Electronic Markets – The International Journal on Networked Business

Full Title of Article:	Network Promoter Score *(NePS)*: An Indicator of Product Sales in E-commerce Retailing Sector
Subtitle (optional):	
Preferred Abbreviated Title for Running Head (maximum of 65 characters including spaces)	*NePS*: An Indicator of Product Sales
Key Words (for indexing and abstract services – up to 6 words):	E-commerce, Product Sales, Review Network, Network Promoter Score, Deep learning technique.
JEL classification	C. Mathematical and Quantitative Methods C4 Econometric and Statistical Methods: Special Topics C45 Neural Networks and Related Topics C5 Econometric Modeling C52 Model Evaluation, Validation, and Selection C53 Prediction Methods L. Industrial Organization L8 Industry Studies: Services L81 e-Commerce
Word Count	Total Words: 10243* Headers: 25 Math Inline: 465 Math Display: 31
Word Processing Program Name and Version Number:	We have used Overleaf (https://www.overleaf.com/) to write our manuscript. This platform generates the word count.

*NOTE: The preferred article length for research papers in EM is around 8,000 words (excl. reference list). Our manuscript *exceeds* the prescribed *word count*. Please consider it, because our research topic demands this length of treatment.

Noname manuscript No.
(will be inserted by the editor)

Network Promoter Score (*NePS*): An Indicator of Product Sales in E-commerce Retailing Sector

Abstract E-commerce companies want to predict their future product sales from the current customers' feedback to frame a better business strategy. However, the conventional way of analyzing rating activities or quality and sentiment of reviews, volume of sales, or product prices is not enough for establishing a strong regression between these parameters and future product sales. Most of the existing works ignore the heterogeneous positional and influential effects of individual customer reviews and ratings. For the realization of these effects, we use review network *i.e.*, a bipartite network between customers and products based on the customers' review activities. In this paper, we present a concept named *Network Promoter Score* (*NePS*) based on the reliability, positional influence of each customer in the network. In-depth experiments on online review datasets show that *NePS* emerges as a strong indicator of product sales and can be remarkably futuristic compared to the existing parameters. Furthermore, we propose a predictive modeling technique to estimate the product sales of a company based on *NePS*.

Keywords E-commerce, Product Sales, Review Network, Network Promoter Score, Deep learning technique.

1 Introduction

Estimating product sales is always one of the top prioritized tasks of any company's or brand's higher management. It helps a manager to keep track of the company's performance in advance to identify any future risk in the company's growth profile (Das & Chen, 2007; Korneta, 2018). In extreme cases, the prediction of product sales warns about unwanted phenomena like bankruptcy and failure (Fisher & Kordupleski, 2019). In general, an estimation of positive product sales gives a good indicator of the company's future revenue growth. It helps to build business and marketing policies and establishes a strong bonding between the customers and the company (Reichheld, 2003; Filieri et al., 2021).

In the conventional offline business world, a popular metric to gauge the loyalty of a brand's customer relationships is *Net Promoter Score (NPS®)*. *NPS®* is a registered trademark of Frederick Reichheld, Bain & Company, and Satmetrix Systems. This concept was introduced in *Harvard Business Review* in 2003 (Reichheld, 2003). Interestingly, it is still widely adopted with more than two-thirds of Fortune 1000 companies (Korneta, 2018; Filieri et al., 2021). It asks just one query: "How likely is it that you would recommend our company/product/service to a friend or a colleague?" in a 0 to 10 scale. A manual survey of customers' feedback is done with that one question about customers' satisfaction. *NPS®* identifies a customer as promoter if the customer gives rating 9 or 10. A customer, who gives a response with 7 or 8, is called a passive customer and the customers having responses with 0 to 6 rating are known as detractors. *NPS®* is defined as % of promoters - % of detractors. A positive *NPS®* is considered to be good and a *NPS®* of 50+ is treated as excellent. At the initial years it was evaluated from the survey of word-of-mouth (WOM) technique (Q. Liu et al., 2019; Baek et al., 2017; Lin et al., 2017). Essentially, what *NPS®* does in traditional offline businesses, the same thing is replicated in online platforms by analyzing reviews and ratings (Fisher & Kordupleski, 2019). The marketing strategy behind *NPS®* is word-of-mouth (WOM) and in E-commerce world it is called electronic word-of-mouth ($eWOM$).

In our daily life, customer-generated contents (CGC) are part of the online activities of online customers. From communications to e-business, everywhere customers are taking a vital role in the spreading and utilization of digital information those are increasing at this age of globalization and specialization. Up from 83 million in 2008, the eMarketer (definitive research source for understanding how digital is changing media, commerce, marketing, and business strategy) calculates that the number of CGC creators is fattened rapidly to 115 million in 2013. The number of US internet customers was 116 million in 2008 and in 2013, the number of US internet customers has reached 155 million. Today whenever people are facing any situation, they try to understand other's opinions by reading some posts, listening to speeches, or watching some programs. Online feedback is a form of customer-generated content that is a key source of information to customers for choosing their preferred products.

Specifically, online E-commerce companies are now using online feedback portals to promote their products for the ultimate goal of increasing product sales. There are many ways to influence the opinion of customers and as a result, the customers are changing their minds to purchase certain products and services. Online Consumer Reviews ($OCRs$) is one of the crucial sources that influence customers to understand the quality of different items and to find the best one. Some studies recommend that customers show more interest in customers' generated product information on online review pages than the features companies provided (Bickart & Schindler, 2001; Fu et al., 2018; Xue et al., 2020). In literature, this inclination of a customer to a specific product or brand is termed as the loyalty of that customer to the brand.

Online feedback platforms of E-commerce companies generally provide two input options for giving feedback: written review and numerical rating. In many research, the authors have experimented with the effectiveness of online customers' reviews on product sales focusing on numeric ratings. Some authors have experimented with rating valence (J. A. Chevalier & Mayzlin, 2006; Duan et al., 2008; Filieri et al., 2018; X. Li et al., 2019), variance in ratings (Godes & Mayzlin, 2004; Chintagunta et al., 2010) and volume of reviews and ratings (Duan et al., 2008). Some researchers have investigated the relationship between customers' interactions and market sales (Godes et al., 2005). Some works show that the power of strongly positive feedback is also related to market sales (Das & Chen, 2007; Colladon et al., 2019). Some researches focus on customers' purchasing behavior in the e-commerce context (Qiu et al., 2015). How negative emotions affect sales rank are discussed in (Fowdur et al., 2009). Some authors have mentioned that numeric ratings cannot capture the context of review texts (Ghose & Ipeirotis, 2011; L. Li et al., 2020). While numeric ratings express customers' preferences, customers' reviews express their preferences more specifically and give an overview about the quality of the products (Ghose & Ipeirotis, 2011; Malik, 2020).

The traditional ways of summarizing customers' ratings, sentiment analysis of reviews, analyzing the quality of reviews, volume of reviews, and product prices are not sufficient to establish a strong regression between these parameters and product sales. Most of the existing works overlook the positional and influential effects of individual customer's reviews on the other customers regarding their purchasing decisions. There are different kinds of reviewers in e-commerce sites that defined their effective influence. In real-world datasets, we observe that there are some positive reviewers (who give always positive or good ratings), some negative reviewers (who give always negative or bad ratings), some reliable reviewers (who give ratings according to the quality of products), and some reviewers are whimsical (whose posted ratings are unpredictable). In general, after analyzing reviews, most of the researchers conclude that "what" and "how" customers express their opinions in reviews are the important factors that affect product sales (Ghose & Ipeirotis, 2011; Netzer et al., 2012; Filieri et al., 2021). In this work, we establish that "who" writes the review is also very important and we evaluate the positional and influential effects of "who" based on the structure of the review network.

The strategies of *NPS*® and other existing works overlook the reliability and positional influence of individual customer's reviews on other customers regarding their purchasing decisions. *NPS*® does not analyze the characteristic of a reviewer regarding the reviewer's reliability or bias. There are different kinds of reviewers in e-commerce sites based on their effective influence. From real-world datasets, we notice that there are some positive reviewers (who give always positive or good ratings), some negative reviewers (who give always negative or bad ratings), some reliable reviewers (who give ratings according to the quality of products), and some reviewers are whimsical (whose posted ratings are unpredictable). In general, existing studies conclude that "what" and "how" customers express their opinions in reviews are the important factors that affect product sales (Ghose & Ipeirotis, 2011; Netzer et al., 2012; Owen, 2019). Existing studies fail to analyze the characteristic of "who" writes the review which is also very important to understand future product sales.

To fulfill this research gap, we introduce *Network Promoter Score* of a customer's review or rating for a particular product. Practically, a positive review or rating of an influential customer can influence other customers to purchase the product, on the other hand, a negative review of an influential customer creates a negative impact on other customers. For example, if 10 customers give positive ratings, 5 customers give negative ratings and 3 customers give neutral ratings, *NPS®* and most of the existing works just count the number, where our work analyses the characteristic of each reviewer in terms of the reviewer's reliability, positional influence in the review network and evaluates *Network Promoter Score*.

To evaluate *Network Promoter Score*, we use two different network parameters, *i.e.*, clustering coefficient and PageRank (In network science community it is known as Eigenvector centrality) from the topological structure of the network. One of the main contributions of our paper is to use the positional influence of a customer to predict future product sales. To measure a customer's influential power, we avoid making any arbitrary choice of parameters. In this work, we try to give due credit to the network parameters as indirect feedback. Network parameters represent the inter-personal relations and interaction among the customers. Our hypothesis is that these interactions have significant effects in shaping the behavior of a customer. From network science point of view two most important parameters are centrality and clustering coefficient. That is the reason behind choosing those two parameters in this work. However, we agree that other network parameters can also be used in this model, albeit their influence would be much less than centrality and clustering coefficient.

Network Promoter Score informs the manager about the heterogeneous reliability and influential power of all customers. This score varies from customer to customer for a particular product. Point to be noted that this score of a customer may vary from product to product as well. That is why it is defined as heterogeneous reliability. Network Promoter Score of a particular review on a product is denoted in the small letter as ***neps***. The monthly average network promoter score for a particular product overall reviews it got in a month is denoted as ***NePS***. Our experiment establishes a strong regression between the *NePS* and product sales. Besides, we perform experiments on the latency time, *i.e.*, how much time the *NePS* will take to affect product sales more positively.

After establishing a strong regression between the *NePS* and product sales, we propose a deep learning model to predict latency time and product sales figures with high accuracy. We design a Long Short Term Memory Model (*LSTM*) that is capable of understanding the trading patterns of sales rank based on the *NePS*. *LSTM*, a variant of Recurrent Neural Network (*RNN*), is discussed in details in (Wu et al., 2016; Elkahky et al., 2015; Goodfellow et al., 2016; Abbasimehr et al., 2020). The predicted sales rank is obtained via a regression layer (L. Zhang et al., 2017).

The experiments on online review datasets show that *NePS* emerges as a strong indicator of product sales and can be remarkably futuristic compared to the existing works. Furthermore, we investigate latency period based on *NePS* and other parameters, *i.e.* average ratings, average sentiment, price, quality of reviews and volume of positive reviews. With the *NePS* as an input in *LSTM* model, we investigate to forecast whether product sales will go up or go down based on different latency periods. The prediction accuracy of *LSTM* model is the best compare to the baselines.

The rest of the paper is arranged as follows. The related literature is discussed in Sec. 2. Then we demonstrate our independent parameter *NePS* and highlight other existing parameters in Sec. 3. After that, our approach is discussed in Sec. 4. In this section, we evaluate the importance of different parameters on product sales. We demonstrate a deep learning technique that estimates product sales based on the *NePS*. The data statistics and baselines are discussed in Sec. 5. Sec. 6 presents the experimental results of our investigation. Finally, we provide some interesting directions for future works and conclude the paper.

2 Related Literature

In (Dellarocas et al., 2007; Malik, 2020), a relation between numeric ratings and sales figure have been discussed. Some researchers also stated that review volume and sales might be positively related (Forman et al., 2008). A better item receives more positive reviews. In (Reichheld, 2003; Forman et al., 2008; Eslami & Ghasemaghaei, 2018), the authors have shown that more positive reviews and product sales are positively related. In some research works (Ghose & Ipeirotis, 2011; Maslowska et al., 2017), the authors have investigated how the textual features of a review affect product sales. Only numeric ratings, the volume of reviews, total number of reviews, the total number of positive ratings are not sufficient to understand the regression with product sales.

Less spelling and grammatical error-based reviews with reasonable length affect product sales positively (X. Li et al., 2019; Siering et al., 2018). Recently, some authors are focusing on the quality title of a review rather than the review text (Salehan & Kim, 2016; Hong et al., 2017). They have observed that customers are influenced more by a quality title of review rather than the review text (Eslami et al., 2018; Xue et al., 2020). In general, customers do not want to spend much time reading a long review text. They prefer review titles as a quick source of information.

Some researchers have focused on the helpful votes of customers' reviews and they have observed that customers are influenced more by the reviews with the high number of helpful votes rather than the content of the reviews (Korfiatis et al., 2012; Stouthuysen et al., 2018). How helpful votes are related

Table 1: Literature reviews of the impact of customer-generated content on Product Sales.

Papers	Experiment on	Sample size
(Reichheld, 2003)	Word-of-mouth (WOM), evaluates the number of promoters and detractors based on ratings. The evaluated score is used as an indicator of product sales.	Satmetrix collects 10,000 to 15,000 responses from more than 400 companies and many other industries.
(Godes et al., 2005)	Social interaction, WOM relate to product sales.	144 (audio & video players), 109 (digital cameras), 158 (DVD players).
(Clemons et al., 2006)	Variance of ratings, volume of influential reviews relate to product sale.	1159 (beer companies).
(Das & Chen, 2007)	Use text mining to extract the sentiment of review text and point out the effectiveness of the sentiment score of reviews on product sales.	145,110 (stocks).
(Duan et al., 2008)	Rating, volume of online posting, suggestion about item, price relate to product sales.	71 movies.
(Forman et al., 2008)	Ratings, analyse reviewer characteristics based on number of purchased items , geographical location-based purchasing behavior relate to product sale.	786 Books.
(Chintagunta et al., 2010)	Valence, volume, and variance of online customer feedback relate to product sales.	148 movies.
(Ghose & Ipeirotis, 2011)	Rating, sentiment of review based on readability and subjectivity, estimate helpfulness score of review. Latency time period is considered.	Audio and video players (144 items), Digital Cameras and related products (109 items), DVDs and related items (158 items).
(Baek et al., 2012)	Review rating, reviewers' activities, the content of reviews and helpfulness of reviews.	75,226 reviews from Amazon of 28 item category.
(Hu et al., 2014)	Rating and sentiment of review; quick source of information influences customers more.	4405 books.
(Chua & Banerjee, 2015)	Sentiment analysis on review text; title of review rather than the review text relates to product sales.	7,897 reviews of 150 books collected from Amazon.com.
(Fang et al., 2016)	Sentiment on review text and how length of reviews relates to product sales.	19,674 online reviews of a tourism destination (i.e., New Orleans) from TripAdvisor.
(Karimi & Wang, 2017)	Customers are influenced more by the reviews with high number of helpful votes rather than the content of the review.	2178 reviews are collected from mobile gaming applications.
(Filieri et al., 2018)	Analyzing context of reviews, purchase decision depends on helpful votes.	Data were collected using Survey Monkey at Hong Kong International Airport.
(X. Li et al., 2019)	Proposing Joint Sentiment-Topic model to understand the contexts and sentiments in review texts and investigating how helpful votes relate to product sales.	Data were collected from Amazon.com.
(Malik, 2020)	Review title, sentiment and polarity of review text and cosine similarity between review text and product title effectively contribute to the helpfulness of customers' reviews.	Reviews from Amazon.

to the review quality are described in (X. Li et al., 2019; Siering et al., 2018; Ren & Hong, 2019). However, how helpful votes affect product sales is not discussed in those works. There are some limited works (Filieri et al., 2018; Sun et al., 2019; Kang & Zhou, 2019; L. Li et al., 2020) which focus on the regression between helpful votes and product sales. In these works, helpful votes are treated as the sole influencing factor.

Table 1 depicts a list of the most important literature on the impact of customer-generated content on product sales. Here we present a brief discussion about the research gap based on Table 1. In the initial years *NPS*® was evaluated from the survey of word-of-mouth (*WOM*) technique (Reichheld, 2003; Godes & Mayzlin, 2004). Nowadays, *NPS*® is also applied on online feedback platforms (Fisher & Kordupleski, 2019) (also known as electronic *WOM* or *eWOM*). All these works evaluate the number of promoters and number of detractors based on customers' rating activities in the *WOM* or *eWOM* technique. These works do not consider the heterogeneous reliability or positional influence of promoters and detractors which can affect future product sales.

In (Clemons et al., 2006), the authors focus on review texts. They identify influential reviews based on positive and negative reviews. The variance of ratings and volume of influential reviews relate to product sales. The authors give the same importance to all positive reviews. In (Clemons et al., 2006), the authors do not consider rating activities. The authors use text mining to extract the sentiment of review text and point out the effectiveness of the sentiment score of reviews on product sales. This work faces the in-authenticity problem of customers' posted reviews.

In (Duan et al., 2008), the authors consider several parameters, *i.e*, rating, the volume of online post, suggestions about products, price relate to product sales. Based on the suggestion from customers about items, the authors try to understand the present quality of items that indicates future product sales. The authors also mention that the current price of items is also a crucial factor to understand product sales. In (Forman et al., 2008), the authors consider ratings, geographical location-based purchasing behavior relates to product sale. The authors consider customers' reliability based on the number of purchased items where purchasing more products means reliable. They do not analyze customers' characteristics in terms of the helpful votes, quality of posted reviews, and influential power.

How valence, volume, and variance of online customer feedback relate to product sales are discussed in (Chintagunta et al., 2010). In (Ghose & Ipeirotis, 2011), the authors have investigated how the textual features of a review affect product sales. Though this work considers the quality of reviews, it overlooks the positional influence of each customer. In (Fang et al., 2016), the authors focus on the length of reviews and its effect on product sales is discussed. They assume that customers always prefer quick sources of information regarding the quality of items. In these works, the reliability of customers is ignored.

In (Baek et al., 2012; Hu et al., 2014; Chua & Banerjee, 2015), the authors have focused on the sentiment score of review and review title. They have also investigated how these factors relate to helpful votes. The authors have observed that customers are influenced more by more helpful votes-based reviews rather than a title of review and the context of review text. How helpful votes are related to product sales is also discussed. In (X. Li et al., 2019; Malik, 2020), the authors focus on the regression between helpful votes and product sales. These works consider helpful votes but it is not the ultimate solution to overcome the reliability problem of customers' reviews. There is no doubt that without having many helpful votes, a review with less spelling and fewer grammatical errors can be effective and can influence other customers to purchase the product.

Unfortunately, due to the "long tail" phenomenon of social systems, there are chances that many good quality reviews would get very few or no helpful votes. So it is very important to consider the quality of reviews to overcome the reliability problem that is ignored in these works. In (Karimi & Wang, 2017; Filieri et al., 2018), the authors show that customers are influenced more by the reviews with a high number of helpful votes rather than the content of the review. How helpful votes relate to product sales is not discussed in these works.

From the above discussion based on Table 1, it is evident that all the existing works overlook the reliability and influential effect of individual customer's reviews on other customers regarding their purchasing decisions. Most of the existing works mainly concentrate on different kinds of independent parameters , *i.e.*, ratings, sentiment analysis of reviews, analyzing the quality of reviews, the total number of reviews that are directly coming from customers or reviewers. There are some deciding parameters that arise from the relations among customers or reviews. To mine these relations we use the concept of *review network* and introduce the parameter *NePS*.

The effects of all the parameters on product sales are not immediate. The latency time (how much time will take to affect product sales) is varied from product to product. In (Ghose & Ipeirotis, 2011; Owen, 2019), the authors have examined whether the reviews' features can be used to understand the effect of a review on product sales. They have observed the latency period. How contents of a review affect other customers' opinions are discussed in (Siering et al., 2018) and they addressed the regression with product sales. In (Ghose & Ipeirotis, 2011; Filieri et al., 2021), the authors wanted to investigate whether they can predict the impact of their parameters on sales. The authors also investigate whether they can predict future product sales rank or product sales based on their investigated parameters. Since the effect of their parameters is not immediate, they investigated the characteristics of latency period.

Although not related to this work, in (Mandal & Maiti, 2020, 2018), some common parameters are used to develop models of recommendation systems. Specifically, the authors considered helpful votes, rank, and currentness of reviews. These works were proposed for rating prediction to recommend the right product to the right customer. In (Mandal & Maiti, 2018), the authors proposed a model named *RHCV-PMF* that considers one explicit feedback and one implicit feedback. In this model, customers' explicit feedbacks' similarity indicates the similarity of their reliability and characteristic and implicit feedback's similarity indicates their preference similarity.

In (Mandal & Maiti, 2020), the authors proposed a Generalized Probabilistic Matrix Factorization ($GPMF$) model which uses multiple parameters of both the types (it may be explicit feedback or implicit feedback or a combination of explicit and implicit feedback) for a better recommendation. This generalization is very important because the effectiveness of each parameter is not the same for recommendations. The model $GPMF$ can estimate the varied effects of different parameters. Based on this effectiveness, more parameters can be added or removed for a better rating prediction. Essentially, these two works (Mandal & Maiti, 2020, 2018) exhibit the importance of different parameters related to reviews and other feedback systems in shaping the mindset of customers.

In the current work, we want to know whether we can predict the financial health of an e-commerce company based on the fact that the customers are the deciding factors for the company's growth. For this we introduce *Network Promoter Score (NePS)* based on the concept of *review network*.

3 Parameters

3.1 Network Promoter Score (*NePS*)

An E-commerce company evaluates customers' behavioral activities based on their reviews and ratings on a brand or brand's items/products. Where all independent or control parameters are directly generated from customers only, here the *NePS* is calculated from a network based structure. Please note that, the Network Promoter Score of a review on a product is denoted in small letter as *neps*. Based on Amazon.com online review dataset, we have developed a bipartite network between customers and products. The name of this network is *review network*. However, this bipartite network structure can be replicated from any review dataset. An edge of this network represents a feedback that is posted by a customer on a product. The network of a brand is built from all the reviews of all the products from that brand. We evaluate the *neps* of a particular customer's review on a product based on other customer's activities and opinions on that review or rating.

Review Network: Before purchasing any product from e-commerce sites, we check previous customers' reviews on that product. A network of customers and products is formed based on their buying activities and timestamps. We name this network as *Review Network*. Fig. 1 presents an example of review network. In this bipartite network, products and customers are denoted by two sets of nodes. If customer u_i posts a review on product p_j then an edge is considered between them. Essentially, each edge denotes a specific review. Please note that each review has a particular time when it is posted by the customers. Each edge is identified by a unique number that presents a time stamp when the particular review is posted. Point to be noted that in this figure, original time stamps are not specified. For understanding purposes, some time directions are assigned in Fig. 1. The edges between customer u_1 - item p_1 and customer u_2 - item p_1 are denoted by 1 and 2 respectively, which indicates that customer u_2 gives rating after customer u_1. Here, we suppose that customer u_2 buys the product after customer u_1 based on the review post timings.

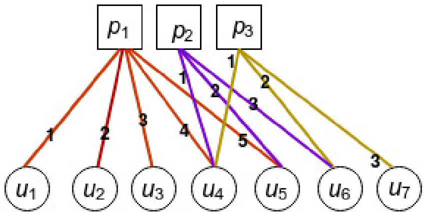

Fig. 1: A bipartite network between Customers and Products. An edge indicates a review that is posted by a customer on the purchased product. Each edge indicates a distinct review. Point to be noted that, each review has a particular time when the review is written or posted. Each edge is denoted by a distinctive numerical digit that indicates when the review is posted. In this figure, original time stamps are not mentioned. For understanding purposes, we have presumed some time stamp.

After buying any item, customers may post rating. We denote r_{ij} as the rating given by customer u_i on product p_j. Without loss of generality, original rating r_{ij} is mapped to an interval of $[-2, 2]$. We choose this specific interval so that negative sign indicates that the customer has given a bad rating and positive sign indicates that the customer has given a good rating. For example, the Amazon.com datasets' ratings range from 1 to 5. In our work, Amazon.com ratings 1 and 2 are mapped into -2 and -1 respectively. Similarly, rating 3 is mapped into 0 which means it is neutral and ratings 4 and 5 are mapped into $+1$ and $+2$, respectively.

The *NePS* is defined using a few component scores. Next, we discuss them one by one and show how to use them to build the *NePS*.

3.1.1 Helpfulness of a review

Before buying a product, customers can check the previous feedback or reviews regarding that particular product. From these reviews, customer can judge the quality of the product and the satisfaction level of the previous customers. In most of the merchandise sites, after each review, sites post a query, "Was this review helpful to you? (Answer Yes/No)". We define helpfulness as a measure for the validation of a review. We normalize the helpfulness h_{ij} of customer u_i for product p_j (i.e., u_i wrote review f_{ij} on p_j) based on how many other customers are helpful from the review and the formula is as follows:

$$h_{ij} = \frac{b_{ij}}{\sum_{y=1}^{n_{p_j}} b_{yj}}, where \tag{1}$$

$$b_{ij} = \frac{|h^*_{ij}|^2}{|v^*_{ij}|}. \tag{2}$$

Here, $|h^*_{ij}|$ indicates the total number of helpful votes on feedback f_{ij} and $|v^*_{ij}|$ indicates the total votes on feedback f_{ij}. The number of customers who buy p_j, is denoted as n_{p_j}. We assign more weight to the particular customers who gain more helpful votes. Thus, Eq. 2 is quadratic in nature. As a special case, if any dataset only contains the helpful vote information without the number of total votes, then the denominator of Eq. 2 will be replaced by the maximum helpful votes gained by any review on product p_j.

3.1.2 Top ranking and most recent reviews

Usually, when we want to read previous reviews for a particular item, we have two options to choose, *i.e.*, i) to read the top-ranking reviews and ii) to read the most recent reviews (Mandal & Maiti, 2020, 2018). In our investigation, it is found that generally customers prefer to read the top-ranking reviews without wasting much time and come to a decision to buy or not to buy. The top-ranking reviews are selected based on helpful votes. Sometimes, customers can prefer to read most recent reviews of a particular product to understand the current quality of the product. Some customers can prefer to read both the top-ranking reviews and most recent reviews to come to a purchasing decision.

Based on the chronological order of all the reviews on product p_j we evaluate the currentness score $most_{ij}$ of review f_{ij} (written by customer u_i on product p_j). First we evaluate non-normalized currentness score z_{ij} as follows:

$$z_{ij} = \sum_{s=1}^{n_{p_j}-i} \frac{1}{s^2}. \tag{3}$$

Here, n_{p_j} denotes the total number of customers who purchase product p_j and $n_{p_j} - i$ is the total number of customers who buy the product after i^{th} customer. Essentially, $n_{p_j} - i$ is the number of customers who were potentially influenced by review f_{ij} to buy product p_j. Here, z_{ij} is denoting the cumulative diminishing currentness measure over all the potential readers of review f_{ij}. In Eq. 3, this diminishing effect is represented by s^2 in the denominator. The normalized z_{ij} is denoted as $most_{ij}$.

$$most_{ij} = \frac{z_{ij}}{\sum_{y=1}^{n_{p_j}} z_{yj}}, \tag{4}$$

where, $most_{ij} \in (0, 1)$ and the denominator is a summation defined by y and y =1 to n_{p_j}.

For example, in Fig. 1, customer u_1 purchases product p_1 and post feedback f_{11}. Now we evaluate the currentness score $most_{11}$. In that figure, $n_{p_1} = 5$; means 5 customers purchase p_1 and $n_{p_1} - 1 = 4$ is the total number of customers who buy the product after u_1. Before purchasing product p_1, customer u_2 may read previous reviews regarding p_1 and for u_2, f_{11} is the most recent review. From u_2, the contribution towards the currentness score of $f_{11} = (1/1)^2$. Similarly, before purchasing product p_1, customer u_3 may read previous reviews f_{11} and f_{21} written by u_1 and u_2 respectively. Note that, for u_3, f_{11} is the second most recent review (f_{21} is the most recent). So, from u_3, the contribution towards the currentness score of $f_{11} = (1/2)^2$. Finally we have $z_{1,1} = \sum_{s=1}^{4} \frac{1}{s^2} = (1/1^2) + (1/2^2) + (1/3^2) + (1/4^2)$. We assign more weight to the most current reviewers regarding the same product. Note that, $most_{ij}$ is a monotonically increasing function over time. However, the increment is diminishing quadratically.

Based on the helpfulness score we evaluate top ranking score top_{ij} of review f_{ij}. First we evaluate non-normalized top ranking score q_{ij} as follows:

$$q_{ij} = (\frac{1}{g_{ij}^2}) * (n_{p_j} - i). \tag{5}$$

Here, g_{ij} denotes the ranking of f_{ij} based on h_{ij} value among all the reviews on product p_j. Highest $h_{ij} \in$ (0,1) value for p_j, means $g_{ij} = 1$. Top rank depends on high helpfulness score. We assign more weight to the top ranking reviews regarding the same product. Low ranking review means more low q_{ij} score and this diminishing effect is represented by g_{ij}^2 in the denominator. So the denominator in Eq. 5 is quadratic of the rank. The normalized q_{ij} (i value range is 1 to n_{p_j}) is denoted as top_{ij}.

$$top_{ij} = \frac{q_{ij}}{\sum_{y=1}^{n_{p_j}} q_{yj}}, \tag{6}$$

where, $top_{ij} \in (0, 1)$.

Based on the top ranking and most current reviews, we evaluate overall ranking score d_{ij} as follows:

$$d_{ij} = \alpha * top_{ij} + (1 - \alpha) * most_{ij}, \tag{7}$$

where α defines the weights of the two components. We take $\alpha = 0.5$, because we want to give same priority to top_{ij} and $most_{ij}$. This value should be decided by the management of the company. The value of $d_{ij} \in (0, 1)$.

3.1.3 Clustering coefficient with review quality

In this subsection, we evaluate a customer's reliability and influential power in the customer's review network based on clustering coefficient concept (Newman, 2018). We assume that there are many customers who give more importance to good quality reviews rather than to the top ranking reviews. So, it is important to evaluate each review's quality along with its writer's influential power.

We evaluate the local clustering coefficient of customer u_i as

$$cc_i = \frac{|\theta_i|}{|\Theta_i|}, \tag{8}$$

where $|\theta_i|$ denotes the number of pairs of u_i's neighbors that are connected and $|\Theta_i|$ denotes number of pairs of customer u_i's neighbors. It is the fraction of pairs of neighbors of u_i that are themselves neighbors.

We calculate clustering coefficient on the one-mode projection of review network. For example, Fig. 2 presents the one-mode projection of the review network from Fig. 1. It is a directed network based on the chronological order of purchases done by the customers. For example u_5, u_6 and u_7 are the three neighbours of u_4 (we only consider the out degree neighbors of u_4 as shown in Fig. 2). Essentially, the count of neighbors of u_4 = the total count of customers who purchase products after u_4. So clustering coefficient of customer u_4 is $cc_4 = 2/3$. Next, we combine clustering coefficient of a customer with the quality of the customer's posted review.

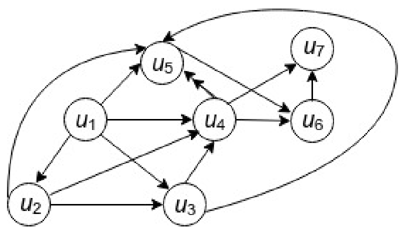

Fig. 2: One mode projection of Fig. 1. It is a directed graph based on customers' purchasing time stamp. $u_1 \rightarrow u_2$ indicates u_2 buys and posts review after u_1 for product p_1. $u_4 \twoheadrightarrow u_5$ indicates u_5 buys and posts one review after u_4 for p_1 and buys and posts another review after u_4 for another product p_2.

The quality of a review is evaluated based on spelling error score and readability of review.

Spelling error: To identify misspelled words, **pyspellchecker**[1] is used. Non-English words presented in the feedback are discarded. The spelling error score (ss_{ij}) of feedback f_{ij} regarding p_j posted by u_i is evaluated as follows:

$$ss_{ij} = \frac{|m^*_{ij}|}{|l^*_{ij}|}. \tag{9}$$

Here, $|m^*_{ij}|$ denotes the number of misspelled words of f_{ij} and $|l^*_{ij}|$ denotes the length of feedback f_{ij} (in words).

Readability: Using a set of metrics *i.e.*, Automated Readability Index, SMOG, Gunning fog index, Flesch-Kincaid Grade level, readability of a feedback can be computed. These metrics indicates how easy to read a feedback

[1] https://pyspellchecker.readthedocs.io/en/latest/.

for customers. In this work, we use Flesch-Kincaid Grade level from (DuBay, 2004). Flesch-Kincaid Grade level score of f_{ij} is denoted as FK_{ij}. The following formula of FK_{ij} is mentioned in (DuBay, 2004):

$$FK_{ij} = 0.39(\frac{|w^*_{ij}|}{|s^*_{ij}|}) + 11.8(\frac{|sy^*_{ij}|}{|w^*_{ij}|}) - 15.59. \tag{10}$$

Here, $|w^*_{ij}|$ denotes the total words of f_{ij}, $|s^*_{ij}|$ denotes the total sentences of f_{ij}, $|sy^*_{ij}|$ denotes the total syllables of f_{ij}. The selection of parameter values are shrewd guess based on the experiments to find out what works well. We scale $FK_{ij} \in (0, 1)$.

Quality of the review f_{ij} is denoted as $rq_{ij} = FK_{ij}$ - ss_{ij}. The value of rq_{ij} is scaled $\in (0, 1)$.

Clustering Coefficient with Review Quality of a review f_{ij} is evaluated as follows:

$$cq_{ij} = cc_i * rq_{ij}. \tag{11}$$

Here, $cq_{ij} \in (0, 1)$.

3.1.4 Centrality with review quality

Here, we evaluate each customer's influential power and position based on the concept of centrality in network (Newman, 2018). Network scientists have proposed different types of centrality. In this work, we use eigenvector centrality which is popularly known as *PageRank* centrality (Newman, 2018) due to its practicality and wide applications.

The eigenvector centrality is a more advanced concept of centrality. A person with few links could have a very high eigenvector centrality if those few links are connected with other highly connected persons. Eigenvector centrality permits connections to have a variable value, so that linking to some vertices has more gain than linking to others. Google's search engine uses the PageRank algorithm that is a variant of Eigenvector Centrality (Hansen et al., 2020).

Though Google uses this algorithm to rank different sites, we use it to evaluate customers' positional influence in a directed network. To evaluate the page-rank centrality of customer u_1, we have to consider: **(i) the number of out-bound customers of u_1 (out degree of u_1 indicates this number),** (ii) the out degree of out-bound customers of u_1, (iii) the page-rank centrality of out-bound customers of u_1 and we have to apply Eq. 14. In our work, we consider out-bound customers, because the network as shown in Fig. 2 is formed based on time stamp and $u_1 \rightarrow u_2$ indicates u_2 buys and posts review after u_1 for product p_1 and u_2 may read review of u_1 and may be influenced by u_1.

Page-rank centrality c_i of customer u_i is evaluated as follows:

$$c_i = \gamma \sum_y M_{iy} \frac{c_y}{out_y} + \varsigma, \tag{12}$$

where γ and ς are positive constants. M_{iy} denotes elements of corresponding adjacency matrix of the directed network (as shown in Fig. 2). The out degree of customer u_y is denoted as out_y. Here, out_y is the out-degree of customer u_y if such degree is positive, or $out_y = 1$ if the out-degree of customer u_y is null. In matrix terms, Eq. 12 is written as

$$c = \gamma M D^{-1} c + \varsigma * \mathbf{1}, \tag{13}$$

with $\mathbf{1}$ being the vector (1,1,1,...) and $\mathbf{D}$ being the diagonal matrix with elements D_{ii}=max(out_i, 1). By rearranging the above equation can be written as

$$c = \varsigma(\mathbf{I} - \gamma M D^{-1})^{-1} * \mathbf{1}, \tag{14}$$

where $\gamma = 0.85$ and $\varsigma = 1$ and there is no rigorous theory behind this value choice (Newman, 2018). Most likely it is just a shrewd guess based on the experiments to find out what works well. The identity matrix is denoted by $\mathbf{I}$.

Centrality with review quality of review f_{ij} is evaluated as follows:

$$cr_{ij} = c_i * rq_{ij}, \tag{15}$$

where, quality of the review is denoted as rq_{ij}. Here, $cr_{ij} \in (0, 1)$.

3.1.5 Calculation of neps

Depending on the structural properties of the Review Network, we evaluate h_{ij}, d_{ij}, cq_{ij}, and cr_{ij} for each review. We define **reliability score** rel_{ij} of review f_{ij} as the average of those four scores:

$$rel_{ij} = \frac{h_{ij} + d_{ij} + cq_{ij} + cr_{ij}}{4}, \tag{16}$$

where the value of $rel_{ij} \in (0, 1)$.

The Network Promoter Score $neps_{ij}$ of review f_{ij} is given as:

$$neps_{ij} = rel_{ij} * r_{ij}. \tag{17}$$

Here r_{ij} is scaled in from -2 to $+2$. So, the *neps* value is always between -2 to $+2$. Here negative sign indicates that the customer posts negative feedback and positive sign indicates that the customer posts positive feedback. We follow this range just to emphasis the positivity or negativity of the customer by the sign of the score.

From our observation we define that if $\mid neps_{ij} \mid > 0.5$, then customer u_i is a reliable reviewer for product p_j and review f_{ij} is a reliable review. The *neps* value higher than 0.5 means that the customer is a positive and reliable reviewer. Similarly, the *neps* value lower than -0.5 means that the customer is a negative and reliable reviewer. Monthly average of network promoter score(*neps*) for a particular product p_j over all reviews it got in t^{th} month is denoted as $NePS_{jt}$. First, we define average of *neps* for p_j over all reviews it got till t^{th} month as:

$$NePS_j^t = \frac{1}{n'} \sum_{i=1}^{n'} neps_{ij}, \tag{18}$$

where n' is the number of reviews posted for p_j till t^{th} month. So, we have

$$NePS_{jt} = NePS_j^t - NePS_j^{t-1}, \tag{19}$$

where $NePS_{jt}$ denotes $NePS$ of t^{th} month. The other parameters used in the existing literature are highlighted in the following subsections.

3.2 Average Rating (AR)

Average rating is an independent parameter. Customers' rating spans of various E-commerce sites are different. We scale customers' ratings from -2 to $+2$ range. For the sparsity of our experimental dataset, we assume that there is no biasness in the dataset. We average rating as:

$$Average\ rating(AR_j) = \frac{1}{n'} \sum_{i=1}^{n'} r_{ij}. \tag{20}$$

3.3 Average Sentiment (AS)

Average sentiment is an independent parameter. Sentiment score of a review text is evaluated in many existing works and relation between sentiment score of reviews and product sales is well studied topic. We identify which review is having a positive sentiment and which one is having a negative sentiment. In (Archak et al., 2011; B. Liu et al., 2010), the authors used feature based sentiment analysis and estimated impact on sales. We have applied the approach (Hu et al., 2014) on the review texts to evaluate sentiment score. As our rating ranges in -2 to $+2$, we scale sentiment score in -2 to $+2$. Further we aggregate sentiment score of all the reviews on a product and compute average score. The expression is given as:

$$Average\ sentiment\ score(AS_j) = \frac{1}{n'} \sum_{i=1}^{n'} sentiment_{ij}, \tag{21}$$

where, $sentiment_{ij}$ denotes the sentiment score of review f_{ij}.

3.4 RS Value

It is an independent parameter. In our work, if $r_{ij} \geq 3$, then customer u_i is considered as a promoter for product p_j. If $r_{ij} < 3$, then customer u_i is considered as a detractor for product p_j. Here, $RS_j =$ %Promoter- %Detractor for product p_j.

3.5 Other parameters

We consider other independent parameters, such as, average quality of reviews (it is discussed in Subsec. 3.1.3), total number of reviews, variance of rating and variance of sentiment. We have also considered two control parameters, price and age of products.

4 Methodology

In this section, we demonstrate an approach to evaluate the importance of each parameter (*i.e.*, *NePS*, age, review volume, price, review quality, avg. sentiment and avg. ratings) on product sales and investigate on some hypotheses. Keeping all these parameters in mind, the following hypotheses are experimented and analyzed in Sec 6. We are interested to identify the parameter which has strongest regression with product sales.
Hypothesis H1: ***An increase in the proportion of the NePS will be positively related to product sales.*** **This hypothesis is tested based on our independent parameter *NePS*. Hypothesis H1 is tested and compared with other existing hypotheses**.

Based on the discussion of related literature in Sec. 2, the existing hypotheses are as follows:
Hypothesis H2: *Product sales increase as average quality of reviews increases.*
Hypothesis H3: *An increase in the proportion of RS score (%promoters-%detractors) will be definitely related to product sales.*
Hypothesis H4: *Product sales increase as average sentiment scores of customers increase.*
Hypothesis H5: *Sales of a product increase as the number of reviews of the particular product increases.*
Hypothesis H6: *Product Sales decrease as price increases.*

Throughout our experiments, it is found that the regression of the *NePS* on product sales is higher than the other existing parameters (discussed in Sec 6). The effects of different parameters (based on our consideration) on product sales are not immediate. The latency time (how much time will take to affect product sales) is varied from parameter to parameter and also product to product.

4.1 Effect on Product sales and Latency Time

First, we estimate the impact of rating score, sentiment value of customers' review texts, quality of review and other parameters on product sales. Now, the question is how we evaluate the product actual sale. Sales rank information is available in the dataset but product sales information is not publicly available. It is not possible to get actual product sales information, because the merchandise companies do not allow to publish this information in dataset. So, based on the concept of (J. Chevalier & Goolsbee, 2003), we assume that product sales rank is associated with the demand levels of products. We assume that product demands and product sales are directly proportional. The product demand in terms of product sales ranking follow a Pareto distribution (i.e., a power-law). Using the following Pareto relationship (J. Chevalier & Goolsbee, 2003; Rossi et al., 2001), product sales rank is converted into demand levels and the expression is as follows:

$$ln(D_{jt}) = a + b * ln(SR_{jt}), \tag{22}$$

where, D_{jt} is the unobserved item demand of product p_j in t^{th} month that indicates product sales of the product in t^{th} month, SR_{jt} is the observed sales rank of product p_j in t^{th} month, and $a > 0$, $b < 0$ are industry specific parameters. In (Brynjolfsson et al., 2003), the authors approximated the above relation by choosing $a = 9.61$ and $b =$ - 0.78. The selection of parameters value is just a shrewd guess based on the experiments to find out what works well. We apply the same values.

How item price, review valence and number of reviews affect product sales is discussed in (J. A. Chevalier & Mayzlin, 2006). In (Hu et al., 2014; Ghose & Ipeirotis, 2011), the authors have focused on sentiment score of review, quality of review, readability and investigated on how these factors affect product sales. To maintain a consistency with these previous works, we consider all these parameters including the proposed parameter *NePS* on a particular set of product categories.

Model Specification: The monthly average sales rank (SR_{jt}) of product p_j in t^{th} month is calculated as follows:

$$SR_{jt} = \frac{1}{n^{p_j}} \sum_{i=1}^{n^{p_j}} SR_{ijt}, \tag{23}$$

where, SR_{ijt} indicates the sales rank of p_j when customer u_i purchases it in t^{th} month. Here, the dependent parameter $ln(SR)_{jt}$ ('ln' means '*log base e*') of sales rank of item p_j at t time) is a linear transformation of the ln of product market demand, as talked about before in Eq. 22. Initially, we assume that the latency time is of one month, which means that considered parameters start their effects on product sales after one month. Later in Sec 6, latency time is varied and we observe the effectiveness of different parameters on product sales based on different latency time.

To measure the impact of the *NePS* and other parameters on product sales, we use a structural equation modeling technique. It is a multivariate statistical analysis technique. It is used to analyze structural relationships. This technique is the combination of factor analysis and **multiple regression analysis (Allison, 1999)** (Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a parameter based on the values of two or more other parameters. The parameter that we want to predict is called the dependent parameter (or sometimes the outcome, target or criterion parameter). The parameters that we are using to predict the value of the dependent parameter are called the independent parameters). The model is proposed as follows:

$$\begin{aligned} ln(SR)_{jt} = \alpha_0 + \alpha_1(AR)_{j(t-1)} + \alpha_2(AS)_{j(t-1)} + \alpha_3 ln(RS)_{j(t-1)} \\ +\alpha_4(NePS)_{j(t-1)} + \alpha_5(AQR)_{j(t-1)} + \alpha_6 ln(PI)_{j(t-1)} + \alpha_7 ln(AI)_{j(t-1)} \\ +\alpha_8 ln(TR)_{j(t-1)} + \alpha_9(VR)_{j(t-1)} + \alpha_{10}(VS)_{j(t-1)} + \mu_j + \varepsilon_{jt}, \end{aligned} \tag{24}$$

where
t: index of month.
j: index of product.
$(SR)_{jt}$: sales rank of product p_j at month t.
$(AR)_{j(t-1)}$: avg. rating of p_j at month $t-1$.
$(AS)_{j(t-1)}$: avg. sentiment score of p_j at month $t-1$.
$(RS)_{j(t-1)}$: %promoter - %detractor of p_j at month $t-1$.
$(NePS)_{j(t-1)}$: *NePS* of p_j at month $t-1$.
$(AQR)_{j(t-1)}$: avg. quality of review of p_j at month $t-1$.
$(PI)_{j(t-1)}$: price of p_j at month $t-1$.
$(AI)_{j(t-1)}$: age of p_j at month $t-1$.
$(TR)_{j(t-1)}$: total reviews of p_j at month $t-1$.
$(VR)_{j(t-1)}$: variance of the rating of p_j at month $t-1$.
$(VS)_{j(t-1)}$: variance of the sentiment score of review of p_j at month $t-1$.
α_0 to α_{10} : estimated regression coefficients of the above mentioned parameters.
μ_j: product fixed effect that accounts for unobserved heterogeneity across products.
ε_{jt}: residual error.

Latency: The impact of different parameters on product sales comes after a latency period. Point to be noted that this latency period also depends on data, season, brand offers, marketing strategy that are not considered in this work. The latency times of different parameters are evaluated using our proposed model as mentioned in Eq. 24 and the results are presented in Sec 6. In Eq. 24, $(SR)_{jt}$ is the dependent parameter and the other parameters are independent. In Eq. 24, we take ln scale of some of the parameters to provide non-linearity and flexibility to large values. Our model is reliable, because it supports all proper specification of multiple regression. i) The dependent variable is measured on a continuous scale. ii) We have more than one independent variables. iii) The model is linear in nature. iv) Homoscedasticity must be assumed; the variance is constant across all levels of the predicted variable. v) Residuals (errors) are approximately normally distributed.

4.2 Predictive Modeling

In this subsection, we construct a deep learning model and examine whether our model can be used to predict future product sales after a certain period (latency time) based on the current *NePS*. So, we examine whether the difference

$$SR_{j(t+1)} - SR_{jt}, \tag{25}$$

is positive or not. Positive difference means the value of sales rank has increased (product sales decreases). Negative difference means product sale has increased. We design a Long Short Term Memory Model (*LSTM*) that is capable of understanding the trading patterns of sales rank based on the *NePS*.

Model Specification: How *LSTM* model learns customers' *NePS* and predict its impact on product sales, is discussed in this subsection. *LSTM*, a variant of Recurrent Neural Network (*RNN*), is discussed in details in (Goodfellow et al., 2016). It is the standard architecture focusing on sigmiod and tanh function[2]. For activation functions, one can freely choose sigmoid, hyperbolic tangent (tanh), and Rectifier (ReLU), among others (Goodfellow et al., 2016). Specifically, sigmoid is used as the gating function for the 3 gates (in, out, forget) in *LSTM*. The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through. A value of zero means "let nothing through," while a value of one means "let everything through." On the other hand, to overcome the vanishing gradient problem we need a function whose second derivative can sustain for a long range before going to zero. Tanh is a good function with the above property and this choice has been widely adopted in many studies (Wu et al., 2016; Elkahky et al., 2015; Goodfellow et al., 2016).

The hidden states of this *RNN* unfold themselves over time, as shown in Fig. 3. When Back-Propagation Through Time (*BPTT*) (Werbos, 1990) is used to train *RNN*, it faces problem with disappearing gradients and fails to overcome long term dependency in a particular time period. Additional gating units in *LSTM* overcome the above problem using the long term memory to understand the product sales' pattern from the *NePS* and sales rank information. Three types of gating units are used to understand product sales' patterns from the *NePS* and sales rank information, using a memory cell at each iteration t, where t indicates month. Input gate (i_t) allows the amount of new information generating into memory cell. The forget gate (f_t) decides which portion of the information should be regulated in the cell and the output gate (o_t) generates the output patterns as shown in Fig. 4.

The *LSTM* can be described as follows: At each iteration t (t indicates t^{th} month), x_t is an input vector that is the *NePS* of item p_j at time t, c_t indicates the memory state vector and h_t is the hidden state vector output from c_t. The mathematical architecture (Goodfellow et al., 2016) is as follows:

$$i_t = sigmoid(W_i x_t + U_i h_{t-1} + b_i) \tag{26}$$

$$f_t = sigmoid(W_f x_t + U_f h_{t-1} + b_f) \tag{27}$$

$$\tilde{c_t} = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{28}$$

[2] http://colah.github.io/posts/2015-08-Understanding-LSTMs/

$$c_t = i_t \circ \tilde{c_t} + f_t \circ c_{t-1} \tag{29}$$

$$o_t = sigmoid(W_o x_t + U_o h_{t-1} + b_o) \tag{30}$$

$$h_t = o_t \circ tanh(c_t), \tag{31}$$

where W_* and U_* indicate weight matrices, b_* denotes bias vectors and we have to initialize it at the first iteration. Sigmoid function sigmoid (.) is used as the activation function for the type of gates. The input modulation $\tilde{c_t}$ and output h_t usually apply the tanh(.) as the activation functions and "$\circ$" refers point wise multiplication.

Regression and Prediction: Predicted sales rank ($\hat{SR}_{t+1}$) is obtained via a regression layer (L. Zhang et al., 2017),

$$\hat{SR}_{t+1} = W_{sr} h_t + b_{sr}, \tag{32}$$

where W_{sr} and b_{sr} are the weight matrix and bias vector, respectively. Each output of training dataset of our model is trained with the original target output (SR_{t+1}) with standard back-propagation technique. For our model, we use *Adam* optimizer, which uses the learning rate for each parameters by executing smaller updates for frequent and larger updates for infrequent parameters and Mean Absolute Error as loss function after experimental observation. Then we examine whether the difference

$$\hat{SR}_{j(t+1)} - \hat{SR}_{jt}, \tag{33}$$

is positive or not.

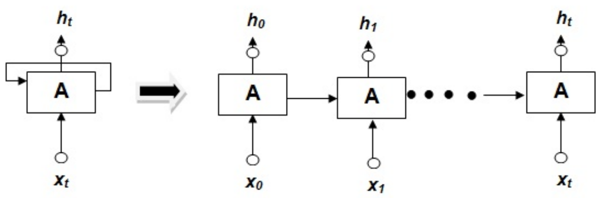

Fig. 3: Recurrent Neural Network Architecture.

5 Evaluation Procedure

Data Statistics: To conduct our experiment, we choose publicly available Amazon.com online review dataset (He & McAuley, 2016; McAuley et al.,

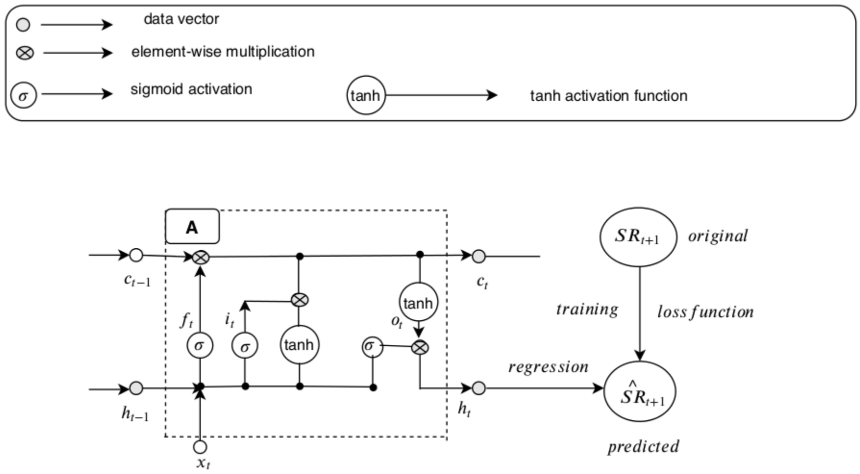

Fig. 4: LSTM Architecture.

2015; Ni et al., 2019)[3]. We perform our experiments on the Home & Kitchen, Digital Camera, and Cell Phones & Accessories category datasets from January 2015 to December 2017. The statistics of the dataset are shown in Tables 2,3,4. This data set provides all the required information to construct the review network. Helpfulness score is calculated from "*helpful*" attribute in the dataset. For example, "*helpful*": $[2, 3]$ indicates two customers are helpful from this review and one customer thinks it is not helpful.

The proposed model in Eq. 24 is used to investigate the effectiveness of the *NePS* and other existing parameters on product sales. For this investigation, we have used the entire dataset, because there is nothing to predict and all information are either available in the dataset (*e.g.*, price, sales rank, age) or evaluated (avg. sentiment, avg. ratings, review quality and *NePS*). The empirical observation of this proposed methodology is performed on *STATA* (*https://www.stata.com/*), a statistical software for data science.

We also construct a deep learning model and examine whether our model can be used to predict future product sales after a certain period (latency time) based on the current *NePS*. To compute the efficiency of *LSTM* model based on the considered datasets with different training size, the five-fold cross validation is applied. As the prediction is time-dependent, cross-validation is on a rolling basis. We start with a small subset of data for training and validation purpose, predict for the later data points and then check the accuracy for the predicted data points. In the next fold, the predicted data points and validation points are then included as part of the next training dataset as shown in Fig. 5.

In this figure, we have only 5 observations in our cross-validation set and we want to perform 5-fold cross-validation. Here in the last fold of five-fold

[3] http://jmcauley.ucsd.edu/data/amazon/
https://nijianmo.github.io/amazon/

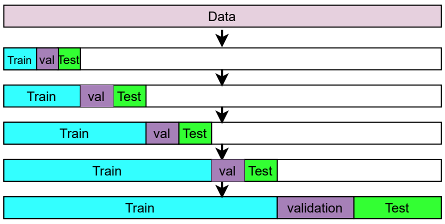

Fig. 5: Five cross-validation is on a rolling basis.

cross validation, 70% of data is selected for training purpose (January 2015 to January 2017), and 15% of our dataset is selected for validation purpose and the remaining is selected for testing phase. Compute the average of the accuracy of the 5 test fold. Actually we follow this strategy to observe our model's performance on small training dataset and we have applied this strategy on baselines also. We have implemented the model on $Keras$ platform. The implementation of this model is publicly available[4].

Baseline: In our work, we perform our investigation on the regression between the *NePS* and product sales. Since we are the first to introduce the *NePS*, there is no baseline that considers this parameter or somewhat similar concept. For comparison purpose, we choose other five baselines that consider other existing parameters *i.e.*, sentiment of review texts, review quality, numeric ratings and number of reviews. Essentially, we compare the effectiveness of the *NePS* with that of the other parameters. The baselines are as follows:

- Reichheld (2003): The difference of the number of promoters and detractors is evaluated and analyze how does it affect product sales. The authors use this approach on offline WOM. But in our experiment, we apply this concept on online review dataset.
- Ghose & Ipeirotis (2011): The authors focus on how subjective statements, readability score and spelling errors in reviews are related to product sale.
- Hu et al. (2014): The authors focus on how sentiment of reviews has direct impact on product sales over ratings.
- Maslowska et al. (2017): The authors focus on the effect of purchasing probability when there are many reviews, the customers read reviews and the product is high priced.
- Filieri et al. (2018): The authors evaluate helpful votes in $eWOM$ contexts and predicts purchase intention.
- X. Li et al. (2019): The findings not only contribute to the knowledge of how $eWOM$ impacts product sales, but also illustrate how numerical rating and textual reviews interplay in shaping product sales. Please note that, online consumer reviews are treated as $eWOM$ (Electronic word-of-mouth).

For the comparison purpose, baseline approaches are applied on the same Amazon.com online review dataset. The same experimental settings as men-

[4] https://github.com/SUPRIYOPHD/An-Indicator-of-Product-Sales

tioned in (Reichheld, 2003; Ghose & Ipeirotis, 2011; Hu et al., 2014; Maslowska et al., 2017; Filieri et al., 2018; X. Li et al., 2019) are followed.

Table 2: Descriptive statistics of Home & Kitchen dataset from Amazon.com for econometric analysis from January 2015 to December 2017.

Parameter	2015			2016			2017		
	Median	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**
Sales Rank	2254.42	3921.12	4350.91	2753.34	3122.94	3957.55	2188 .75	2997.71	4176.09
Average rating(-2 to +2 scale)	0.08	1.03	2.97	0.17	1.18	2.54	0.02	0.97	2.17
Average sentiment (-2 to +2 scale)	0.07	1.17	2.30	0.21	1.73	2.91	0.04	1.32	2.92
RS	39.21	43.45	51.67	41.39	39.11	41.23	29.91	37.91	42.01
NePS	0.93	1.01	0.54	1.02	1.17	0.37	1.10	1.31	0.67
Avg. quality of review	9.78	13.09	25.55	10.23	14.01	31.23	11.43	14.11	8.34
Price (US dollars)	16.26	18.19	22.98	18.08	21.08	23.75	11.76	14.20	21.02
Age (in days)	122.76	322.72	577.02	167.93	421.34	567.92	301.52	511.71	411.81
Total reviews	199.34	273.56	457.09	178.03	294.45	477.03	211.72	251.67	571.04
Variance of rating	1.22	1.54	0.65	0.98	1.33	0.75	1.65	1.22	0.49
Variance of sentiment	0.63	0.77	0.31	0.61	0.65	0.29	0.71	0.62	0.33
Sample size	**8,892**			**11,711**			**13,189**		

Table 3: Descriptive statistics of Digital Camera dataset from Amazon.com for econometric analysis from January 2015 to December 2017.

Parameter	2015			2016			2017		
	Median	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**
Sales Rank	1288.54	1988.34	2845.12	1219.03	1987.34	3377.23	1922.93	2397.29	4165.34
Average rating(-2 to +2 scale)	-0.13	0.97	1.99	0.09	1.38	1.83	-0.06	1.06	1.05
Average sentiment (-2 to +2 scale)	-0.17	0.81	1.02	0.21	1.73	1.91	-0.04	1.32	1.92
RS	39.21	43.45	51.67	41.39	39.11	41.23	29.91	37.91	42.01
NePS	0.88	1.10	0.66	0.90	1.21	0.56	0.95	1.27	0.63
Avg. quality of review	10.23	17.10	35.67	10.23	15.12	33.45	8.79	12.87	18.75
Price (US dollars)	97.12	165.02	388.43	101.22	167.32	378.90	99.01	187.59	353.11
Age (in days)	122.76	322.72	577.02	167.93	421.34	567.92	301.52	511.71	411.81
Total reviews	178.21	255.23	411.33	170.61	288.34	429.71	215.91	201.39	512.34
Variance of rating	1.12	1.33	0.35	0.91	1.01	0.43	1.33	1.01	0.22
Variance of sentiment	0.53	0.63	0.29	0.53	0.61	0.21	0.66	0.69	0.35
Sample size	**4,117**			**5,429**			**6,094**		

Table 4: Descriptive statistics of Cell Phones & Accessories datset from Amazon.com for econometric analysis from January 2015 to December 2017.

Parameter	2015			2016			2017		
	Median	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**	**Median**	**Mean**	**Std. Dev.**
Sales Rank	6178.67	7599.81	11873.06	5122.87	7119.67	12331.98	3499.12	4993.15	9827.71
Average rating(-2 to +2 scale)	0.07	0.83	1.03	0.03	1.17	1.02	0.11	0.97	0.83
Average sentiment (-2 to +2 scale)	0.11	0.73	0.99	0.11	1.45	1.04	0.03	1.19	0.79
RS	40.11	53.32	33.87	41.34	55.67	32.45	29.91	42.78	29.34
NePS	0.73	1.02	0.63	0.93	1.37	0.67	0.89	1.19	0.70
Avg. quality of review	9.31	17.45	21.45	11.72	17.22	31.76	8.23	15.12	11.54
Price (US dollars)	239.37	395.34	348.71	294.91	343.19	381.23	211.23	387.12	390.23
Age (in days)	117.34	351.34	611.34	133.45	411.34	601.38	401.23	511.71	221.11
Total reviews	252.89	311.45	311.45	288.45	323.67	282.67	317.34	465.01	353.93
Variance of rating	0.99	1.21	0.93	1.01	1.43	0.65	1.23	1.02	0.66
Variance of sentiment	0.58	0.71	0.43	0.67	0.71	0.39	0.61	0.67	0.23
Sample size	**13,174**			**17,523**			**21,284**		

6 Empirical Result

In this section, we demonstrate the results of our investigation on the regression between each parameter and product sales and test our hypotheses (mentioned in Sec. 4). Our results show which parameters are important to understand the strong regression with product sales. The performances of our approach based on the *NePS* is compared with baselines. In this section, we also investigate on latency time of different product categories based on different parameters. We have also presented the performances of our predicting model based on products' *NePS* values and compared with baselines.

6.1 Regression between the Parameters and Product Sale

Our proposed approach estimates the regression between the parameters and product sales using **Two Stage Least Squares (2SLS)** with product level fixed effects. The empirical observation is performed on *STATA*. *STATA* fits fixed-effects (within), between-effects, and random-effects (mixed) models. The product level fixed effects mean that the parameters related to product are fixed or non-random quantities. The estimation of our model (mention in Eq. 24) is shown in Table 5. Here significance: ***p < .001; **p < .01; *p < .05. The level of statistical significance is often expressed as a p-value between 0 and 1. The smaller the p-value, the stronger the evidence that we should reject the null hypothesis. Most authors refer to statistically significant as $p < 0.05$ and statistically highly significant as $p < 0.001$ (less than one in a thousand chance of being wrong).

In Table 5, the negative sign of the coefficient of a parameter indicates that the value of sales rank of products decreases (means an increase in product sales for the products) when the value of the parameter increases. In our hypothesis, we have mentioned that sales rank decreases as the value of the *NePS* score increases. Similarly, sales rank decreases as the value of average quality of reviews, *RS* score (%promoters-%detractors), average sentiment scores, average rating scores and the number of reviews increase. On the other hand, sales rank increases as price, variance of rating and variance of sentiment increase. Based on our experiment, we investigate our hypotheses.

Hypothesis H1: This hypothesis is performed based on the independent parameter *NePS*. In Table 5, the coefficient of the *NePS* is negative, which indicates that an increase in the *NePS* leads to an increase in sales for products. In our experiment, it is very significant result that *Hypothesis H1* is supported in all the three categories , *i.e.*, Home & Kitchen, Digital Camera and Cell phones & Accessories.

We investigate why the *NePS* achieves a stronger regression with product sales in all the three categories compare to the other parameters. From our observation it is found that in the Home & Kitchen category dataset, for a particular time period the average quality of reviews increase but product sales decrease after a particular latency time period. For this contradictory result, we analyze each review very carefully and it is found that there are some negative reviews whose quality, reliability and helpfulness score are higher than some positive reviews. Some customers may be influenced by these negative reviews. For this reason product sales decrease after a time period. At this point, it is challenging to claim the effectiveness of the *NePS* on product sales and it is not clear whether the *NePS* influences sales. But practically a strong regression between the *NePS* and product sales is established. Point to be noted that data, season, brand offers, marketing strategy also affect product sales. We do not consider these factors in our current work but we would like to investigate on these factors in our future work.

We investigate the Digital Camera category and it is found that for a particular time period the number of reviews are increasing but after a time period product sales decrease. We analyze each review very carefully and it is found that the customers are purchasing this category item but many of them are not satisfied with the quality of items. In the online review portal they post negative reviews and other customers give helpful votes on those reviews. As a result, the reliability and helpfulness score of these negative reviews is higher than the reliability of posted positive reviews. With the influential effect of these negative feedback, product sales decrease after a time period.

We also analyze the Cellphones & Accessories category and it is found that in this category, for a particular time period the number of positive reviews is higher than the number of negative reviews but product sales decrease after some time period. The reason is that the effectiveness of these negative reviews are higher than positive reviews and customers may be influenced by these negative reviews and do not purchase the products. For this reason product sales decrease.

From this observation, our opinion is that product sales do not always depend on the average rating or sentiment score or avg. quality of reviews. The reliability of customers' reviews is also a crucial factor to establish a strong regression with product sales. Consequently, *Hypothesis H1* is supported in all the categories. For example in the Digital Camera category data, the impact of the *NePS* **(-0.356**)** on sales rank is higher than that of avg. rating (-0.016*), avg. sentiment score (-0.033*), *RS* value (-0.193*) and avg. quality of reviews (-0.153**) on sales rank.

Hypothesis H2: Table 5 reports the regression between avg. quality of reviews and product sales. It is clear that *Hypothesis H2* is supported in the Digital Camera and Cell Phones & Accessories categories. Point to be noted that *NePS* achieves more strong regression compare to avg. quality of reviews with product sales.

For example, Digital Camera category, impact of *NePS* **(-0.356**)** on sales rank is higher than impact of avg. quality of review (-0.153**). The regression between avg. quality of review and sales rank is statistically insignificant for the Home & Kitchen category. So, *Hypothesis H2* is not supported in this case. From our observation it is found that in Home & Kitchen category dataset, for a particular time period the average quality of reviews increase but product sales decrease after a particular latency time period. For this contradictory result, we analyze each review very carefully and it is found that there are some negative reviews whose quality, reliability and helpfulness score are higher than some positive reviews. The effectiveness of these negative reviews are higher than positive reviews and may be other customers are influenced by these negative reviews and do not purchase items. For this contradictory result, the reason is that the customers are purchasing this category item but many of them are not satisfied with the quality of items. In online review portal they post negative reviews with good quality and other customers give helpful votes on these reviews. As a result the reliability and helpful score of these negative reviews is higher than the reliability of posted positive reviews. With the influential effect of these negative feedback, the product sales decrease after a latency time period. From this observation, our opinion is that product sales do not always depend on avg quality of review.

Hypothesis H3: The regression between *RS* value and sales rank is statistically significant for the Digital Camera category and *Hypothesis H3* is supported only for this category as shown in Table 5. But, the impact of *RS* value is statistically insignificant for the Cell phones & Accessories and Home & Kitchen categories.

Hypothesis H4: Table 5 shows that average sentiment of reviews is statistically significant for the Digital Camera category and *Hypothesis H4* is supported only for this category. But, it is statistically insignificant for the Cell Phones & Accessories and Home & Kitchen and *Hypothesis H4* categories.

Hypothesis H5: *Hypothesis H5* is supported for none of the three product categories. From our observation, we find that product sales do not depend on the number of reviews.

Hypothesis H6: In Table 5, the positive sign of the coefficient of the parameter price indicates that the value of sales rank of products increases (means decrease in sales for the products) when the price of products increases. *Hypothesis H6* is supported for none of the three product categories. The summary of all the hypotheses is presented in Table 6.

Comparison with the Baselines: After testing the hypotheses, it is cleared that there is a strong regression between the *NePS* and product sales and this parameter achieves the best regression compare to all the other parameters. In Table 7, we have presented the performances of our approach based on the *NePS* and it is compared with the other baselines. From Table 7, it is clear that product sales do not always depend on the existing parameters which are considered in baselines (Reichheld, 2003; Ghose & Ipeirotis, 2011; Hu et al., 2014; Maslowska et al., 2017; X. Li et al., 2019). Why product sales do not always depend on these parameters is already discussed in hypotheses discussion part.

6.2 Latency Time

The effect of different parameters on product sales based on different latency time period is shown in Tables 8, 9, 10. Here, @1month indicates the effect of a parameter on product sales after one month latency period. For example, we consider the value of the parameter for January 2015 and consider the value of product sales for February 2015.

For example, in Table 9, the effect of different parameters on the product sales of the Digital Camera category based on the different latency periods is observed. The best effect of avg. rating is found in @3 month latency period

Table 5: The coefficient values from the Two Stage Least Squares (2SLS) Regressions with Product-Level Fixed Effects based on our proposed model (mention in Eq. 24). The dependent parameter is ln(Sales Rank) depends on the independent parameters $i.e$ average rating, avg. sentiment, ln(RS), $NePS$, ln(Total reviews), and control parameters $i.e$ ln(Price), ln(Age). We take ln scale of some parameters to provide non-linearity and flexibility to large values. In parenthesis robust standard errors are mentioned. Increases in the value of sales rank mean decreases product market sales. A negative coefficient of a parameter indicates that increases in a parameter lead to decreases in the value of sales rank of items that means increases in sales for items. Significance: ***$p < .001$; **$p < .01$; *$p < .05$. The R-square includes fixed effects in R-square computation. The highest effectiveness values are marked in bold.

Parameters	Home & Kitchen	Digital Camera	Cell Phones & Accessories
***NePS* (our parameter) (Hypothesis H1)**	**-0.297** (0.117)*	**-0.356** (0.012)**	**-0.379** (0.017)***
Avg. quality of review (Hypothesis H2)	0.027 (0.012)	-0.153 (0.021)**	-0.113 (0.005)*
ln(RS) (Hypothesis H3)	0.057(0.012)	-0.193(0.070)*	0.103(0.076)
Average sentiment (Hypothesis H4)	0.039 (0.030)	-0.033 (0.011)*	0.028 (0.128)
ln(Total reviews) (Hypothesis H5)	0.011 (0.101)	0.076 (0.016)	0.235 (0.021)
ln(Price) (Hypothesis H6)	-0.095(0.007)	-0.173(0.015)	-0.083(0.017)
Average rating	0.003 (0.007)	-0.016 (0.012)*	0.009 (0.011)
ln(Age(in days))	0.120(0.060)	0.105(0.033)	0.039(0.020)
Variance of rating	-0.123 (0.004)	0.118 (0.106)*	-0.087 (0.013)
Variance of sentiment	-0.016 (0.017)	0.021 (0.109)*	-0.072 (0.210)
R-square (with fixed effects)	0.81	0.88	0.92

Table 6: Summary of hypotheses. Hypothesis H1 is performed based on the $NePS$ and it is supported for all Categories. It is marked in bold.

Hypotheses	Home & Kitchen	Digital Camera	Cell Phones & Accessories
Hypothesis H1	**supported**	**supported**	**supported**
Hypothesis H2	not supported	supported	supported
Hypothesis H3	not supported	supported	not supported
Hypothesis H4	not supported	supported	not supported
Hypothesis H5	not supported	not supported	not supported
Hypothesis H6	not supported	not supported	not supported

Table 7: Two Stage Least Squares (2SLS) regressions with product-level fixed effects for baselines and our Model (mention in Eq. 22). In parenthesis robust standard errors are mentioned. Increases in the value of sales rank mean decreases product sales. A negative coefficient of a parameter indicates that increases in a parameter lead to decreases in the value of sales rank of items that means increases in sales for items. Significance: ***$p < .001$; **$p < .01$; *$p < .05$. The highest effectiveness values are marked in bold.

Model	Home & Kitchen	Digital Camera	Cell Phones & Accessories
(Reichheld, 2003)	0.057 (0.012)	-0.193 (0.070)*	0.103 (0.076)
(Ghose & Ipeirotis, 2011)	0.027 (0.012)	-0.153 (0.021)**	-0.113(0.005)*
(Hu et al., 2014)	0.057(0.127)	-0.043(0.070)*	0.163(0.076)
(Maslowska et al., 2017)	0.197 (0.136)	-0.096 (0.022)*	0.179 (0.120)
(Filieri et al., 2018)	-0.113 (0.167)*	-0.211 (0.021)**	-0.139 (0.027)*
(X. Li et al., 2019)	0.003 (0.107)	-0.050 (0.012)*	0.009 (0.111)
[our]	**-0.297** (0.102)*	**-0.356** (0.012)**	**-0.379** (0.017)***

on product sales. The same thing happens for avg. sentiment score. The $NePS$ is strongly supported on three month latency time period **(-0.356**)** compared to other latency periods. Similarly, in Table 8 and Table 10, the effect of different variables on Home & Kitchen and Cell Phones & Accessories product sales, respectively, based on different latency periods is observed. In Table 8, the $NePS$ is strongly supported on two month latency time period **(-0.297*)** compared to other latency periods for Home & Kitchen related product sale. In Table 10, the $NePS$ is strongly supported on two month latency time period **(-0.379**)** compared to other latency periods for Cell Phones & Accessories related product sale. The latency time period is not fixed. It varies from product to product depending on season, brand offers, marketing strategy. If a parameter supports a hypothesis, then the best regression is marked in bold.

Table 8: Fixed effect on product sales of Home & Kitchen category in our model (mention in Eq. 24) based on different latency time. We consider the value of the parameters from January, 2015 to December, 2017. If the parameter supports hypothesis, then highest regression for this parameter is marked in bold.

Parameter	**2SLS Regressions with Product level fixed effect based on different latency periods**						
	@1month	@2 month	@3 month	@4 month	@5 month	@6 month	@7 month
Average rating	0.007	0.003	0.010	0.209	0.112	0.115	0.023
Average sentiment	0.076	0.055	0.039	0.102	0.123	0.293	0.109
ln(RS)	0.144	0.163	0.057	0.057	0.193	0.188	0.201
NePS	-0.107	**-0.297***	-0.253*	-0.205*	0.102	0.119	0.277
Avg. quality of review	0.127	0.151	0.027	0.097	0.192	0.199	0.110
ln(Price)	-0.015	-0.095	-0.023	-0.007	-0.032	-0.119	-0.091
ln(Age(in days))	0.120	0.122	0.132	0.177	0.141	0.223	0.211
ln(Total reviews)	0.011	0.223	0.176	0.098	0.107	0.025	0.072
Variance of rating	-0.003	-0.123	-0.028	-0.109	-0.097	-0.019	-0.055
Variance of sentiment	-0.007	-0.016	-0.014	-0.002	-0.002	-0.010	-0.011

Table 9: Fixed effect on product sales of Digital Camera category in our model (mention in Eq. 24) based on different latency time. We consider the parameters from January, 2015 to December, 2017. If the parameter supports hypothesis, then highest regression for this parameter is marked in bold.

Parameter	**2SLS Regressions with Product level fixed effect based on different latency periods**						
	@1month	@2 month	@3 month	@4 month	@5 month	@6 month	@7 month
Average rating	0.016	0.003	**-0.016***	-0.009*	0.011	0.005	0.017
Average sentiment	0.013	0.041	**-0.033***	-0.010*	0.003	0.138	0.101
ln(RS)	0.093	0.066	**-0.193***	-0.010*	-0.076*	0.168	0.197
NePS	0.056	-0.210*	**-0.356****	-0.309*	-0.112	0.037	0.156
Avg. quality of review	-0.073*	**-0.153****	-0.113*	-0.017*	0.127	0.127	0.009
ln(Price)	-0.033	-0.173	-0.008	-0.019	-0.163	-0.109	-0.051
ln(Age(in days))	0.115	0.108	0.105	0.117	0.193	0.151	0.243
ln(Total reviews)	0.086	0.161	0.076	0.111	0.103	0.115	0.207
Variance of rating	-0.018	0.003	**0.118***	-0.013	-0.176	-0.291	-0.011
Variance of sentiment	-0.011	**0.021***	0.010	-0.007	-0.011	-0.121	-0.271

Table 10: Fixed effect on product sales of Cell Phones & Accessories category in our model (mention in Eq. 24) based on different latency time. We consider the value of the parameters from January, 2015 to December, 2017. If the parameter supports hypothesis, then highest regression for this parameter is marked in bold.

Parameter	**2SLS Regressions with Product level fixed effect based on different latency periods**						
	@1month	@2 month	@3 month	@4 month	@5 month	@6 month	@7 month
Average rating	0.019	0.009	0.111	0.101	0.183	0.213	0.217
Average sentiment	0.078	0.052	0.028	0.117	0.204	0.288	0.171
ln(RS)	0.133	0.197	0.211	0.103	0.233	0.147	0.107
NePS	-0.209*	**-0.379****	-0.205*	-0.191*	0.003	0.147	0.137
Avg. quality of review	0.013	-0.107*	**-0.113***	-0.109*	0.116	0.121	0.002
ln(Price)	-0.083	-0.033	-0.022	-0.017	-0.049	-0.006	-0.011
ln(Age(in days))	0.089	0.039	0.117	0.108	0.127	0.132	0.212
ln(Total reviews)	0.275	0.235	0.393	0.309	0.254	0.259	0.291
Variance of rating	-0.107	-0.118	-0.087	-0.095	-0.093	-0.182	-0.127
Variance of sentiment	-0.102	-0.072	-0.091	-0.121	-0.123	-0.176	-0.219

6.3 Performance of the Predictive Model

We observe a strong regression between the *NePS* and product sales. In this subsection, with the *NePS* as an input in *LSTM* model, we want to forecast whether product sales will go up or go down based on different latency periods. In Table 11, accuracy and area under the *ROC* curve for product sales effect classification are shown. In Table 11, the high accuracy of prediction indicates that we can accurately predict the sales figure based on the *NePS* value. the *NePS* influences sales, or whether it is just a sign of the *NePS* underlying product market sales movement. Here, sales effect classification means the difference between sales ranks of the next month and the current month is positive or not. If the difference is positive, then it is labeled by 0, otherwise 1.

For the Home & Kitchen dataset, the highest accuracy and AUC are 85.78% and 0.90, respectively, when the latency time is 2 months. For the Digital Camera dataset, the highest accuracy and AUC are 80.23% and 0.85, respectively, when the latency time is 3 months. For the Cell Phones & Accessories dataset, the highest accuracy and AUC are 87.21% and 0.91, respectively, when the latency time is 2 months.

We can predict the exact sales rank using $LSTM$ model based on the *NePS* value. Table 12 presents the MAE between the original sales figure and predicted sales figure. In this table, it is shown that for the Home & Kitchen dataset, the MAE is minimum at 12.73% when the latency period is 2 month. Similarly, for the digital camera dataset, the MAE is minimum at 14.76% when the latency period is 3 month and for the Cell Phones & Accessories dataset, the MAE is minimum at 11.56% when the latency period is 2 month.

We compare the performance of our model with two baselines and it is presented in Table 13. Here in the last fold of five-fold cross validation, 70% of data is selected for training purpose, and 15% of our dataset is selected for validation purpose and the remaining is selected for testing phase. We perform random sampling independently five times and evaluate the mean as the final output for each investigation.

In Table 13, the performances of the baseline (Z. Zhang & Varadarajan, 2006) is presented based on the features of customers' posted review texts. The other baseline model (Ghose & Ipeirotis, 2011) investigates the model's performances using different features of reviewers and review texts. The model in (Ghose & Ipeirotis, 2011) achieves best performance using reviewers' helpful votes and subjectivity of posted reviews and it is presented here based on the best latency time. Both baselines are experimented with Random Forests technique. Our approach is experimented with $LSTM$ based on the *NePS*. The performances of our model is the best compare to the two baselines.

Table 11: Accuracy and area under the ROC curve for the product sales effect classification based on our model (mention in Fig. 4, Subsection 4.2). The best performances are marked in bold.

Dataset	**Metric**	**Latency time**						
		@1month	@2 month	@3 month	@4 month	@5 month	@6 month	@7 month
Home & Kitchen	Accuracy	74.35%	**85.78%**	81.09%	79.45%	41.76%	31.81%	31.33%
	AUC	0.80	**0.90**	0.87	0.87	0.45	0.37	0.37
Digital Camera	Accuracy	31.34%	71.56%	**80.23%**	77.92%	60.88%	37.09%	31.35%
	AUC	0.39	0.76	**0.85**	0.83	0.66	0.42	0.38
Cell Phones & Accessories	Accuracy	80.45%	**87.21%**	80.90%	71.54%	33.33%	33.17%	21.93%
	AUC	0.88	**0.91**	0.88	0.76	0.39	0.39	0.27

Table 12: Experimental result on predicting sales figure based on our model (mention in Fig. 4, Subsection 4.2). Best performances are marked in bold.

Dataset	**Metric**	**Latency time**						
		@1month	@2 month	@3 month	@4 month	@5 month	@6 month	@7 month
Home & Kitchen	MAE	23.98%	**12.73%**	18.84%	20.08%	59.56%	61.05%	65.82%
Digital Camera	MAE	63.11%	24.47%	**14.76%**	21.75%	37.07%	58.56%	65.71%
Cell Phones & Accessories	MAE	17.34%	**11.56%**	18.41%	25.67%	59.04%	61.19%	65.55%

Table 13: Accuracy and area under the ROC curve for the product sales effect classification of our model (mention in Fig. 4, Subsection 4.2) and baselines. Best performances are marked in bold.

Model	Metric	Home & Kitchen	Digital Camera	Cell Phones & Accessories
(Z. Zhang & Varadarajan, 2006)	Accuracy	67.87%	64.81%	69.45%
	AUC	0.70	0.68	0.72
(Ghose & Ipeirotis, 2011)	Accuracy	77.39%	74.65%	76.16%
	AUC	0.83	0.80	0.82
[our]	Accuracy	**85.78%**	**80.23%**	**87.21%**
	AUC	**0.90**	**0.85**	**0.91**

7 Conclusion and Future Work

In this work, we propose the *Network Promoter Score* (*NePS*) to evaluate the reliability, positional influence of customers and reviews posted by customers for E-commerce products. The significant performances of our approach establish the strong regression between the *NePS* and product sales.

Our key contributions can be outlined as follows:

– When we investigate *Hypothesis* 1, we observe that an increase in the *NePS* is related with an increase in product sales.
– The significant performances of our methodology establish the strong regression between the *NePS* and product sales and the result is better than the other independent and control parameters.
– The *NePS* can be used as a strong indicator of product sales for all the considered datasets and can be remarkably futuristic compared to the other parameters.
– Since the effect of the *NePS* is not immediate, we investigate on latency period.

To the best of our knowledge, this is the first work in observing regression between reviews' influential power based on the structure of the review network and product sales. There are many interesting directions for future works. There are some limitations in our research imposed by the nature of the dataset. Publicly available datasets do not provide much information. So in some cases, we have to make some realistic assumptions. In this paper, product sales rank is used as a proxy for product demand. Depending on the availability of data, we can perform our investigation on actual product sales. To the best of our knowledge, sales rank information is available only in Amazon.com as a publicly available dataset. For this reason most of the existing product sales related works have performed (as shown in Table 1) experiments on Amazon dataset only. In future we would explore more datasets and investigate the performance of our methodology on suitable dataset if they become available.

When we are analyzing each customer's positional influence in the network, we can choose many interesting directions. Several E-commerce companies have their own social media pages and customers share their opinion on these platforms. If the social network related dataset is available, then it can be used to understand the customer-customer trust relationships.

We have investigated regression between the *NePS* and product sales based on different latency periods. This latency period depends on data, season, brand offers, marketing strategy. We would like to investigate these control parameters and try to understand the contribution of these parameters on latency period.

Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using lstm network for demand forecasting. *Computers & industrial engineering*, *143*, 106435. https://doi.org/10.1016/j.cie.2020.106435.

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management science*, *57*(8), 1485–1509. https://doi.org/10.1287/mnsc.1110.1370.

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99–126. https://doi.org/10.2753/JEC1086-4415170204.

Baek, H., Oh, S., Yang, H.-D., & Ahn, J. (2017). Electronic word-of-mouth, box office revenue and social media. *Electronic Commerce Research and Applications*, *22*, 13–23. https://doi.org/10.1016/j.elerap.2017.02.001.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of interactive marketing*, *15*(3), 31–40. https://doi.org/10.1002/dir.1014.

Brynjolfsson, E., Hu, Y., & Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, *49*(11), 1580–1596. https://doi.org/10.1287/mnsc.49.11.1580.20580.

Chevalier, J., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and barnesandnoble. com. *Quantitative marketing and Economics*, *1*(2), 203–222. https://doi.org/10.1023/A:1024634613982.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, *29*(5), 944–957. https://doi.org/10.1287/mksc.1100.0572.

Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal* *of the Association for Information Science and Technology*, *66*(2), 354–362. https://doi.org/10.1002/asi.23180.

Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of management information systems*, *23*(2), 149–171. https://doi.org/10.2753/MIS0742-1222230207.

Colladon, A. F., Guardabascio, B., & Innarella, R. (2019). Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems*, *123*, 113075. https://doi.org/10.1016/j.dss.2019.113075.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, *53*(9), 1375–1388. https://doi.org/10.1287/mnsc.1070.0704.

Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, *21*(4), 23–45. https://doi.org/10.1002/dir.20087.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? an empirical investigation of panel data. *Decision support systems*, *45*(4), 1007–1016. https://doi.org/10.1016/j.dss.2008.04.001.

DuBay, W. H. (2004). The principles of readability. *Online Submission*.

Elkahky, A. M., Song, Y., & He, X. (2015). A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th international conference on world wide web* (pp. 278–288). Retrieved from https://doi.org/10.1145/2736277.2741667.

Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, *45*, 74–80. https://doi.org/10.1016/j.jretconser.2018.08.003.

Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? a multi-method investigation. *Decision Support Systems*, *113*, 32–42. https://doi.org/10.1016/j.dss.2018.06.012.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506. https://doi.org/10.1016/j.tourman.2015.07.018.

Filieri, R., Galati, F., & Raguseo, E. (2021). The impact of service attributes and category on ewom helpfulness: An investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior*, *114*, 106527. https://doi.org/10.1016/j.chb.2020.106527.

Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, *55*(8), 956–970. https://doi.org/10.1016/j.im.2018.04.010.

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of net promoter score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138–151. https://doi.org/10.1002/asmb.2417.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research*, *19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193.

Fowdur, L., Kadiyali, V., & Narayan, V. (2009). The impact of emotional product attributes on consumer demand: An application to the us motion picture industry. *Johnson School Research Paper Series*(22-09). Retrieved from http://dx.doi.org/10.2139/ssrn.1407520

Fu, D., Hong, Y., Wang, K., & Fan, W. (2018). Effects of membership tier on user content generation behaviors: evidence from online reviews. *Electronic Commerce Research*, *18*(3), 457–483. https://doi.org/10.1007/s10660-017-9266-7.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE transactions on knowledge and data engineering*, *23*(10), 1498–1512. https://doi.org/10.1109/TKDE.2010.188.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing science*, *23*(4), 545–560. https://doi.org/10.1287/mksc.1040.0071.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., ... Verlegh, P. (2005). The firm's management of social interactions. *Marketing letters*, *16*(3-4), 415–428. https://doi.org/10.1007/s11002-005-5902-4.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (http://www.deeplearningbook.org)

Hansen, D. L., Shneiderman, B., Smith, M. A., & Himelboim, I. (2020). Chapter 3 - social network analysis: Measuring, mapping, and modeling collections of connections. In D. L. Hansen, B. Shneiderman, M. A. Smith, & I. Himelboim (Eds.), *Analyzing social media networks with nodexl (second edition)* (Second Edition ed., p. 31-51). Morgan Kaufmann. Retrieved from https://doi.org/10.1016/B978-0-12-817756-3.00003-0.

He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th international conference on world wide web* (pp. 507–517). Retrieved from https://doi.org/10.1145/2872427.2883037

Hong, H., Xu, D., Wang, G. A., & Fan, W. (2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, *102*, 1–11. https://doi.org/10.1016/j.dss.2017.06.007.

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision support systems*, *57*, 42–53. https://doi.org/10.1016/j.dss.2013.07.009.

Kang, Y., & Zhou, L. (2019). Helpfulness assessment of online reviews:

The role of semantic hierarchy of product features. *ACM Transactions on Management Information Systems (TMIS)*, *10*(3), 1–18. https://doi.org/10.1145/3365538.

Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems*, *96*, 39–48. https://doi.org/10.1016/j.dss.2017.02.001.

Korfiatis, N., GarcíA-Bariocanal, E., & SáNchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205–217. https://doi.org/10.1016/j.elerap.2011.10.003.

Korneta, P. (2018). Net promoter score, growth, and profitability of transportation companies. *International Journal of Management and Economics*, *54*(2), 136–148. https://doi.org/10.2478/ijme-2018-0013.

Li, L., Goh, T.-T., & Jin, D. (2020). How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Computing and Applications*, *32*(9), 4387–4415. https://doi.org/10.1007/s00521-018-3865-7.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, *56*(2), 172–184. https://doi.org/10.1016/j.im.2018.04.007.

Lin, C.-Y., Liaw, S.-Y., Chen, C.-C., Pai, M.-Y., & Chen, Y.-M. (2017). A computer-based approach for analyzing consumer demands in electronic word-of-mouth. *Electronic Markets*, *27*(3), 225–242. https://doi.org/10.1007/s12525-017-0262-5.

Liu, B., et al. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*(2010), 627–666.

Liu, Q., Zhang, X., Zhang, L., & Zhao, Y. (2019). The interaction effects of information cascades, word of mouth and recommendation systems on online reading behavior: An empirical investigation. *Electronic Commerce Research*, *19*(3), 521–547. https://doi.org/10.1007/s10660-018-9312-0.

Malik, M. S. I. (2020). Predicting users' review helpfulness: The role of significant review and reviewer characteristics. *Soft Computing*, 1–16. https://doi.org/10.1007/s00500-020-04767-1.

Mandal, S., & Maiti, A. (2018). Explicit feedbacks meet with implicit feedbacks: A combined approach for recommendation system. In *International conference on complex networks and their applications* (pp. 169–181). Retrieved from https://doi.org/10.1007/978-3-030-05414-4_14

Mandal, S., & Maiti, A. (2020). Explicit feedback meet with implicit feedback in GPMF: a generalized probabilistic matrix factorization model for recommendation. *Appl. Intell.*, *50*(6), 1955–1978. https://doi.org/10.1007/s10489-020-01643-1.

Maslowska, E., Malthouse, E. C., & Viswanathan, V. (2017). Do customer reviews drive purchase decisions? the moderating roles of review exposure and price. *Decision Support Systems*, *98*, 1–9. https://doi.org/10.1016/j.dss.2017.03.010.

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval* (pp. 43–52). Retrieved from https://doi.org/10.1145/2766462.2767755

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521–543. https://doi.org/10.1287/mksc.1120.0713.

Newman, M. (2018). *Networks*. Oxford university press.

Ni, J., Li, J., & McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 188–197). Retrieved from https://doi.org/10.18653/v1/D19-1018

Owen, R. (2019). Net promoter score and its successful application. In *Marketing wisdom* (pp. 17–29). Springer. Retrieved from https://doi.org/10.1007/978-981-10-7724-1_2

Qiu, J., Lin, Z., & Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context. *Electronic commerce research*, *15*(4), 427–452. https://doi.org/10.1007/s10660-015-9191-6.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, *81*(12), 46–55.

Ren, G., & Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing & Management*, *56*(4), 1425–1438. https://doi.org/10.1016/j.ipm.2018.04.003.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20–31. https://doi.org/10.1198/016214501750332668.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30–40. https://doi.org/10.1016/j.dss.2015.10.006.

Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, *108*, 1–12. https://doi.org/10.1016/j.dss.2018.01.004.

Stouthuysen, K., Teunis, I., Reusen, E., & Slabbinck, H. (2018). Initial trust and intentions to buy: The effect of vendor-specific guarantees, customer reviews and the role of online shopping experience. *Electronic Commerce Research and Applications*, *27*, 23–38. https://doi.org/10.1016/j.elerap.2017.11.002.

Sun, X., Han, M., & Feng, J. (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. *Decision Support Systems*, *124*, 113099. https://doi.org/10.1016/j.dss.2019.113099.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560. https://doi.org/10.1109/5.58337.

Wu, Y., DuBois, C., Zheng, A. X., & Ester, M. (2016). Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the ninth acm international conference on web search and data mining* (pp. 153–162). Retrieved from https://doi.org/10.1145/2835776.2835837

Xue, J., Liang, X., Xie, T., & Wang, H. (2020). See now, act now: How to interact with customers to enhance social commerce engagement? *Information & Management*, *57*(6), 103324. https://doi.org/10.1016/j.im.2020.103324.

Zhang, L., Aggarwal, C., & Qi, G.-J. (2017). Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 2141–2149). Retrieved from https://doi.org/10.1145/3097983.3098117

Zhang, Z., & Varadarajan, B. (2006). Utility scoring of product reviews. In *Proceedings of the 15th acm international conference on information and knowledge management* (pp. 51–57). Retrieved from https://doi.org/10.1145/1183614.1183626