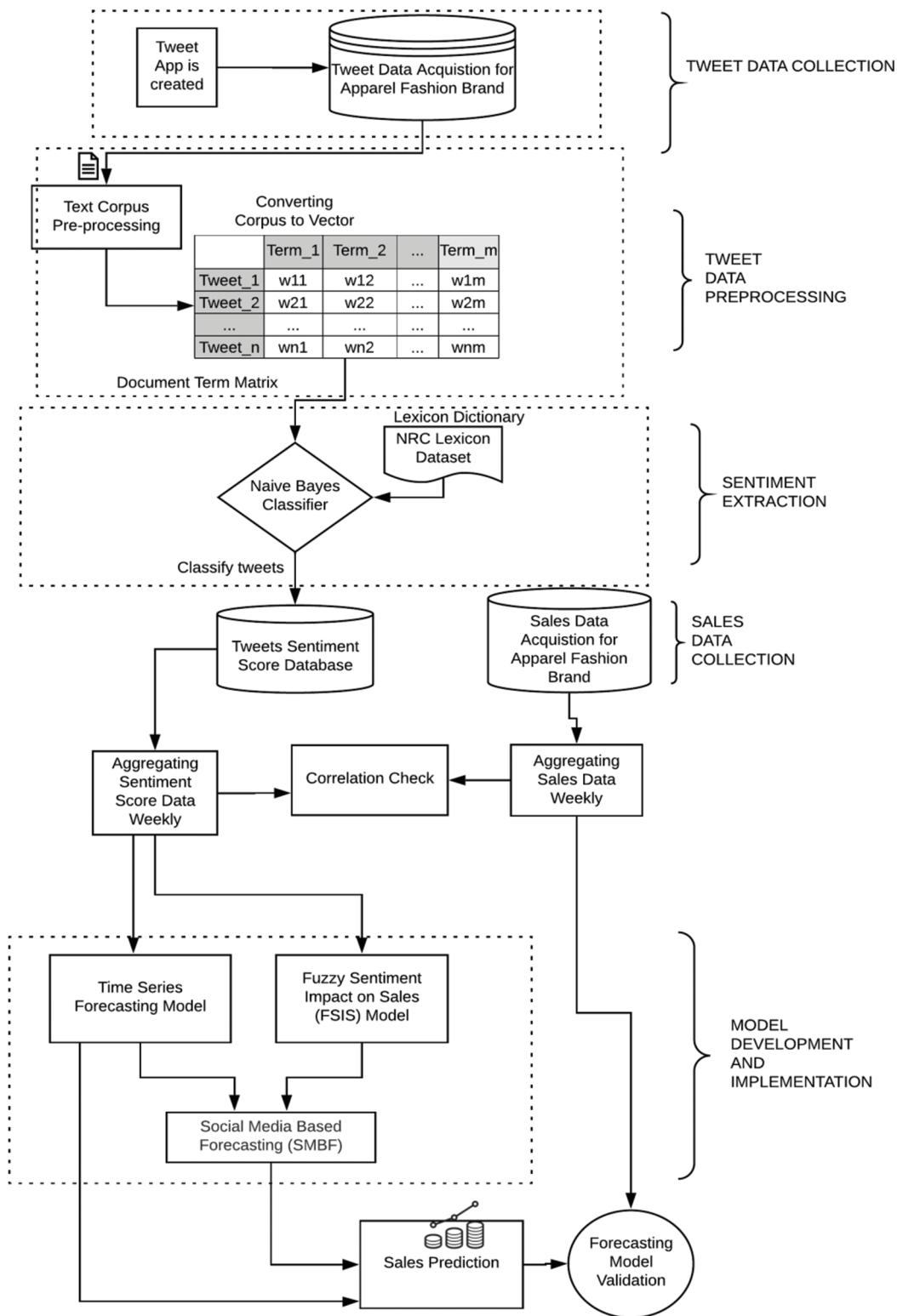


Figure 1 | Research framework.

Table 1 | Attributes of Tweet data.

Serial No.	Attribute	Type	Description
1	Text	String	This is the text of the Tweet.
2	Favorited	Boolean	Specifies if the Tweet has been liked by the authenticated user
3	Favorite_count	Integer	Specifies the count of the Tweet that has been liked by the users
4	ReplyToSN	String	If the represented Tweet is a reply, this field will contain the screen name of the original Tweet's author
5	Created	String	This field states creation of Tweet at UTC time
6	Truncated	Boolean	If the original tweet exceeds the limit of 140 characters, the attribute "text" will be truncated and it will be represented by ellipsis "..."
7	ReplyToSID	String	If the specified Tweet is a reply, this field will contain the string representation of the original Tweet's ID.
8	Id	Integer	It represents a unique identifier of a tweet
9	ReplyToUID	Integer	If the tweet is a reply, this attribute will give the integer representation of the original Tweet's author ID.
10	StatusSource	String	It represents source of the Tweet created via Web, Android, iphone, etc.
11	ScreenName	String	This field gives the "screen name" of the user
12	RetweetCount	Integer	Number of times particular Tweet has been retweeted
13	Place	Places	User's location
14	Retweeted	Boolean	It indicates if the Tweet has been retweeted
15	Longitude/latitude	Coordinates	Indicates user's geographical location

- II. Data Preprocessing and Sentiment extraction—In this step, text mining technique is used for cleaning tweets. These cleaned tweets were then used for sentiment extraction using an enhanced NRC lexicon data and NB classifier. Classification results into three sentiment indicators Positive (S_p), Negative (S_n) and Neutral (S_{ne}).
- III. Examine correlation—Extracted sentiments from earlier stage were aggregated weekly and the relationship between the sentiments of the current week and sales of the upcoming week was examined. This step was carried out to investigate relationship between tweets and sales.
- IV. SMBF Model—After exploring the relationship between tweet sentiments and sales, this model is designed to forecast garment sales based on historical sales and sentiments from social media. To attain this, FSIS model was built to identify the uncertainty of sentiments in terms of corrective coefficient (V_s), which represents the sales variation factor with actual sales and this is implemented on Exponential Forecast (EF) to predict the final forecast from SMBF model.
- V. Performance Validation—The model performance was validated using forecasting measure, Mean Absolute Performance Error (MAPE).

3.2. Data Collection

Data used in this work was collected for one of the Fashion apparel brands, spanning time period of six months from April 2018 and September 2018. This study uses two data forms, first, Twitter data, which were collected using tweet API by creating tweet app to get authentication key [22]. Open source tool "Rstudio" and its package "TwitterR" were used to request API to Twitter using OAuth access [20]. Semantic keywords with hashtag "Brand name" were applied on tweet retrieval string. Collected tweets attributes is shown in the Table 1 and total number of tweets collected for six months interval was 838 as shown in Table 2.

Table 2 | Collected Tweets.

Month	Collected Tweets
April	142
May	166
June	97
July	152
Aug	143
Sep	138
Total	838

In the next step, sales data for the same interval of time as the Twitter data was collected, which accounts for the daily sales of garments.

3.3. Twitter Data Preprocessing

Twitter data collected in previous step was in an unstructured form as it contained many duplicates, urls, punctuation, etc. It was necessary to clean and normalize tweet for analysis. As the first step of preprocessing, manual screening of tweets were performed to confirm whether or not it belongs to the fashion apparel brand that we targeted. While exploring the tweets, synonyms of selected brand name were found in some tweets. Moreover, some of the tweets were also related to politics, games and some other issues, and therefore such tweets were removed. Secondly, as only text is required for sentiment analysis, first attribute shown in the Table 1 was chosen. Text mining was applied to remove duplicates, whitespaces, hashtags, urls, stop words, retweet entities and numbers [20]. As a result, the total count of tweets reduced to 313, which is 37.35 % of the actual collected data.

3.4. Algorithm Formalization for Sentiment Extraction (Or Mining?) from Tweets

This section discusses the methodology applied for the classification of tweets. Lexical approach was used to extract sentiments from

tweets [23]. NRC lexicon dictionary [24] was used, which contains 5636 words labeled with “Positive” and “Negative.” For the tweet classification task, a probabilistic classifier “NB” algorithm [25] was used as it is the effective classification model for text classification. Moreover, it is highly scalable and it allocates the class labels of a finite set to feature vectors [23]. The algorithmic representation of NB is shown in Eq. (1).

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (1)$$

where,

- $P(X|Y)$ is the posterior probability of a target class given attribute.
- $P(Y|X)$ is the likelihood, a probability of attribute given class.
- $P(X)$ is the prior probability of class.
- $P(Y)$ is the prior probability of predictor.

For the mathematical formulation of this research task Twitter dataset T is considered, which is a set of tweets t and can be represented as $T = \{t_1, t_2, t_3 \dots t_n\}$. It was converted into corpus using text mining as explained in the Section 3.3. This corpus is a structured form of text, and is applicable for mathematical operations. Hence, each tweet could be considered as a set of words, i.e., $t = \{w_1, w_2, w_3 \dots w_n\}$. Similarly, Lexical labeled dictionary or dataset L [24] is composed of two attributes: words (W') and sentiment class (C), and it is represented as $L = \{W', C\}$, where, $W' = \{w'_1, w'_2, \dots w'_n\}$ is the set of words in lexicon dictionary and $C = \{c_1, c_2, c_3 \dots c_n\}$ is labeled sentiment class in lexicon dictionary. The sample of the lexicon dataset is shown in Table 3. Lexicon dictionary is labeled with two classes such as “Positive” and “Negative.”

For each given tweet t in Twitter dataset T , NB computes posterior probability of classes of sentiments, $c \in C$, where $C = \{positive, negative\}$ and assigns the predicted class \hat{C} , which has the maximum posterior probability to the word of a given tweet. Accordingly, NB algorithm is represented in Eq. (1) can be rewritten as shown in Eq. (2).

$$\hat{C} = \underset{c \in C}{\operatorname{argmax}} P(c|t) = \underset{c \in C}{\operatorname{argmax}} \frac{P(t|c)P(c)}{P(t)} \quad (2)$$

Further, from Eq. (2), denominator is dropped as $P(t)$ will remain same for all sentiment class and it will have an impact on “argmax” [26]. Therefore, it can be transformed as shown in Eq. (3).

$$\begin{aligned} \hat{C} &= \underset{c \in C}{\operatorname{argmax}} P(t|c)P \\ &= \underset{c \in C}{\operatorname{argmax}} P(w_1, w_2, w_3 \dots w_n|c)P(c) \end{aligned} \quad (3)$$

Table 3 | Example of the lexicon dictionary dataset.

Words (W')	Sentiment Class (C)
Good	Positive
Sad	Negative
Honest	Positive

where,

$$P(w_1, w_2, w_3 \dots w_n|c) = P(w_1|c) \times P(w_2|c) \dots \times P(w_n|c)$$

Therefore, Eq. (3) can be rewritten as below

$$\hat{C} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{w \in t} P(w|c) \quad (4)$$

This model follows the assumptions: position of the features (words) are not taken into account and the feature probabilities $P(w_n|C_n)$ are independent of the given class $c \in C$ [26]. “Prior” and “Likelihood” probability values are calculated for the occurrence of the most probable class and score is assigned.

Furthermore, formalization of the above equations was implemented for the ease of applying it in programming language illustrated in the Figure 2. It is assumed that the algorithm will calculate the maximum posterior probability for the occurrence of a word in a given tweet that is present in a lexical dataset. Therefore, $P(c)$ is calculated, which is a prior probability of sentiment class and it can be calculated using the below Eq. (5) and likelihood using Eq. (6).

$$\operatorname{prior}[c] = P(c) = \frac{Nc}{N} \quad (5)$$

where,

N = Total words in lexicon dictionary

Nc = counts of words from lexicon dictionary for each class c

$$\begin{aligned} \operatorname{Likelihood} &= \prod_{w \in t} P(w|c) = P(w_1, w_2, w_3 \dots w_n|c) \\ &= \frac{\operatorname{count}(w, c)}{\sum_{w' \in W'} \operatorname{count}(w', c)} \end{aligned} \quad (6)$$

where,

(w, c) = number of occurrences of w in a Tweet t

(w', c) = occurrence of w' in lexicon dictionary L labelled with class c

If the tweet does not have the identical word that exists in the lexical dictionary, then the likelihood in Eq. (6) will be zero as shown in Eq. (7) and this cannot be conditioned, therefore, Laplace smoothing [26] is introduced by adding 1 illustrated in Eq. (8).

$$\prod_{w \in t} P(w|c) = \frac{\operatorname{count}(w, c)}{\sum_{w' \in W'} \operatorname{count}(w', c)} = 0 \quad (7)$$

$$\prod_{w \in t} P(w|c) = \frac{\operatorname{count}(w, c) + 1}{\sum_{w' \in W'} \operatorname{count}(w', c) + 1} \quad (8)$$

While applying this equation in programming, a problematic condition could ascend when count of terms increases. As we know that each probability value of term $P(w|c)$ will vary between 0 and 1, and if we multiply them, the product result will start approaching towards zero value. This leads to the problem in the programming language where representation of extremely small number is given as “double” or “long double” type that relies on standard data type “floating,” which is eventually fixed number and its precision reduces with miniscule numbers. The common way to combat this problem is to write the probabilities in “log-probabilities.” This transformation is considered due to number of reasons [27];

- The value of probability lies between 0 and 1, whereas log-probability can lie between $-\infty$ and 0.
- The log-probabilities are easily comparable between smaller and larger numbers, that argmax does in probabilities for example $m < n$, then $\ln(m) < \ln(n)$.
- Arithmetic works well, i.e., logarithm of the multiplication of the variable is the addition of the logarithms, i.e.,

$$\ln(m.n) = \ln(m) + \ln(n)$$

Logarithmic representation of the Eq. (4) is represented below comprehensively with the final Eq. (9).

$$\hat{C} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{w \in t} P(w|c)$$

$$\hat{C} = \ln \left[\prod_{w \in t} P(w|c) \right] + \ln P(c)$$

$$\hat{C} = \left[\sum_{w=1}^{w=n} \ln P(w|c) \right] + \ln P(c)$$

$$\hat{C} = \ln \left[\frac{\operatorname{count}(w, c) + 1}{\sum_{w' \in W} \operatorname{count}(w', c) + 1} \right] + \left[\frac{Nc}{N} \right] \quad (9)$$

Algorithm presented in Figure 2 assigns score to a word according to its frequency that matches with the words in a lexical dataset. The final score is the absolute of addition of logarithmic calculation of prior and likelihood as shown in Eq. (10). So, the algorithm assigns the class score for all words in a tweet. And, we are interested to know if the given tweet, which is as a whole sentence or statement,

is whether positive or negative. Therefore, a new metrics “Best fit” is used for categorizing the tweets as “Positive” or “Negative.”

$$\operatorname{score}[c] = \operatorname{abs}(\operatorname{prior}[c] + \operatorname{likelihood}[w, c]) \quad (10)$$

“Best fit” metrics for a given tweet is calculated as the ratio of the scores of positive and negative class as illustrated in Eq. (11). Classes for a given tweet will be assigned based of the conditions of score shown in Table 4. That is, if the value of “Best fit” score is greater than 1, then the tweet is positive; if it is less than 1 then it will be categorized as “Negative” tweet; and if the score is equal to 1, then it will be categorized as “Neutral” tweet.

$$\operatorname{Best_Fit} = \frac{\operatorname{Total\ score\ of\ Positive\ words\ in\ a\ tweet}}{\operatorname{Total\ Score\ of\ Negative\ words\ in\ a\ tweet}} \quad (11)$$

3.5. Forecasting Model

This research study uses Exponential forecasting (EF) as a base model for forecasting the sales quantity. Exponential smoothing is a well-known forecasting method and was introduced by [28] and [29]. It can be used as an alternative to group of ARIMA models [30]. EF model equation is shown in the Eq. (12):

$$F_{t+1} = \alpha A_t + (1 - \alpha) F_t \quad (12)$$

where,

F_{t+1} = Forecast for the time $t + 1$

Table 4 | Best fit criteria.

Best Fit	Tweet
Greater than 1	Positive
Less than 1	Negative
Equal to 1	Neutral

PSEUDOCODE:

INPUT :

1. $T = \{t_1, t_2, t_3, \dots, t_n\}$ #Tweet Cleaned Dataset where, $t = \{w_1, w_2, w_3 \dots \dots w_n\}$, set of words in a tweet

2. $L = \{W', C\}$ # Classes of Sentiments, where, $C = \{c_1, c_2, c_3 \dots \dots c_n\}$ and $W' = w'_1, w'_2, \dots w'_n$

PROCESS:

#Calculate

$$1. \operatorname{prior}[c] = \ln \frac{Nc}{N}$$

N, Nc = Total words in dataset and counts of words from Dataset in each class c (NRC dataset)

$$2. \operatorname{likelihood}[w, c] = \ln \left[\frac{\operatorname{count}(w, c) + 1}{\sum_{w' \in W} \operatorname{count}(w', c) + 1} \right]$$

$$3. \operatorname{score}[c] = \operatorname{abs}(\operatorname{prior}[c] + \operatorname{likelihood}[w, c])$$

For each word w in L (L = Lexicon of dataset D), $\operatorname{count}(w, c)$ = number of occurrences of w in Tweets and $\operatorname{count}(w', c)$ = number of occurrences of w in L .

OUTPUT : Predicted Class c for the words w in a tweet t with a score

Figure 2 | Algorithm for assigning score to the word class.

α = Smoothing constant $0 \leq \alpha \leq 1$

A_t = Actual data

F_t = Forecast data for time t

Sales data of fashion garment industry was aggregated weekly and an EF was performed. “Corrective Coefficient (V_s)” was calculated, which represents the variation in sales as shown in Eq. (13). V_s is used as an output variable for FSIS, mapping the impact of tweets optimizing the predicting ability of the forecast model.

$$\text{Corrective Coefficient } (V_s) = \frac{\text{Actual Sales} - \text{EF Sales Forecast}}{\text{EF Sales Forecast}} \quad (13)$$

3.6. Modelling Tweet’s Sentiment Uncertainty for Predicting Sales Using Fuzzy Theory

This section discusses, in detail, the process involved in modelling tweet sentiments. This section is divided into two sub-sections briefing the details of model design and defines linguistic variables for analyzing the impact of Tweet sentiments. Section 3.6.1 outlines the architecture of FSIS model and section 3.6.2 explains the fuzzy linguistic variables for the FSIS model.

The uncertainty is represented by fuzzy sets in classical theory and each element in a set hold a membership degree [31–33]. Fuzzy set can be denoted as a pair (U, m) , where U is a set and m is a membership function of a fuzzy set, which assigns the value between 0 and 1 and it can be denoted as $U \rightarrow [0, 1]$ [34]. Therefore, for a fuzzy set $P = (U, m)$, membership function $m = \mu(A)$ maps the function between 0 to 1. For a finite set $U = \{a_1, \dots, a_n\}$, it can be represented as $\{m(a_1)/a_1, \dots, m(a)/a_n\}$. Let $a \in U$, then α will not be considered in the fuzzy set (U, m) if $m(a) = 0$, will be fully considered if $m(a) = 1$ and it will be partly considered if $0 < m(\alpha) < 1$ [35]. It follows the mathematical operations of complement, intersection and union. For the fuzzy set P and Q, $P, Q \subseteq U, u \in U$, complement operation can be represented as $\mu_P(u) = 1 - \mu_P(u)$, intersection operation can be represented as $\mu_{P \cap Q}(u) = \min\{\mu_P(u), \mu_Q(u)\}$ and Union of two fuzzy sets can be shown as $\mu_{P \cup Q}(u) = \max\{\mu_P(u), \mu_Q(u)\}$. Fuzzy set has “Commutative,” “Associative,” “Distributive,” “Identity” and “Transitive” properties. Triangular membership [36] is popular to solve engineering optimization problems due to its ability to provide solution instantly. Degree of Membership function for a fuzzy set P can be illustrated in Fig. 3 and Eq. (14) defined by the lower and upper limit, a and b where, $a < b < c$.

$$\mu_{P(x)} = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ \frac{c-x}{c-b} & b < x < c \\ 0 & x \geq c \end{cases} \quad (14)$$

3.6.1. FSIS architecture

FSIS architecture is defined using classical “Mamdani” Fuzzy inference proposed by [32,37], the first control system base, fashioning it by if-then rules, which can be gained from human knowledge. This work was inspired by [31] and realized its application of fuzzy inference for complex decision mechanism. Mamdani inference maps the input attributes to the output using fuzzy logic inference which involves steps as follows;

- (a) First step involves defining the input and output linguistic variables followed by defining membership function for each linguistic variables
- (b) Fuzzification—In this step, input parameters are mapped to appropriate linguistic variables based on pre-defined membership function that can be viewed as fuzzy set. Membership functions are linked with the weight factors that regulate the influence for each defined rule.
- (c) Knowledge base—It has a database of fuzzy set and defined rules. “Facts” are symbolized through linguistic variables and reasoning is followed by defined if-then rules.
- (d) Inference Engine—Rule inference is managed by inference engine, where human knowledge can be combined without any difficulty using linguistic rules.
- (e) Defuzzification—Aggregation of all fuzzy sets to give the single output to draw a conclusion. Centroid method is used for obtaining the crisp number as shown in Eq. (15).

$$x^* = \frac{\int \mu_P(x) .xdx}{\int \mu_P(x) .dx} \quad (15)$$

Model design of FSIS involves above steps and this is explained in Section 3.6.2 with input parameters as sentiments and output parameter as corrective coefficient as explained in the Section 3.5, and corresponding defined membership functions.

3.6.2. Defining universal and fuzzy set for sentiments for FSIS

For formalizing and modeling twitter data with garment sales performance, we used intelligent fuzzy technique as discussed in the Section 3.6.1. FSIS constitutes three fuzzy sentiment input variables I as S_p (Positive), S_n (Negative), S_{ne} (Neutral) and Output O

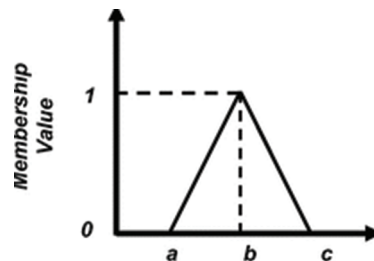


Figure 3 | Triangular membership.

as V_s (Corrective coefficient). Triangular membership function is assigned to input variables in the range between [0 1] and output variable in the range [-25 25]. Aggregated weekly sentiment data were used for all three input variables to feed into model. The key objective in modelling sentiments, viz., S_p (Positive); S_n (Negative); and S_{ne} (Neutral) retrieved from tweets with fuzzy inference is to capture human emotions on social media weekly and to forecast the sales variation based on these consumer emotions. Human emotions are uncertain and they change from time to time. So, for example, maybe in one week, sentiment "Positive" could be high and it could lead to a positive effect on sales of upcoming week. Similarly, if it is "Negative," then it could lead to a negative impact on sales of upcoming week. The designs of the linguistic variables for sentiments are taken as "Low," "Medium" and "High." Thus, I is a set of sentiments, i.e., $Input(I) = \{S_p, S_n, S_{ne}\}$ and membership sets for I is defined in three linguistic variables as "Low," "Medium" and "High" for each input variables and can be written as $S_p = \{Low, Medium, High\}$, $S_n = \{Low, Medium, High\}$, $S_{ne} = \{Low, Medium, High\}$, and for output variable $Output(O) = \{V_s\}$, membership sets is defined as $V_s = \{Low, Medium, High\}$, where, V_s is a corrective coefficient and it is calculated as a difference between actual sales and EF sales as shown in Eq. (13).

Triangular membership function for the input variable in set I is defined between limit [0 1] and the output variable in the limit [-25 25], which is a corrective coefficient in percentage of sales and it will predict sales variation from -25% to 25%. (Table 5 describes) the membership variables and values for the I and O respectively are shown in Table 5. Figures 4 and 5 depict the triangular membership function for variables I and O.

Calculation of the membership function value for the Input variable " $S_p = \{Low, Medium, High\}$ " is illustrated below:

$$\mu(S_p(Low)) = \frac{0.5 - x}{0.5} \quad 0 \leq x \leq 0.5$$

Table 5 | Membership function for input variable I (Input) and O (Output).

Input/ Membership	Low	Medium	High
$(S_p)_w$	[0 0 0.5]	[0 0.5 1]	[0.5 1 1]
$(S_n)_w$	[0 0 0.5]	[0 0.5 1]	[0.5 1 1]
$(S_{ne})_w$	[0 0 0.5]	[0 0.5 1]	[0.5 1 1]

Output/ Membership	Low	Medium	High
$(V_s)_{w+1}$	[-25 -25 0]	[-25 0 25]	[0 25 25]

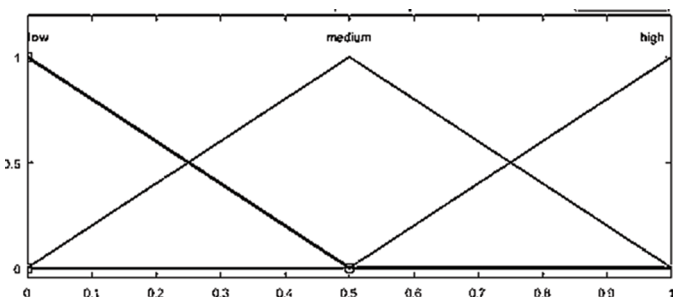


Figure 4 | Triangular membership for input variables (S_p, S_n, S_{ne}).

$$\mu(S_p(Medium)) = \frac{x}{0.5} \quad 0 \leq x \leq 0.5 \quad \text{and} \quad \frac{1-x}{0.5} \quad 0.5 \leq x \leq 1$$

$$\mu(S_p(High)) = \frac{x-0.5}{0.5} \quad 0.5 \leq x \leq 1$$

Similarly, membership function values is calculated for S_n and S_{ne} . Besides, the membership function value for corrective coefficient " $V_s = \{Low, Medium, High\}$ " is calculated as below:

$$\mu(V_s(Low)) = \frac{0 - x}{0 - (-25)} \quad -25 \leq x \leq 0$$

$$\mu(S_p(Medium)) = \frac{x - (-25)}{0 - (-25)} \quad -25 \leq x \leq 0 \quad \text{and}$$

$$\frac{25 - x}{25 - 0} \quad 0 \leq x \leq 25$$

$$\mu(S_p(High)) = \frac{x - 25}{25 - 0} \quad 0 \leq x \leq 25$$

Considering the impacts of the weekly sentiments, fourteen conditional rules are defined by taking into account human knowledge as shown in Table 5. These rules are defined as

"Rule i : IF condition i THEN action I"

To demonstrate the functioning mechanism of rules consider first row of the Table 6, which states that if $S_p = High$, $S_n = low$ and $S_{ne} = Medium$, then $V_s = High$, i.e., if the aggregated weekly positive sentiment is high, negative sentiment is low, and the neutral sentiments is medium, then the sales performance will be high and its values will lie between 0 % to 25 % range, meaning there is an incremental variation in sales than the previous week. Similarly, expected sales variation decision will be taken by the model considering the inference of input variables and defined rules. The final score of the output, i.e., defuzzification of the fuzzy set and the crisp number is calculated by the centroid method using Eq. (15). Architecture of FSIS is depicted in Figure 6.

3.7. SMBF Model

SMBF model is a time series model, which takes into account the impact of sentiments on sales using FSIS. Parameters of the FSIS are optimized to improve the forecast ability of the model and it is implemented on the exponential time series forecasting and the formula its forecast estimate is given in Eq. (16), where V_s is corrective coefficient, calculated using Eq. (13) and EF is sales volume forecasted by the EF model as explained in section 3.5.

$$SMBF = (1 + V_s) \times EF \tag{16}$$

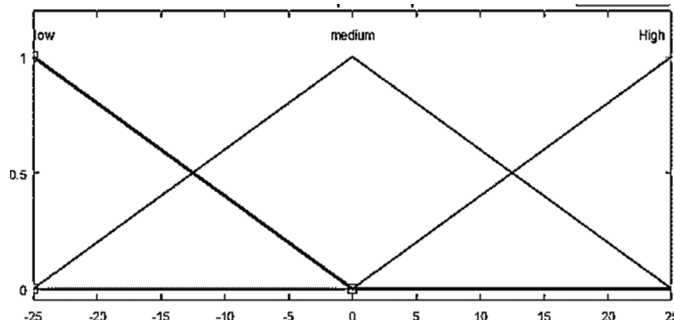


Figure 5 | Triangular membership for output variables (V_s).

where,

SMBF = sales Forecast from fuzzy time series model

V_s = corrective coefficient predicted by the FSIS model

EF = sales forecasted by Exponential Model

FSIS model parameters, i.e., “Input,” “Output” and “Rules” defined in Section 3.6.2 were optimized using the local optimization

Table 6 Rules defined for Fuzzy Sentiment Impact on Sales (FSIS).

if	And	And	Then
$(S_p)_w$	$(S_n)_w$	$(S_{ne})_w$	$(V_s)_{w+1}$
High	Low	Medium	High
High	Medium	Low	High
Low	High	Medium	Low
Medium	High	Low	Medium
Low	High	Low	Low
Medium	Low	High	Medium
Low	Medium	High	Medium
Low	Low	Medium	High
Medium	Low	Medium	High
Low	Medium	Medium	Low
Medium	Medium	Low	Medium
Medium	Low	Low	Medium
Low	Low	High	Medium
High	Low	Low	High

method known as “pattern search method” [38], which is useful for the fast convergence. The model was tuned for 500 iterations, and as a result, model was optimized. Optimized parameters for “Input” and “Output” variables are shown in Table 7 and optimized rules are shown in the Table 8.

3.8. Validation Method and Performance Measure (MAPE)

Mean Absolute Percentage Error (MAPE) is a performance metrics for evaluating the performance of the forecasting models and it is denoted as in Eq. (17).

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \tag{17}$$

where,

A_t = actual sales at time t

F_t = forecasted sales at time t

Table 7 | Optimized model parameter of FSIS.

Input/ Membership	Low	Medium	High
$(S_p)_w$	[0 0 0.5]	[0 0.875 1]	[0.75 1 1]
$(S_n)_w$	[0 0 0.5]	[0 1 1]	[0.125 0.75 1]
$(S_{ne})_w$	[0 0 0.5]	[0.125 0.375 1]	[1 1 1]
Output/ Membership	Low	Medium	High
$(V_s)_{w+1}$	[-25 -23.5 20.5]	[-15 -14.375 -14.25]	[22.5 25 25]

FSIS, Fuzzy Sentiment Impact on Sales.

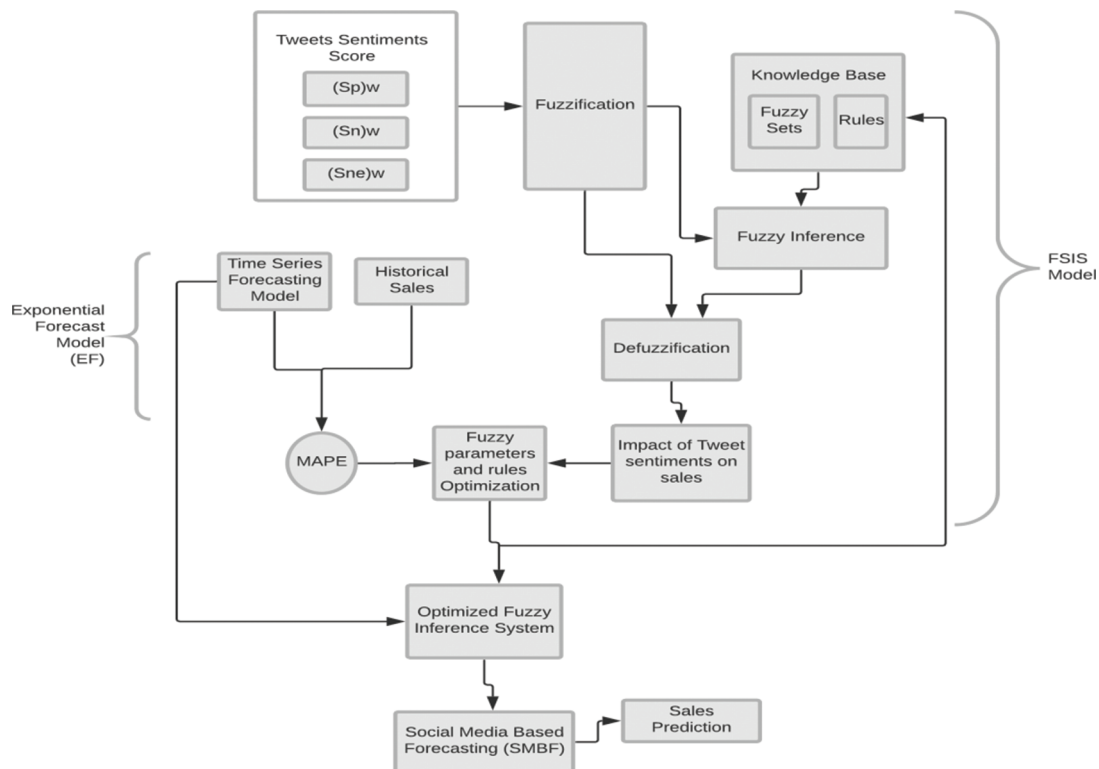


Figure 6 | Social Media Based Forecasting (SMBF) model.

Table 8 | Optimized rules defined for FSIS.

$(S_p)_w$	$(S_n)_w$	$(S_{ne})_w$	$(V_s)_{w+1}$
High	Medium	Medium	High
High	Medium	High	High
High	High	-	High
-	High	Medium	Low
High	High	High	Low
Medium	Low	High	Medium
High	High	Medium	Medium
High	Low	Medium	High
Medium	Low	Medium	High
Low	-	Low	Low
-	Medium	Low	Medium
High	High	Low	Medium
Low	Low	High	Medium
Low	High	Low	High

FSIS, Fuzzy Sentiment Impact on Sales.

4. EXPERIMENTAL RESULTS

This section presents the results of all the models discussed in previous section. This section is divided into three subsections: Section 4.1 summarizes the result of NB classifier, which is described in Section 3.4. It is used for extracting sentiments from the tweets and assigning sentiment score to a word in a tweet that matches with the lexical dictionary. Based on sentiment score, tweets are classified as “Positive,” “Negative” and “Neutral” as explained in Section 3.4. Further, in Section 4.2, computational mechanism of correlation between weekly aggregated Twitter sentiments for the current week and sales of the next week is discussed. Lastly, Section 4.3 evaluates the forecast performance of SMBF and EF models.

4.1. Tweet Classification

Tweets were assigned sentiment class as explained in Section 3.4. Logarithmic values of “Prior” and “Likelihood” were calculated of the lexicon dataset [24] and it is shown in the Table 9. The total count of the words labeled “Positive” is 2312 and “Negative” is 3324 in lexical dataset. (To see in detail) Detailed mechanism (of how the) of algorithm is explained in section 3.4. (Let us take one) Consider, for example, in a tweet, “style made genuine matching lining,” the word “genuine” exists in the lexical database as “Positive,” therefore, the score for this positive word in a tweet will be calculated using Eq. (10). As there is only one positive word that matches with the words in lexical dictionary and not with the negative word in a tweet, the score for the “Positive” class will be calculated by using logarithmic value in Table 9 as “ $0.891062 + (1 \times 7.74586823) = 8.636929$ ”; and for the “Negative” class, it is “ $0.528006 + (0 \times 8.1088924156) = 0.528006$,” and the final score for the best fit will be calculated using Eq. (11) and the result is shown in Table 10. In Table 10, “5” indicates the tweet number in a Tweeter data T . In this case, as the score is more than 1, the assigned class for this tweet is “Positive.” Similarly, all tweets were assigned scores and classified as “Positive,” “Negative” and “Neutral” based on their “Best Fit” score. Overall classification of cleaned tweets is illustrated in the Figure 7, and it can be seen that about 45% tweets were classified as “Positive”; 27% tweets were classified as “Negative” and approximately 28% of tweets were classified as “Neutral.” The classification results indicate that the chosen fashion brand has quite a positive

Table 9 | Log prior and log likelihood of lexical dictionary.

NRC Lexicon	Count of Words (Nc)	$\log(Nc/N)$	\log likelihood (w, c)
Positive	2312	0.891062	7.74586823
Negative	3324	0.528006	8.1088924156
Total (N)	5636		8.636929873

Table 10 | Result interpretation for a tweet has one positive word.

▲ POS	NEG	POS/NEG	BEST_FIT
5	8.63692987301857	0.528005717043232	16.357644612987
			positive

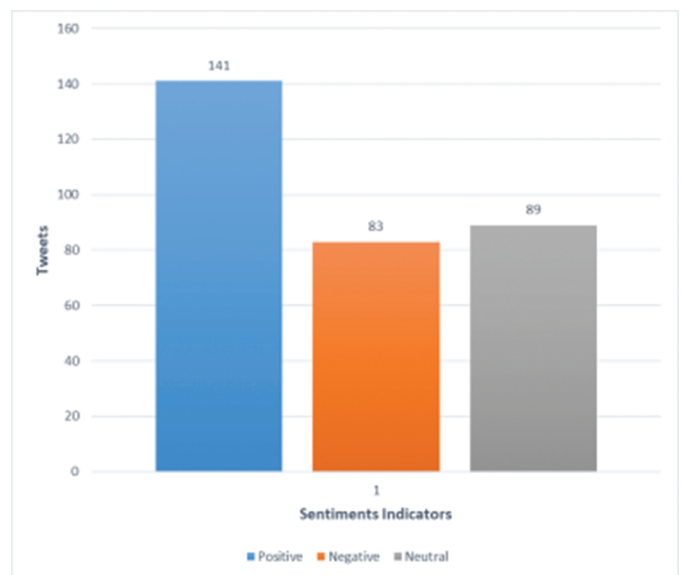


Figure 7 | Classification of tweets for fashion brand.

outlook from customer perspectives as only 27% of total tweets were negative.

4.2. Correlation Test

The main aim of this study was to investigate if the customer sentiments from tweets collected in the current week have any influence on the upcoming week’s sales. For this, assumption was made and it was verified using Pearson correlation [39]. Following assumptions were made to check the correlation between sentiments indicators S_p, S_n, S_{ne} and sales volume S_v , where S_p = aggregated “Positive” tweets for a week “ w ,” S_n = aggregated “Negative” tweets for a week “ w ,” S_{ne} = aggregated Neutral tweets for a week “ w ,” S_v = Change in volume of sales for week $w + 1$.

- If the correlation between S_p and S_v is positive, then the sales will increase.
- If the correlation between the S_n and S_v is negative the sales will be decrease.
- If the correlation between the S_{ne} and S_v is positive the sales will be increase.

Correlation values, as shown in the Table 11, indicate that there is a moderate positive correlation between S_{ne} and S_v that is 32%, weak

positive correlation between S_p and S_v with a value of 22% and negative correlation between S_n and S_v with -13% . This result clearly shows that if there is more number of positive tweets and neutral tweets, it will have positive effect on sales, whereas if there is more number of negative tweets, it will have negative effect on sales.

4.3. SMBF Model Performance

In this research FSIS is combined and implemented on EF as explained in Section 3.7 and final forecast was achieved by SMBF using Eq. (16). Model performance of SMBF was evaluated by calculating MAPE values as explained in Section 3.8. MAPE values for EF model and SMBF model is shown in the Table 12. This clearly shows that adding the impact of the sentiments by FSIS model on EF model, exhibit the forecasting capability of model. It can be observed that the performance of the SMBF is better than EF, which solely relies on the historical sales information, whereas SMBF model counts on both historical sales data and social media data. This result illustrates that by adding the features of social media in our forecasting model, the performance of the model can be improved significantly.

The forecasted sales by EF and SMBF models were plotted to compare it with actual sales as shown in Figure 8, which clearly shows that forecast achieved by

SMBF model is significantly better than EF. The prediction of the SMBF model was very close to the actual as shown in Figure 8. As the collected data was during summer season, there is a peak in sales for week 27 and 34 while there is a slight drop in week 36. This could be attributed to the promotional or sales discount effect on the sales. These factors are not considered in this study. However, model SMBF tried to capture the trend to some extent.

5. DISCUSSION

This study is conducted on real tweets and sales data of fashion brand for spring and summer assortment. When collected tweets were analyzed after cleaning, the number of tweets was reduced to only 37.5% of the actual tweets in the collected data, which indicates that this brand is in establishing phase on social media. Data pre-processing of tweets was done in two ways: manually; and using text mining to ensure that the collected tweets belonged to the chosen brand. As the enhanced version of lexicon dataset (NRC lexicon) was used, NB classifier classified tweets quite satisfactorily. Most of the tweets were classified as “Positive” and “Neutral,” which represents the popularity of the brand amongst its consumers. This research investigated and established the relationship between the social media data and sales data. Although, the relationship between the social media data and sales data was moderate and weak, we can conclude that there is a relationship between tweet data and fashion apparel sales data, and this indicates that tweet data can be used to forecast garment sales. To deal with the fuzziness of human sentiments and fashion apparel sales, FSIS model was designed, which analyzes the impact of sentiments on sales. FSIS model was combined with EF model to form the SMBF model. The model parameters for SMBF model were optimized using pattern search method used for local optimization reducing the errors between the actual and predicted sales of the model. SMBF model performance is illustrated in Figure 8 and Table 12. SMBF outperforms the EF. There are two peaks in sales for week 28 and week 35, which are summer weeks and model has tried to capture the sales peak during summer to some extent. This shows that the garment sales can be influenced by the social media and if

Table 11 | Correlation table.

Variables	Correlation
S_p and S_v	0.22
S_n and S_v	-0.13
S_{ne} and S_v	0.32

Table 12 | Model evaluation.

Forecasting Model	MAPE
EF	14.7911567205165
SMBF	10.1143453396686

EF, Exponential Forecast; SMBF, Social Media Based Forecasting.

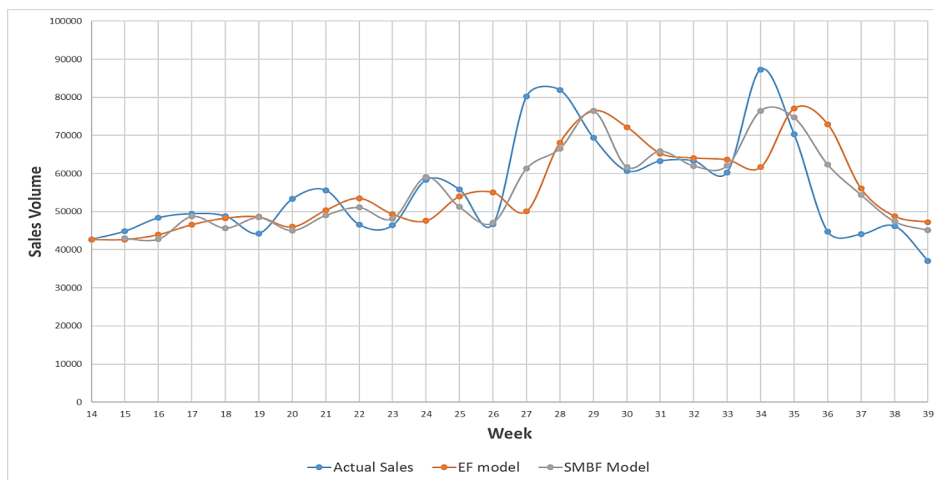


Figure 8 | Forecasted sales by Exponential Forecast (EF) model and Social Media Based Forecasting (SMBF) model with actual sales.

its behavior is modelled in existing time series model, we can considerably improve the performance of the traditional forecasting of garment sales.

6. CONCLUSION AND FUTURE WORK

Taking into account SMBF model performance and results, we conclude that this model could be effective and beneficial to the fashion garment industry. This research proves that social media data can be used for forecasting sales volume. As the forecasting model will be sensitive to social media data, it is the responsibility of a company to increase its visibility on social media in terms of its services and advertisement.

This will improve the quality of collected social media data. The limitation of this study, firstly, lies in the fact that classification of tweets could be improved by adding more words to the dictionary for the class “Positive” and “Negative.” Secondly, as for the presented model, only “Twitter” as a social media data platform was considered, and if we combine other social media platforms such as “Facebook” and “Instagram,” model performance could be enhanced further.

The experiment in this study was focused on weekly forecast, and therefore, overall sales data of a brand was used for analysis. We would like to extend this work in future to create a short term forecasting model to forecast according to the product category. In this experiment, we combined FSIS model with EF model to build SMBF model for forecasting garment sales. In future, we will incorporate other forecasting models to enhance this study further.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGMENTS

This research work is conducted under the framework of SMDTex-Sustainable Management and Design in Textiles. We are grateful to the company, Evo Pricing, for providing us with the data to carry out this research work.

REFERENCES

- [1] Z.P. Fan, Y.J. Che, Z.Y. Chen, Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis, *J. Bus. Res.* 74 (2017), 90–100.
- [2] Q. Ye, R. Law, B. Gu, The impact of online user reviews on hotel room sales, *Int. J. Hosp. Manage.* 28 (2009), 180–182.
- [3] N. Amblee, T. Bui, The impact of electronic-word-of-mouth on digital microproducts: an empirical investigation of Amazon shorts, in 2007 40th Annual Hawaii International Conference on System Sciences (HICSS07), 2007.
- [4] C. Giri, S. Thomassey, X. Zeng, Customer analytics in fashion retail industry, in: A. Majumdar, D. Gupta, S. Gupta (Eds.), *Functional Textiles and Clothing*, Springer, Singapore, 2019, pp. 349–361.

- [5] S. Yu, S. Kak, A Survey of Prediction Using Social Media, 2012. <https://arxiv.org/abs/1203.1647>.
- [6] M.E. Nenni, L. Giustiniano, L. Pirolo, Demand forecasting in the fashion industry: a review, *Int. J. Eng. Bus. Manage.* 5 (2013).
- [7] M. Xia, Y. Zhang, L. Weng, X. Ye, Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs, *Knowl. Based Syst.* 36 (2012), 253–259.
- [8] S. Thomassey, Sales forecasts in clothing industry: the key success factor of the supply chain management, *Int. J. Prod. Econ.* 128 (2010), 470–483.
- [9] V. Setyani, Y.-Q. Zhu, A.N. Hidayanto, P.I. Sandhyaduhita, B. Hsiao, Exploring the psychological mechanisms from personalized advertisements to urge to buy impulsively on social media, *Int. J. Inf. Manage.* 48 (2019), 96–107.
- [10] S. Asur, B.A. Huberman, Predicting the future with social media, in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, 2010, pp. 492–499.
- [11] N. Archak, A. Ghose, P.G. Ipeirotis, Deriving the pricing power of product features by mining consumer reviews, *Manage. Sci.* 57 (2011), 1485–1509.
- [12] Y. Liu, X. Huang, A. An, X. Yu, ARSA: a sentiment-aware model for predicting sales performance using blogs, in SIGIR '07 Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, 2007.
- [13] M. Mendoza, B. Poblete, I. Valderrama, Nowcasting earthquake damages with Twitter, *EPJ Data Sci.* 8 (2019).
- [14] O. Şerban, N. Thapen, B. Maginnis, C. Hankin, V. Foot, Real-time processing of social media with SENTINEL: a syndromic surveillance system incorporating deep learning for health classification, *Inf. Process. Manage.* 56 (2019), 1166–1184.
- [15] G. Chen, Q. Kong, N. Xu, W. Mao, NPP: a neural popularity prediction model for social media content, *Neurocomputing.* 333 (2019), 221–230.
- [16] Z. Wang, S.-B. Ho, Z. Lin, Stock market prediction analysis by incorporating social and news opinion and sentiment, in 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 2018, pp. 1375–1380.
- [17] R. Vijayan, G. Mohler, Forecasting retweet count during elections using graph convolution neural networks, in 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, 2018, pp. 256–262.
- [18] W. Du, S.Y.S. Leung, C.K. Kwong, A multiobjective optimization-based neural network model for short-term replenishment forecasting in fashion industry, *Neurocomputing.* 151 (2015), 342–353.
- [19] E.W.K. See-To, E.W.T. Ngai, Customer reviews for demand distribution and sales nowcasting: a big data approach, *Ann. Oper. Res.* 270 (2018), 415–431.
- [20] C. Giri, N. Harale, S. Thomassey, X. Zeng, Analysis of consumer emotions about fashion brands: an exploratory study, in Conference on Data Science and Knowledge Engineering for Sensing Decision Support, Belfast, 2018, pp. 1567–1574.
- [21] S. Ren, H.-L. Chan, T. Siqin, Demand forecasting in retail operations for fashionable products: methods, practices, and real case study, *Ann. Oper. Res.* (2019).
- [22] J. Gentry, Twitter client for R, n.d. <http://geoffjentry.hexdump.org/twitterR.pdf>.

- [23] S.J. Russell, P. Norvig, E. Davis, *Artificial Intelligence A Modern Approach*, Pearson Education India, 2015.
- [24] M. Saif, NRC Emotion Lexicon. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.
- [25] M.E. Maron, *Automatic indexing: an experimental inquiry*, J. ACM. 8 (1961), 404–417.
- [26] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, 2008.
- [27] S. Russell, P. Norvig, *Artificial Intelligence—A Modern Approach*, third ed., Pearson Education, 2013. <https://www.pearson.com/us/higher-education/program/Russell-Artificial-Intelligence-A-Modern-Approach-3rd-Edition/PGM156683.html>
- [28] R. Brown, *Exponential Smoothing for Predicting Demand*, Arthur D. Little, Cambridge, 1956.
- [29] C.C. Holt, *Forecasting seasonals and trends by exponentially weighted moving averages*, Int. J. Forecast. 20 (2004), 5–10.
- [30] A. Pankratz, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, John Wiley & Sons, New York, 2009.
- [31] L.A. Zadeh, *Outline of a new approach to the analysis of complex systems*, IEEE Trans. Syst. Man Cybern. SMC-3 (1973), 28–44.
- [32] S. Gottwald, *An early approach toward graded identity and graded membership in set theory*, Fuzzy Sets Syst. 161 (2010), 2369–2379.
- [33] L.A. Zadeh, *The concept of a linguistic variable and its application to approximate reasoning—I*, Inf. Sci. (Ny). 8 (1975), 199–249.
- [34] D. Dubois, H. Prade, *Fuzzy Sets and Systems Theory and Applications*, Academic Press, New York, 1980.
- [35] I. Beg, S. Ashraf, *Similarity measures for fuzzy sets*, Appl. Comput. Math. 8 (2009), 192–202.
- [36] W. Pedrycz, *Why triangular membership functions?*, Fuzzy Sets Syst. 64 (1994), 21–30.
- [37] E.H. Mamdani, S. Assilian, *An experiment in linguistic synthesis with a fuzzy logic controller*, Int. J. Man. Mach. Stud. 7 (1975), 1–13.
- [38] C. Audet, J.E. Dennis, *Analysis of generalized pattern searches*, SIAM J. Optim. 13 (2002), 889–903.
- [39] K. Yeager, *LibGuides: SPSS Tutorials: Pearson Correlation*. <https://www.semanticscholar.org/paper/LibGuides%3A-SPSS-Tutorials%3A-Pearson-Correlation-Yeager/7c8f5092aa5a6b963f5e71abb4021344296eecd8#related-papers>.