Social Media Big Data Analytics for Demand Forecasting:

Development and Case Implementation of an Innovative Framework

Rehan Iftikhar, Maynooth University, Maynooth, Ireland

Mohammad Saud Khan, Victoria University of Wellington, New Zealand

ABSTRACT

Social media big data offers insights that can be used to make predictions of products' future demand and add value to the supply chain performance. The paper presents a framework for improvement of demand forecasting in a supply chain using social media data from Twitter and Facebook. The proposed framework uses sentiment, trend, and word analysis results from social media big data in an extended Bass emotion model along with predictive modelling on historical sales data to predict product demand. The forecasting framework is validated through a case study in a retail supply chain. It is concluded that the proposed framework for forecasting has a positive effect on improving accuracy of demand forecasting in a supply chain.

KEYWORDS

Apparel Supply Chain, Bass Emotion Model, Big Data, Demand Forecasting, Emotion Enhanced Model, Sentiment Analysis, Social Media, Supply Chain Management

INTRODUCTION

Big data represents a tremendous opportunity for companies, as it can help to make better decisions in an operational, tactical and strategic level (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012), with direct impact on business profitability (Waller & Fawcett, 2013). The ability to draw insights from different types of data creates huge value for a firm (Dijcks, 2013; Kiron & Shockley, 2015). Big data presents a far greater opportunity than what is being utilized. Only 0.5% of big data is being utilized and analysed while there is potential for so much more (Guess, 2015). Bearing in mind this huge potential, literature providing empirical evidence of the business value added by big data analytics in a supply chain remains little and even poor (Wamba, 2017).

All supply chain operations and activities are set in motion by the final customers' demand (Syntetos et al., 2016). Demand forecasting is used as a basis to make supply chain strategy (Marshall, Dockendorff, & Ibáñez, 2013) and forecasting weaknesses is one of the main reasons for supply chain failures (Zadeh, Sepehri, & Farvaresh, 2014). Demand Forecasting can be improved significantly by using big data (Chao, 2015), especially the big data from social media (Arias, Arratia, & Xuriguera, 2014). With an increase in social media activity, there has been an emergence of academic and industrial research that taps into these social media data sources. However, the utilization of these data sources remain at an early stage and outcomes are often mixed (Yu, Duan, & Cao, 2013).

DOI: 10.4018/JGIM.2020010106

Companies face a challenge in forecasting with regards to analysing their historical data in the same breath as big data from social media (Papanagnou & Matthews-Amune, 2017). There has been an increased focus from supply chain practitioners to leverage effects from unstructured big data such as social media data, but there is very little support in terms of empirical evidence (Syntetos et al., 2016). Integration of social media analytics and supply chain management is needed to comprehensively establish 'what can be actually done' in the field of forecasting with the help of analytics. There is a paucity of predictive frameworks for forecasting using social media big data. This paper aims to bridge the gap between traditional forecasting techniques and big data analytics utilization and contributes towards a forecasting platform using social media big data as well as historical sales data.

This work presents a framework to utilize social media big data in Bass-Emotion Model introduced by Fan, Che, & Chen (2017). The proposed framework uses the results of sentiment analysis on Facebook and Twitter for demand forecasting. This work provides empirical evidence on the usage of social media big data for demand forecasting in supply chain management (Choi, 2018; Schaer, Kourentzes, & Fildes, 2018). It is one of the first studies that incorporates word analysis, topic modelling and sentiment analysis to provide social media data parameters to the Bass- Emotion model.

LITERTATURE REVIEW

Big Data Analytics in Supply Chain Management

Diverse, massive and complex data on different domains of business and technology which cannot be efficiently addressed by the traditional technologies, skills, and infrastructure is referred to as big data. Most big data researchers and practitioners in general agree on three dimensions that characterize big data: volume, velocity and variety (Zikopoulos & Eaton, 2011). Big data analytics in supply chain management can be described as applying analytical techniques on big data to facilitate optimization and decision making in a supply chain (Souza, 2014). The use of big data analytics can help us understand 'what has happened, what is happening at the moment, what will happen and why things happen' (Feki & Wamba, 2016 p.1127). Three distinct analytics approaches for answering these questions have been classified as descriptive, predictive, and prescriptive analytics (Hahn & Packowski, 2015). The most valued use of big data analytics in a supply chain is the ability it provides to analysts in predicting a reaction or an event by detecting changes based on current or historical data (Sanders, 2014). The utilization of current data, is very effective in improving a supply chain which is seeing a start in its use now in industry. Amazon has patented 'Anticipatory Shipping' which predicts based on an analysis of previous orders and other factors such as customers' shopping trend to anticipate that when and by whom a certain product will be bought and ship it in advance and deliver it instantly after the order has been placed (Kopalle, 2014). Another example is that of DHL. DHL is implementing big data analytics to re-route their vehicles and re-define the delivery/picking sequence to save significant time; additionally, DHL has also developed 'MyWays': a crowd-based platform that assigns the parcels to daily commuters, students and taxi drivers by their geo-location and usual routes which in turn improves the efficiency of the last-mile delivery (Jeske, Grüner, & WeiB, 2013).

Most important aspect which hinders maximum utilization of big data is the lack of analytical techniques and applications which could be used to convert the unstructured data from various sources to business intelligence for the user (Sanders, 2014). This calls for more practical applications and techniques to be introduced which use big data analytics for improving decision making in supply chain management. To cater for this call, this paper introduces a framework which utilizes social media big data to update the demand forecast while also using information from the related product's sale. The proposed framework will generate direct implications to supply chain practitioners who are keen to utilize customers' opinions for improving their demand forecasting.

Social Media Analytics

Social Media is defined as " a conversational, distributed mode of content generation, dissemination, and communication among communities" (Zeng et al., 2010 p. 13). Social Media is an effective sensor when it comes to receiving signals from potential customers. Social media data contains emotions, opinions, and preferences which makes it potentially useful as a market sensing platform but with social media data being qualitative, unstructured and subjective form of big data, it calls for a different analytics approach from traditional approach used in big data (Wong, Chan, & Lacka, 2017). Descriptive analytics, network analytics and content analytics have been identified as three major type of analytics which can be used to create value from social media data (Chae, 2015). As the concern of this study is analysis of the text on Twitter and Facebook, content analytics will be used. Three main dimensions have been identified in the content analytics domain through which social media data can be used to create value for a supply chain forecasting in the proposed framework which are sentiment analysis, word analysis and topic modelling.

Sentiment Analysis

Analysing people's opinion, sentiment, evaluation, attitude, judgment and emotions towards tangible or intangible objects, issues or attributes, such as, product, service, organizations, individuals, events, topics is known as Sentiment Analysis (Liu, 2012). Twitter and Facebook are a very tempting source for sentiment analysis due to the variety, velocity and volume (3vs of big data) of the available content. But informal style of posts and tweets, length of tweets, the resulting use of special symbols in posts makes it challenging to extract high performance result from analysis on these sources. Appraisal theory (Scherer, 2005) describes a way to extract sentiment from text. Arnold and Plutchik (1964) introduced the basic concept of the theory. The theory lays basis for structured sentiment extraction that is based on appraisal expression, a basic grammatical unit by which an opinion is expressed . Korenek and Šimko (2014) utilized appraisal theory to analyse microblogs using sentiment analysis and categorize sentiments as positive, negative and neutral. The sentiments have been categorized in the proposed framework utilizing concepts from appraisal theory. Various organizations from different sectors have used sentiment analysis for gathering information, predicting market response, election results, product innovation, improving customer service, stock forecasting and supply chain management as shown in Table 1. Machine learning, lexicon based, statistical and rule based approaches are the most widely used methods for sentiment analysis (Medhat et al., 2014) but n-gram analysis and artificial neural networks methods have also been used (Ghiassi, Skinner, & Zimbra, 2013). Fan et al. (2017) used Naïve Bayes (NB) algorithm for sentiment analysis on online reviews for use in product forecasting. NB algorithm is better suited to classifications where text is treated independently. Cui et al. (2017) used Support Vector Machine (SVM) for classifying text from social media for event detection. In the proposed framework, both NB and SVM algorithm are used but different from all it is being applied on social media data from Twitter and Facebook and is used in conjunction with trend and word analysis results.

Topic Modelling

Social media sources provide huge amount of information every day and with proper tools an understanding of the trends of that information for actionable insights can be developed. Topic Modelling is typically used to uncover industry data across a certain topic or domain (Kwak, Lee, Park, & Moon, 2010), such as product demands, consumer insights, and service quality of an industry. It can help business managers or decision makers to predict the future behaviours or trends of a community based on a relevant set of data. Lansley and Longley (2016) demonstrates a way to use Twitter information to analyse and present geographical trends using Latent Dirichlet Allocation (LDA). Blei, Ng and Jordan (2003) describes LDA as an unsupervised model which is used to find possible topics from collections of text.

Table 1. Studies based on sentiment analysis

Research Topic	Previous work with description
Stock Forecasting	Arias et al. (2013) and Bollen et al. (2011) have used social media analytics for stock forecasting using twitter information. Srivastava et al. (2016) and (Zhang, Xu, & Xue, 2017) used sentiment analysis and transaction data to predict market trends for stock market customers. Ren, Wu and Liu (2018) used SVM with sentiment analysis to predict market movements.
Brand management	Ghiassi et al. (2013) have used sentiment analysis from twitter data for brand management employing techniques such as n-gram analysis and artificial neural networks.
Election results	Oliveira, Bermejo and dos Santos (2017) compared results from sentiment analysis on social media data to traditional opinion surveys and found it 1 to 8% more accurate for predicting election results. Giglietto (2012) used likes on Facebook pages to the study the predictive power of Facebook to forecast Italian elections in 2011.
Product Innovation	KIA motors and The Royal Bank of Canada, have used sentiment analysis to innovate new products (Kite, 2011).
Supply Chain Management	Singh et al. (2017) presented a framework for improving supply chain management in food industry using sentiment analysis. Swain and Cao (2017) explored the sharing of information by supply chain members on social media and by using sentiment analysis gauged its association with supply chain performance.
Box Office Forecasting	Asur and Huberman (2010) presented a study to use data from Twitter for Box Office forecasting using sentiment analysis.
Customer Service	Bank of America used sentiment analysis to recognize key issues facing their customers by collecting and analysing texts from different social media sources (Purcell, 2011). Malhotra et al. (2012) used sentiment analysis to implement improved marketing methods using Twitter.

Word Analysis

Word analysis of social media data encompasses term frequency analysis, word cloud formation and clustering (Chae, 2015). Term frequency is used to identify key words and phrases from the dataset by use of algorithms such as n-gram. N-gram combines adjacent words of length ‘n’ from the given dataset to capture the language structure from statistical point of view. Word cloud is a visually appealing method to get an overview of the text (Heimerl et al., 2014). Word analysis have been used frequently in literature for text summarization (Kuo, Hentrich, Good, & Wilkinson, 2007), opinion mining (Wu et al., 2010) and text visualization (Stasko, Görg, Liu, & Singhal, 2007), patent analysis (Koch et al., 2011) and investigative analysis (Stasko et al., 2007). In the proposed framework, word analysis is used to get an overview of the text being used for the selected keywords and to identify related words to add to the search.

Table 2. Use of social media analytics in supply chain

Research Topic	Previous work with description	Used Feature
Supply Chain Forecasting	(Chong, Li, Ngai, Ch'ng, & Lee, 2016) conducted a study using neural network and sentiment analysis to see effect of online user generated contents on product sales.	Three-layered neural network Sentiment Analysis
	Choi (2016) analytically explored the impact of positive sentiment on social media on market demand of fashion retailers.	Word Analysis
	Beheshti-Kashi (2015) explored whether microblogging websites such as Twitter can be used for predicting fashion trends.	Trend Analysis
	Boldt et al., (2016) tested utilization of Facebook data for predicting sales of Nike Products and the effects of events on activity on Nike's Facebook pages.	Event Study
Supply Chain Management	Chae (2015) developed a framework to study usefulness of twitter information in supply chain management.	Descriptive Analytics Content Analytics Network Analytics
	Sianipar and Yudoko (2014) concluded in their work that social media integration with a supply chain can be helpful to improve collaboration among supply chains and to increase the agile response of a supply chain.	Content Analysis
	Singh et al. (2017) presented a framework for improving supply chain management in food industry using sentiment analysis	Sentiment Analysis

Social Media Analytics in Supply Chain

Getting accurate information from extremely noisy data such as social media data, is a big challenge and as is unifying all social media data and making sense of it, which hinders wide use of social media analytics. Table 2 lists the major studies which have used social media big data in supply chain management. In the last few years, there has been a growing interest in utilizing value from social media data in supply chain management as evident from Table 2. But there is still a lack of accurate models for supply chain management which utilize social media data. One of the reason is that with extremely noisy sources such a social media getting the external casual factors right is a big challenge. Making sense of all the casual data (particularly social media) poses a big question for supply chain practitioners and software developers and requires further research (Syntetos et al.,2016). The framework proposed in this paper tries to address this issue.

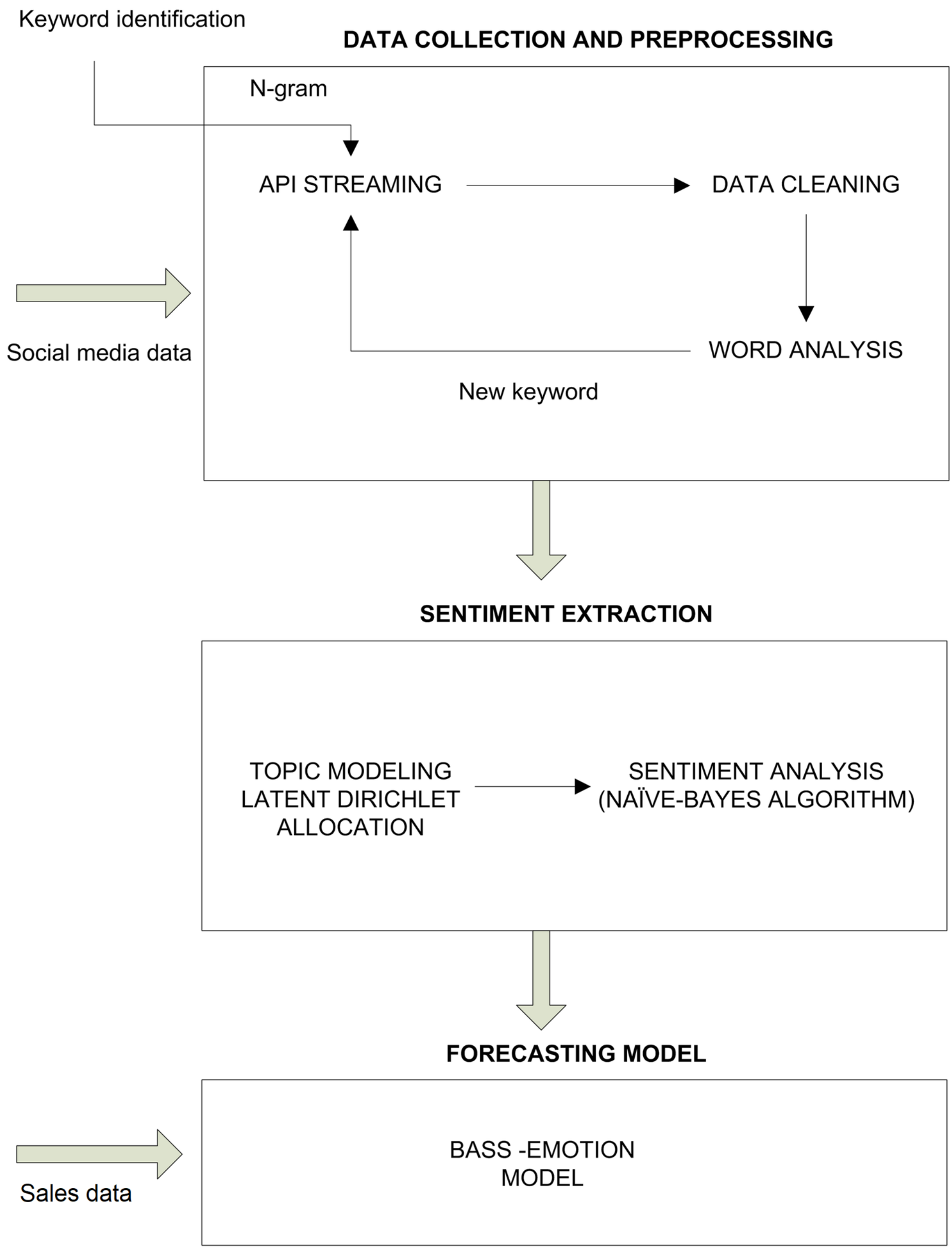
FRAMEWORK

The authors have developed a framework for extracting maximum benefits out of social media in terms of product forecasting. Three main dimensions were identified from the literature and experimentation through which social media data can be used to create value in demand forecasting which are sentiment analysis, word analysis and topic modelling. The framework utilizes these dimension for using social media analytics to improve demand forecasting. The framework consists of data collection and preprocessing, sentiment extraction and building of forecasting model as shown in Figure 1.

Data Collection and Preprocessing

Data is collected and preprocessed using following methods in the given order.

Figure 1. Overview of the demand forecasting framework using social media big data



Keywords Identification

The first step is to identify the initial keywords to be provided by the user. Keywords are used to harvest public data from Facebook and Twitter which are selected after input from the user. N-gram is then applied.

API Streaming

The process of getting data from Twitter and Facebook is the next step and it starts authentication from Twitter and Facebook APIs and establishing a connection. After the authentication, data can be captured using different platforms such as R and Python.

Data Cleaning

The Twitter and Facebook data extracted contains a lot of details (tweets, posts, number of comments, coordinates, embedded URLs, hashtags, retweet count, number of follower, username, location). This data is then transformed using data parsing, data cleansing and noise cancellation to get only relevant data for analysis. All those SMDs (Social Media datasets) collected from Facebook and Twitter are to be neglected which contained less than three words as they didn't represent the customer comments in focus. SMDs from users with 2000 plus posts or tweets are also discarded. If a user is tweeting or posting on the same subject with high frequency those will also be discarded to prevent bias as the results which include these are skewed by the company's marketing campaign. Beheshti-Kashi, Karimi, Thoben, Lütjen, & Teucke (2015) had similar results in their study when they found URLs linked of such tweets and posts to eBay shops. In the final step of data cleansing, the pre-processing of the collected data is done which is mainly cleaning the data. This includes removing URL links, symbols, punctuation and spaces to transform cases.

Word Analysis

Word analysis of social media data encompasses term frequency analysis, word cloud formation and clustering (Chae, 2015). Term frequency is used to identify key words and phrases from the dataset by use of algorithms such as n-gram. In the proposed framework, n-grams that occur with frequency above the selected threshold are selected. This step involves identifying keywords for the products using word analysis. It is then later compared to quantitative result from the sentimental analysis obtained by rating positive and negative words being used. Bounding Boxes and restricting region approach is used which helps in extracting more useful data from the API (Singh et al., 2017). Specific keywords and exact regions are used to make sure of the accuracy of the data.

Sentiment Extraction

In the second major part of the framwork topic modelling is performed to form different groups of text extraced from Facebook and Twitter in terms of product type, colour and brand.

Topic Modelling

LDA is used in the proposed framework to identify topics related to a product and then perform sentiment analysis on the groups. It is described as an unsupervised model which is used to find possible topics from text collections (Blei et al., 2003). LDA is applied using R and the library 'topicmodels'.

Sentiment Analysis

Liu (2012) provides an English Lexicon of about 6800 words which has been amended and used for the purpose of Sentiment Analysis . NB method (Yu et al., 2013) is used for polarity classification with the aim of obtaining a sentiment index for each SMD. Three categories of sentiment are positive, negative and neutral. The value of W_{tk} is calculated using the NB and SVM method. 'R' is the software used in this study. NB is applied using 'E1071' library in R and SVM using 'caret' package in R. 'Caret' package has in built algorithms for different machine learning algorithms including decision tree, K-Nearest Neighbours(KNN) and SVM. In this instance, the authors are using only SVM from caret package.

The sentiment index in time period t, W_t, is calculated by $W_t = \sum_h (W_{tk} \times c)$ where value of 'c' is from 1 to -1 depending on the category of W_{tk} i.e. sentiment value of the SMD(positive, negative, neutral) and h is the number of SMDs.

Forecasting Model

In this framework, the Bass Emotion Model (Fan et al., 2017) is extended to include sentiment analysis results from SMDs collected in the first step. In the Bass model (Bass, 2004), potential buyers are classified as innovators and imitators, and then the general form of the Bass model is as follows.

$$S(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} \times e^{-(p+q)t}}$$

where S(t) is the cumulative sales by the end of time period t. p refers to the coefficient of innovation, q refers to the coefficient of imitation, and m refers to the total number of potential adopters. m and p are calculated using historical sales data. q is related to the sentiment and can be perceived as a function of the social media sentiment $q = f(W_t)$. From the SMDs, if positive sentiment is obtained it means that social media users are talking positively about the product and it gives a potential increase in adopters q and vice versa. The function is described as

$$q = \frac{q^m q^0}{q^0 + \left(q^m - q^0\right) e^{-\gamma W_t}}$$

where q denotes the effect of word of mouth via social media. q^0 refers to the minimum of q, q^m refers to the maximum of q. Υ is a constant that represents the slope of the sales curve. Υ is calculated using historical product data.

CASE STUDY

The study was conducted at an apparel retail company. Focal company's business model is buying and selling apparel products. The suppliers are from different countries encapsulating Far East, South Asia and Europe. Clothes are imported from these countries as well as bought from the local market and then sold to more than 60 countries throughout the world. The complete supply chain is huge spanning four continents. The focal apparel retail company was chosen because of importance of customer-oriented content in apparel industry and because of the focal company's significant presence on social media.

It is difficult to coordinate longer apparel supply chains, so it becomes really important to have very accurate demand forecasting (Syntetos et al., 2016). Traditional forecasting methods like time series data don't work particularly will in an apparel industry as designs and items of one season are typically replaced next season by new collections and trends, and therefore, companies often face a lack of historical sales data (Thomassey, 2010). Moreover, demand in the industry is significantly influenced by additional factors such as the economic situation, events or changing weather conditions (Thomassey, 2014). Many practitioners have been using univariate method (Au et al., 2008) for supply chain forecasting in apparel industry which utilizes historical sales data and it is assumed that the underlying variation of data is constant. For instance, Wong and Guo (2010) utilized one-step-ahead sales data to predict the sales of medium-priced fashion products in Mainland China. Au et al. (2008) used previous time series data to predict the sales of T-shirt and jeans from several shops with the use of neural networks. The sales of products in apparel industry are volatile, often influenced by changing trends and weather conditions and events. So, for the forecasting purposes, it is not right to hypothesize that the trend of time series sales data is unchanged. To cope with this, researchers integrate other influencing factors as the inputs of forecasting models besides the historical time series data, which is known as multivariate forecasting. Beheshti-Kashi (2015) has presented current fashion forecasting approaches in the industry and academia. Most successful techniques surveyed were Extreme machine learning(Sun, Choi, Au, & Yu, 2008), evolutionary neural network (ENN) (Au et al., 2008; Wong & Guo, 2010), Thomassey and Happiette fuzzy inference systems (Thomassey, Happiette, & Castelain, 2005) and hybrid intelligent sales forecasting model (Aburto & Weber, 2007).

Most of the forecasting models discussed above give reliable results for middle and long-term forecasting. But due to a very competitive market and short selling span accurate and customer centric and short-term forecasting is necessary. With the advent of information technology and affordable information systems, most companies (big and small) have developed or implemented information systems from which they get sales reports, graphs and even forecasts. With the advent of social media data, this is not enough to be competitive. Data gathered by the companies needs to add the information circulating on social media, which could deliver another type of insight for forecasting and result in the increased competitiveness especially for creative industry such as apparel industry with the involvement of potential customers in style design, colour preference and judging trends, and scope for new products (Banica & Hagiu, 2016).

Short term forecasting methods have not been explored as much (N. Liu, Ren, Choi, Hui, & Ng, 2013). Short term forecasting is very important in the apparel industry because of the ever-changing trends and short selling times. For this purpose, Beheshti (2015) suggested adding social media to the discussion of fashion forecasting and Syntetos et al. (2016) predicted that future of supply chain forecasting will include predictive analytics based on social media data. For an apparel supply chain, there can be multiple topics of interest which are being discussed in social media. The authors try to utilize these topics to make this data viable using the proposed framework for supply chain forecasting in apparel industry.

For the implementation of the framework, company sales and social media data i.e. Twitter and Facebook data was collected. This data was collected for a period of six weeks. Data collection for this study began in July 2016 and data was collected till August 15, 2016. Beheshti-Kashi (2015) did a study for exploration of trends using twitter and found out it hard to present the finding in quantitative form. To cater for this issue, the authors expanded the study by analysing specificities and increased the amount of data collection by including both Facebook and Twitter so results could be presented in quantitative form. The period of six weeks was chosen with the insights from the user, which in this case is the supply chain manager of the focal company. 'Shorts' were selected as the product to be used for the study. For collection of data from social media i.e. Twitter and Facebook, APIs were used and the related SMDs was analysed. Only those SMDs were selected which were either brand related, product type related, or a fashion trend related. Data was collected every 7 days as twitter allowed tweets to be collected which were 7-8 days old. SMDs were extracted for brand and products. Hashtags and texts for the brands sold by the focal company were analysed. The total number of tweets analysed were 1,208,650. For the category product type shorts were chosen as they were the most selling item as the data was collected in summers. SMDs were collected against different type of shorts as shown in Table 3 and for different brands as shown in Table 4. As this data of brands was analysed there were a lot of data which wasn't related to the brand or products of the focal company. One such example was #next being used for election campaign in United States. After extraction of text, it was used to form word clouds which can be helpful in manual inspection of the data gathered as the viewer can get a general idea about the kind of words being used and this can later be used for cross checking the results obtained by sentimental analysis to make sure no anomaly has occurred during the process. Word Clouds were formed before and after processing and cleaning of data to investigate manually the dataset being used for sentiment extraction. Figure 2 displays a word cloud for keyword ‘nike’ before data cleaning process. The noise in this dataset is evident as there are words from different languages and some completely unrelated words. Figure 3 displays the word cloud after data cleaning which removes all the unrelated SMDs.

For a period of 6 weeks, the SMDs were analysed and then compared to the sales period for that period as well as next 6 weeks. Table 5 shows the sentiment analysis score for different product categories after application of SVM and then calculation of parameter q. Analysis of sentiment score show that the amount of sales had a co relation with the sentiment around that particular brand or colour. There was no co relation found when sentiment analysis was done for the product type which could be attributed to the noise in the data as single word or single product search was susceptible to much more noise than a search using words for multiple characteristics. Multiple character searches with positive sentiment lead to an increase in sale and the negative sentiment lead to a decrease. Analysing the tweets and Facebook comments for running shorts and running a sentiment analysis on it using SVM and NB methods. Comparison of the results of these models have been shown in Table 7.

Figure 2. Word cloud for brand ‘Nike’

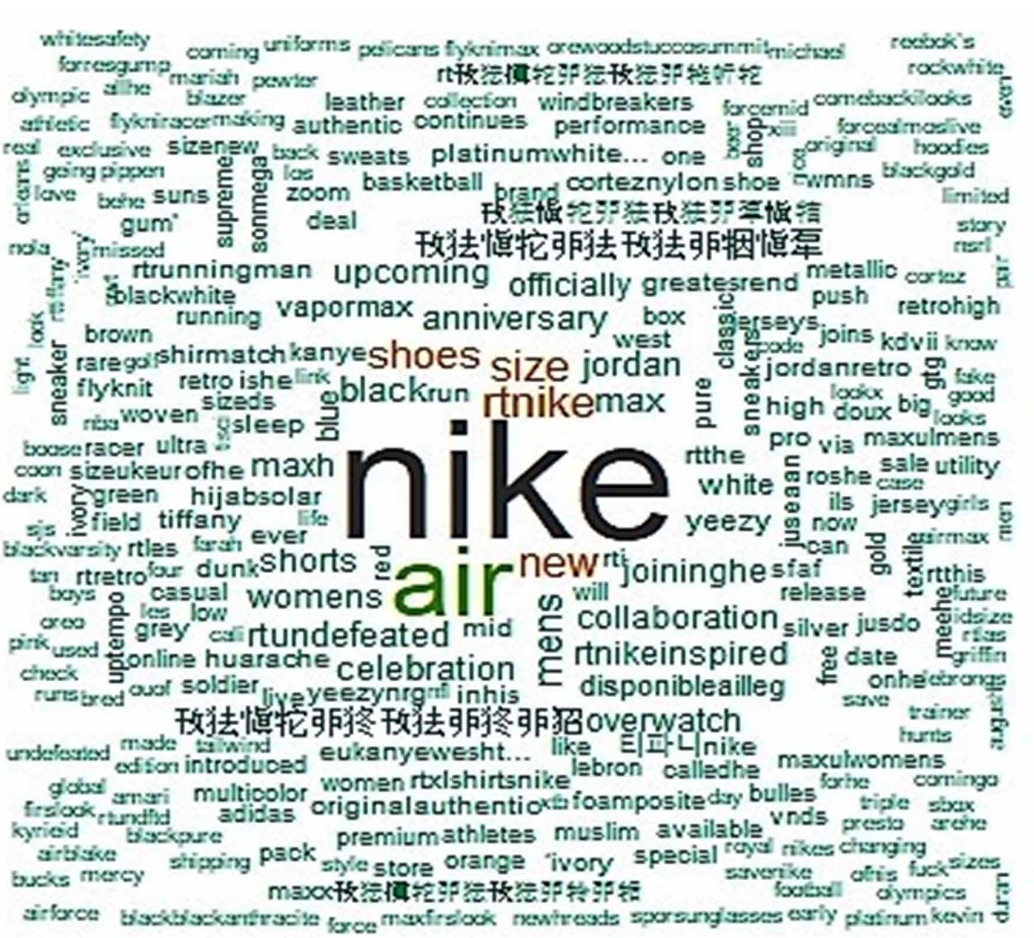


Figure 3. Word cloud after data cleaning

The results from sentiment analysis were then used in Bass Emotion model to predict the sales. The parameters m,p and γ for Bass- Emotion model were calculated using historical sales data and q was calculated using sentiment analysis from SMDs. Parameters calculated are represented in Table 8. All these parameters were calculated using R. Table 6 shows the forecasting accuracy of the proposed emotion enhanced model which is a significant improvement on the forecasting accuracy of original Bass Model. Figure 4 displays the forecasted values using proposed model compared to actual values.

CONCLUSION

This paper introduced a framework that provides a way of utilizing social media big data in Bass-Emotion Model for demand forecasting using results from sentiment analysis on Facebook and Twitter data. As social media data is very noisy, it is difficult to make accurate predictions from social media data about products in general but if the products are broken down and multiple characteristics search is applied then the information which is collected can be converted as a demand forecasting and market or trend sensing tool. The major factor in extracting value from the social media is to apply multiple data cleaning techniques in conjunction with one another, so the data subjected to

Table 3. Keywords used for SMDs extraction for 'shorts'

Shorts#nike	Shorts#green	Shorts#swimming	zara#swimmingshorts
Shorts#adidas	Shorts#navy	Shorts#running	zara#runningshorts
Shorts#reebok	Shorts #jersey	nike#jerseyshorts	zarablack#jerseyshort
Shorts#next	Shorts #cargo	nike #cargoshorts	zarablack#cargoshorts
Shorts#blue	Shorts#jorts	nike #jorts	zarablack#jorts
Shorts#black	Shorts#fleece	nike #fleeceshorts	zarablack#fleeceshort
Shorts#grey	Shorts#gym	nike #gymshorts	zarablack#gymshorts
Shorts#swimming	nike#swimmingshort	Shorts#swimming	adidas#swimmingshor
Shorts#running	nike#runningshorts	Shorts#running	puma#runningshorts
nike#jerseyshorts	nikeblack#jerseyshor	adidas#jerseyshorts	nikeblack#jerseyshort
nike#cargoshorts	nextblack#cargoshor	adidas#cargoshorts	pumablack#cargoshts
nike #jorts	nike black#jorts	adidas #jorts	nike black#jorts
nike #fleeceshorts	nikeblack#fleeceshor	adidas#fleeceshorts	nikeblack#fleeceshort
adidasShorts#ru	nike#runningshorts	adidasShorts#runni	puma#runningshorts
next#jerseyshorts	nikeblack#jerseyshorts	adidas#jerseyshorts	pumablack#jerseyshorts
next #cargoshorts	nextblack#cargoshors	adidas#cargoshorts	pumblack#cargoshorts
next #jorts	nike black#jorts	adidas #jorts	puma black#jorts
next #fleeceshorts	nikeblack#fleeceshorts	adidas#fleeceshorts	pumablack#fleeceshorts
next #gymshorts	nikeblack#gymshorts	adidas #gymshorts	pumablack#gymshorts

Table 4. Number of Brands and Product Related SMDs for week 1

Brand	# of SMDs	Product Type	# of SMDs
Zara	12,456	#jerseyshorts	651
Nike	29,435	#cargoshorts	543
Adidas	36,792	#jorts	189
NEXT	71,234	#gymshorts	984
BHS	61,281	#swimmingshorts	429
Puma	23,124	#runningshorts	183

later analysis gives reliable results as described in the framework presented in the paper. More than 1200,000 tweets, posts and comments from Facebook and Twitter were analysed in the case study. The study showed that social media big data is extremely useful for apparel industry and can be very effective if used to support demand forecasting. With proper modelling and implementation of right techniques, social media big data has the potential to help forecast with accuracy. Results from this study shows a co relation between customers opinion on Facebook and Twitter to actual sales. The framework presented in this study can be further verified and improved with the help of case studies to make it a reliable mechanism for using social media big data in demand forecasting.

As this a relatively new research area, there is a considerable need for enhancing our understanding social media data in supply chain contexts. One area which needs urgent work, is developing detailed,

Table 5. Product type with sentiment analysis score

Product Type	Sales	Number of SMDs	Sentiment Analysis Score	Product Type	Sales	Number of SMDs	Sentiment Analysis Score
Nike Jersey Shorts	1120	651	0.23	Adidas Jersey Shorts	983	156	0.64
Nike Cargo Shorts	2832	543	0.12	Adidas Cargo Shorts	811	531	0.12
Nike Denim Shorts	563	189	0.70	Adidas Denim Shorts	641	145	0.53
Nike Fleece Shorts	212	84	0.34	Adidas Fleece Shorts	1212	821	0.31
Nike Gym Shorts	984	984	0.05	Adidas Gym Shorts	1944	547	0.43
Nike Swimming Shorts	1367	429	0.76	Adidas Swimming Shorts	937	122	0.53

Table 6. Comparison of forecasted and actual values for Bass Model and proposed Emotion Enhanced Model

Forecasting week	1	2	3	4	5	6
Actual value	712.3409	817.6867	921.2260	843.5641	926.7657	923.9208
Forecasted value (Bass Model)	704.5435	810.4631	927.0904	841.5382	922.7238	918.6123
Forecasted value (Proposed Model)	708.6674	816.5294	923.1996	844.2350	926.8046	922.7927

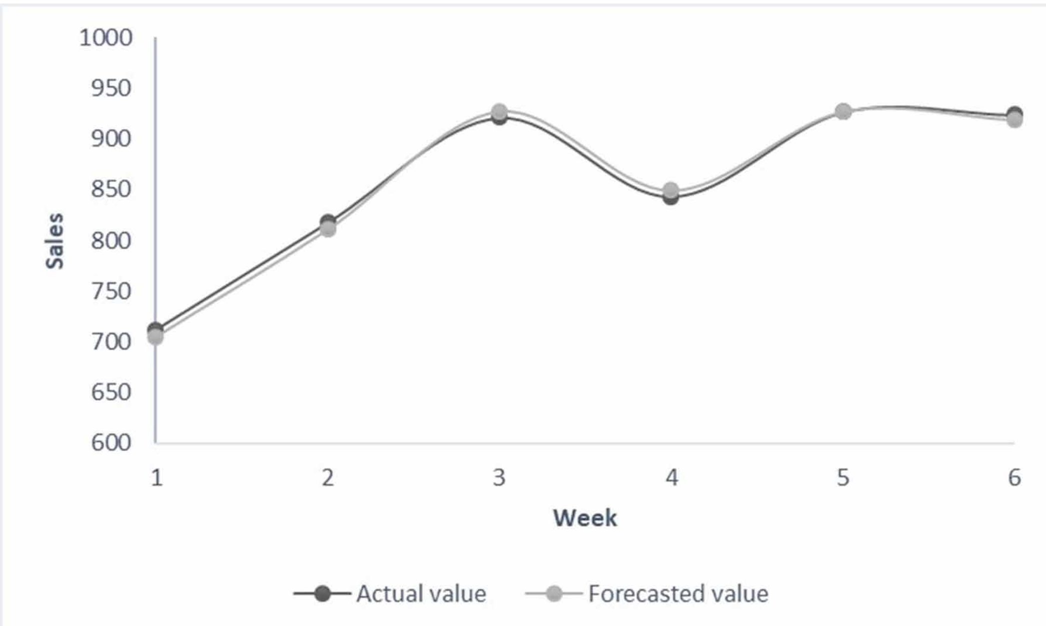
Table 7. Comparison of SVM and NB Methods

Product Brand	Algorithm	Accuracy
Nike	NB	67.21
	SVM	69.24
Adidas	NB	67.46
	SVM	75.12
Puma	NB	65.24
	SVM	71.81
BHS	NB	69.42
	SVM	78.10
Next	NB	63.41
	SVM	63.51
Zara	NB	75.87
	SVM	75.11

Table 8. Parameter for bass model

Parameter	Results
m	887.0306
p	0.023777
q^0	0.090407
q^m	0.093113
γ	0.170784

Figure 4. Results of Forecasting Model of Emotion Enhanced Model



practical guidelines, which can help companies in designing industry applications, using Facebook, Twitter and other social media platforms, for diverse supply chain activities, including new product development, stake holder engagement, supply chain risk management, and market sensing. Further research is needed in the implementation of this framework on other industries and using cloud-based systems. Moreover, sentiment extraction could be improved by including other social media platforms including YouTube, google trends and Instagram. Sentiment analysis can be implemented on videos and pictures posted instead of limiting it only to the text. This could further improve the results as it will take into consideration users from other platforms as well, painting a more accurate picture of customers sentiment.

REFERENCES

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*. doi:10.1016/j.asoc.2005.06.001

Arias, M., Arratia, A., & Xuriguera, R. (2014). Forecasting with Twitter Data. *ACM Transactions on Intelligent Systems and Technology*. doi:10.1145/2542182.2542190

Arnold, M. B., & Plutchik, R. (1964). The Emotions: Facts, Theories and a New Model. *The American Journal of Psychology*. doi:10.2307/1421040

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *Journal of Interactive Marketing*. doi:10.1007/978-1-4419-7142-5

Au, K. F., Choi, T. M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2007.06.013

Banica, L., & Hagiu, A. (2016). Using big data analytics to improve decision-making in apparel supply chains. In Information Systems for the Fashion and Apparel Industry. doi:10.1016/B978-0-08-100571-2.00004-X

Bass, F. M. (2004). A New Product Growth for Model Consumer Durables. *Management Science*. doi:10.1287/mnsc.1040.0264

Beheshti-kashi, S. (2015). *Twitter and Fashion Forecasting : An Exploration of Tweets regarding Trend Identification for Fashion Forecasting*. Academic Press.

Beheshti-Kashi, S., Karimi, H. R., Thoben, K.-D., Lütjen, M., & Teucke, M. (2015). A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering: An Open Access Journal*. 10.1080/21642583.2014.999389

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 10.1162/jmlr.2003.3.4-5.993

Boldt, L. C., Vinayagamoorthy, V., Winder, F., Schnittger, M., Ekran, M., Mukkamala, R. R., & Vatrapu, R. (2016). Forecasting Nike's sales using Facebook data. In *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*. IEEE. doi:10.1109/BigData.2016.7840881

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*. doi:10.1016/j.jocs.2010.12.007

Chae, B. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2014.12.037

Chao, L. (2015). *Big Data Brings Relief to Allergy Medicine Supply Chains - WSJ*. Retrieved September 18, 2017, from https://www.wsj.com/articles/big-data-brings-relief-to-allergy-medicine-supply-chains-1432679948

Choi, T.-M. (2016). *Incorporating social media observations and bounded rationality into fashion quick response supply chains in the big data era*. 10.1016/j.tre.2016.11.006

Choi, T. M. (2018). Incorporating social media observations and bounded rationality into fashion quick response supply chains in the big data era. *Transportation Research Part E, Logistics and Transportation Review*. doi:10.1016/j.tre.2016.11.006

Chong, A. Y. L., Li, B., Ngai, E. W. T., Ch'ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*. doi:10.1108/JFM-03-2013-0017

Cui, W., Wang, P., Du, Y., Chen, X., Guo, D., Li, J., & Zhou, Y. (2017). An algorithm for event detection based on social media data. *Neurocomputing*. doi:10.1016/j.neucom.2016.09.127

Dijcks, J.-P. (2013). *Oracle : Big Data for the Enterprise. Academic Press.*

Fan, Z.-P., Che, Y.-J., & Chen, Z.-Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*. doi:10.1016/j.jbusres.2017.01.010

Feki, M., & Wamba, S. F. (2016). Big Data Analytics-enabled Supply Chain Transformation : A Literature Review. *49th Hawaii International Conference on System Sciences*, 1123–1132. https://doi.org/ doi:10.1109/HICSS.2016.142

Fosso Wamba, S. (2017). Big data analytics and business process innovation. *Business Process Management Journal*. doi:10.1108/BPMJ-02-2017-0046

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*. doi:10.1016/j.eswa.2013.05.057

Guess, A. R. (2015). *Only 0.5% of All Data is Currently Analyzed - DATAVERSITY*. Retrieved September 4, 2017, from http://www.dataversity.net/only-0-5-of-all-data-is-currently-analyzed/

Hahn, G. J., & Packowski, J. (2015). A perspective on applications of in-memory analytics in supply chain management. *Decision Support Systems*, *76*, 45–52. doi:10.1016/j.dss.2015.01.003

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1833–1842. doi:10.1109/HICSS.2014.231

Jeske, M., Grüner, M., & Wei, B. F. (2013). *Big data in logistics: A DHL perspective on how to move beyond the hype*. DHL Customer Solutions & Innovation.

Khalil Zadeh, N., Sepehri, M. M., & Farvaresh, H. (2014). Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach. *Mathematical Problems in Engineering*. doi:10.1155/2014/420310

Kiron, D., & Shockley, R. (2015). Creating business value with analytics. *MIT Sloan Management Review*.

Koch, S., Bosch, H., Giereth, M., & Ertl, T. (2011). Iterative integration of visual insights during scalable patent search and analysis. *IEEE Transactions on Visualization and Computer Graphics*. doi:10.1109/TVCG.2010.85

Kopalle, P. (2014). *Why Amazon's Anticipatory Shipping Is Pure Genius*. Retrieved September 4, 2017, from https://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/#5056b0bf4605

Korenek, P., & Šimko, M. (2014). Sentiment analysis on microblog utilizing appraisal theory. *World Wide Web (Bussum)*. doi:10.1007/s11280-013-0247-z

Kuo, B. Y.-L., Hentrich, T., & Good, B. M., & Wilkinson, M. D. (2007). Tag clouds for summarizing web search results. *Proceedings of the 16th International Conference on World Wide Web - WWW '07*. doi:10.1145/1242572.1242766

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is Twitter, a Social Network or a News Media?* Network. doi:10.1145/1772690.1772751

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*. doi:10.1016/j.compenvurbsys.2016.04.002

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. doi:10.2200/S00416ED1V01Y201204HLT016

Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: A review. *Mathematical Problems in Engineering*. doi:10.1155/2013/738675

Malhotra, A., Kubowicz, C., & See, A. (2012). How to Get Your Messages Retweeted. *MIT Sloan Management Review*. https://doi.org/1532-9194

Marshall, P., Dockendorff, M., & Ibáñez, S. (2013). A forecasting system for movie attendance. *Journal of Business Research*, *66*(10), 1800–1806.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. 10.1016/j.asej.2014.04.011

Oliveira, D. J. S., Bermejo, P. H. de S., & dos Santos, P. A. (2017). Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *Journal of Information Technology & Politics*. doi:10.1080/19331681.2016.1214094

Papanagnou, C. I., & Matthews-Amune, O. (2017). Coping with demand volatility in retail pharmacies with the aid of big data exploration. *Computers & Operations Research*.

Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine. *IEEE Systems Journal*.

Sanders, N. R. (2014). *Big data driven supply chain management: A framework for implementing analytics and turning information into intelligence*. Pearson Education.

Schaer, O., Kourentzes, N., & Fildes, R. (2018). Demand forecasting with user-generated online information. *International Journal of Forecasting*.

Scherer, K. R. (2005). Appraisal Theory. In Handbook of Cognition and Emotion. https://doi.org/doi:10.1002/0470013494.ch30

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: The real-world use of big data.* IBM Global Business Services Saïd Business School at the University of Oxford.

Sianipar, C. P. M., & Yudoko, G. (2014). Social media: Toward an integrated human collaboration in supply-chain management. WIT Transactions on Information and Communication Technologies. doi:10.2495/Intelsys130221

Singh, A., Shukla, N., & Mishra, N. (2017). Social media data analytics to improve supply chain management in food industries. *Transportation Research Part E: Logistics and Transportation Review*. https://doi.org/https://doi.org/10.1016/j.tre.2017.05.008

Souza, G. C. (2014). Supply chain analytics. *Business Horizons*. doi:10.1016/j.bushor.2014.06.004

Stasko, J., Görg, C., Liu, Z., & Singhal, K. (2007). Jigsaw: Supporting investigative analysis through interactive visualization. *VAST IEEE Symposium on Visual Analytics Science and Technology 2007, Proceedings*. https://doi.org/ doi:10.1109/VAST.2007.4389006

Sun, Z.-L., Choi, T.-M., Au, K.-F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*. doi:10.1016/j.dss.2008.07.009

Swain, A. K., & Cao, R. Q. (2017). Using sentiment analysis to improve supply chain intelligence. *Information Systems Frontiers*. doi:10.1007/s10796-017-9762-2

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*. doi:10.1016/j.ejor.2015.11.010

Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2010.07.018

Thomassey, S. (2014). Sales Forecasting in Apparel and Fashion Industry. *Intelligent Fashion Forecasting Systems: Models and Applications*. 10.1007/978-3-642-39869-8

Thomassey, S., Happiette, M., & Castelain, J. M. (2005). A global forecasting support system adapted to textile distribution. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2004.03.001

Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, *34*(2), 77–84. doi:10.1111/jbl.12010

Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2016.03.014

Wong, T. C., Chan, H. K., & Lacka, E. (2017). An ANN-based approach of interpreting user-generated comments from social media. *Applied Soft Computing*. doi:10.1016/j.asoc.2016.09.011

Wong, W. K., & Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2010.07.008

Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., & Qu, H. (2010). OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*. doi:10.1109/TVCG.2010.183

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*. doi:10.1016/j.dss.2012.12.028

Zeng, D., Chen, H. C. H., Lusch, R., & Li, S.-H. (2010). Social Media Analytics and Intelligence. *IEEE Intelligent Systems*.

Zhang, G., Xu, L., & Xue, Y. (2017). Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. *Cluster Computing*. doi:10.1007/s10586-017-0803-x

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Rehan Iftikhar is a Marie-Curie Research Fellow and a 2nd year PhD student at School of Business, Maynooth University. He holds a Master's degree in Engineering Management from University of Exeter. His current research interests include digital retail, information systems and big data. His work has appeared in various journals and conference proceedings including Journal of Global Information Management, British Food Journal, Academy of Management Global Proceedings and International Conference on Information Systems Development. Rehan is the corresponding author and can be contacted at: rehan.iftikhar@mu.ie

Mohammad Saud Khan, PhD, is a Senior Lecturer in the area of Strategic Innovation and Entrepreneurship at Victoria University of Wellington, New Zealand. Before taking up this role, he was positioned as a Postdoctoral Researcher at the University of Southern Denmark. Having a background in Mechatronics (Robotics & Automation) Engineering, he has worked as a field engineer in the oil and gas industry with Schlumberger Oilfield Services in Bahrain, Saudi Arabia, and the United Kingdom. His current research interests include innovation management (especially the implications of big data and 3D printing), technology, and social media entrepreneurship.