

Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews

Abstract

This study aims to investigate the contributions of online promotional marketing and online reviews as predictors of consumer product demands. Using electronic data from Amazon.com, we attempt to predict if online review variables such as valence and volume of reviews, the number of positive and negative reviews, and online promotional marketing variables such as discounts and free deliveries, can influence the demand of electronic products in Amazon.com. A Big Data architecture was developed and Node.JS agents were deployed for scraping the Amazon.com pages using asynchronous Input/Output calls. The completed web crawling and scraping datasets were then preprocessed for Neural Network analysis. Our results showed that variables from both online reviews and promotional marketing strategies are important predictors of product demands. Variables in online reviews in general were better predictors as compared to online marketing promotional variables. This study provides important implications for practitioners as they can better understand how online reviews and online promotional marketing can influence product demands. Our empirical contributions include the design of a Big Data architecture that incorporate neural network analysis which can be used as a platform for future researchers to investigate how Big Data can be used to understand and predict online consumer product demands.

Keywords: Product demands; online reviews; promotional marketing; online marketplace; big data; neural network

1.0 Introduction

Businesses today operate in an increasingly competitive and dynamic environment. Traditionally, manufacturers compete with each other by lowering their production costs and having better product qualities. However, competition on costs and product qualities are becoming more challenging as manufacturers move from offering standardized products and services to one that focuses on customizations (Dobrzykowski et al. 2014). In order to achieve competitive advantages over their rivals, manufacturers now aim to improve the efficiencies of their supply chain, and many achieved this by having better understanding of their customer demands (Chong, Ooi, and Sohal 2009). Bullwhip effect, a traditional challenge faced by manufacturers, can be overcome by having a better understanding and forecasting of customer demands.

The advent of Information Technology (IT) and Data Sciences has empowered companies with the ability to understand and predict customer demands more accurately using quantitative approaches. For example, Radio Frequency Identification Tag (RFID) allows companies to obtain real-time inventory data and their spatial mobility so as to help companies understand and improve product demand forecasts. An emerging IT trend which has captured the attentions of researchers and practitioners is the application of Big Data to better understand business processes and performances. Big Data technologies have the ability to help companies understand complex business relationships by providing useful information that were previously not available to them (Bollen, Mao, and Zeng 2011). The use of data analytics to better understand business processes is not new. Companies such as Wal-Mart and Kohl use various sales, pricing, economic, and demographic data to understand customer behaviors and

product demands. Big Data technologies and the Internet however, provide companies with enhanced abilities to obtain and analyze data from multiple channels, resulting in opportunities for discovering untapped business information. An example of how Big Data can provide new insights can be found in a recent study by Bollen, Mao, and Zeng (2011), who found that the moods of Twitter can influence and predict stock market.

Previous studies have shown that a manufacturer can obtain competitive advantage over its rivals by having an efficient supply chain, and this can be achieved by better understanding the demands of products. An area which has not been studied extensively by previous researchers is how data from online marketplace or e-commerce offers manufacturers the opportunity to better understand product demands (Cao and Schniederjans 2004). Specifically, could information from online user-generated data be useful for manufacturers to understand and forecast product demands better? Online user-generated contents are beginning to have a bigger influence on consumer decisions than traditional media such as newspaper and television. Statistics have shown that as many as 53% of posters on Twitter recommend products or brands in their tweets, and 48% of those who receive the tweets follow through on the recommendations (Flannagan 2011). The state of online retail is that when products are being sold online, the inclusion of user reviews has become a common practice. Besides online reviews, companies selling products online will also include product information such as price and descriptions, as well as promotional marketing information such as the availability of discounts, or 'savings' from discounts of products. The volume of available data related to products, marketing promotions, and online reviews can potentially be used to predict the demand of products and help companies plan their logistics better.

The goal of this study is to examine the comparative influence of promotional marketing strategies such as discounts and the provision of free delivery options, and online reviews information such as the ratings of the products and the percentage of positive and negative reviews on products. The demand of products in this study is measured by a product's online sales rank. We designed a series of Big Data algorithms which sit on a Big Data architecture used for Web data and social media analytics (Ch'ng, 2014). The algorithms use asynchronous I/O (input/output) to request, extract and preprocess data in real-time from Amazon.com. The categories used for our study are electronic devices such as camera, computers and televisions. Besides examining the influence of online promotional strategies and online reviews on product demand, this study also examines if their interaction effects (e.g. when both of them are being offered concurrently online to users) can improve the demand of products (Lu et al. 2013). Recent articles have stated that manufacturers, through the collection of social media chatter, can help understand the real time demand and trends of their products, and thus help to mitigate bullwhip effect (Suominen, 2014). In particular, it is challenging for companies to detect sudden input swings in demand in real time for even small number of items which can result in bullwhip effect, but this can be solved with Big Data (Chase, 2013). Our proposed approach will allow a better understanding of customer demand through the use of online marketplace information, which in turn reduce the risks of bullwhip effect. One of the reasons why our approach reduces bullwhip effect risk is that new and real time data can be extracted and feed into our architecture, and predictions and decisions can be made instantly instead of having the delays of compiling data from various sources which we may find in many legacy systems.

This research makes several important contributions. First, although understanding and predicting product demand is an important topic in operations management, few studies have examine whether the Internet and e-commerce data can help predict product demand. Second,

this study examines whether combining marketing promotional strategies together with online reviews can result in better predictions of product demand. Third, this study demonstrates how Big Data technologies and architecture can be applied to extract data online, in combination with neural networks to predict how product demands can be influenced by promotional marketing and online reviews variables.

2.0 Literature Review

2.1 Online promotional marketing

The shorter product life cycles today, especially in the electronic industry, means that manufacturers are faced with a greater pressure to sell their products in a shorter span of time. As compared to the past, the increase in product information means that more options are available to consumers. For example, consumers are now able to efficiently compare product prices and features online. Due to these business pressures, companies are spending significant amount of their resources to promote and advertise their products online. A popular promotional marketing strategy implemented by companies is to offer price discounts.

2.1.1 Discount Value

Price discounts are popular as they are able to stimulate short term, immediate increase in sales of a product (Gendall et al. 2006). Transaction posit utility theory stated that consumers' demand for a product will increase when the product is being given a higher discount, resulting in consumers believing they have received a bargain (Lichtenstein, Netemeyer, and Burton 1990). The effects of price discounts are often measurable, and because they are able to increase the store traffics, they are able to support relationships between manufacturers and retailers, and ensuring that a particular brand is well stocked and has adequate shelf space in the retail stores (Gendall et al. 2006; Liu et al. 2015). Despite numerous studies on the influence of discount on product demands, results have shown inconsistent and contradictory outcomes on the effects of price discounts on product demands (Drozdenko and Jensen 2005). Marshall and Leng (2002), for example, found that although product sales increased when discount is being offered from 10% to 50%, additional increase at between 60 to 70% have no effect on product sales. Consumers may occasionally use a product's price information to determine the quality of the product instead of just using it to determine monetary gain or loss (Suri, Manchanda, and Kohli 2000). Thus a high price with no discount may indicate high product quality, while a product without discount could mean high monetary sacrifice to the consumers (Suri, Manchanda, and Kohli 2000).

On the other hand, based on attribution theory, when information is being thoroughly processed, consumers are able to rely on attribute information besides price to evaluate a product's quality (Drozdenko and Jensen 2005). When purchasing products online, there are various information about the product, as well as online reviews which may result in customers viewing discount as a monetary gain, and may increase the intention to purchase the product. Furthermore, Jensen et al. (2003) when comparing the prices of online stores versus brick and mortar store, found that having an external reference price has a lesser effects for online stores. These contradictory findings have not been verified in the context of an online environment (The study used price discount as a predictor of a consumer's intention to purchase product). This is important for manufacturers as discounts are often used as a strategy by companies when they are trying to clear their stocks for example, or when they have less stocks on a

particular product, and will use discounts to encourage customers to buy the alternative product under the company, thus giving them more time to replenish their inventory.

2.1.2 Discount Rate

Another predictor used in the present research is the ratio of discount when compared to the actual price of a product. A consumer's perception of price in terms of absolute or relative sense has the ability to influence their price discount perceptions (Chen, Monroe, and Lou 1998). The "psychophysics-of-price-heuristics" theory stated that consumers' psychological utility derived from saving a fixed amount of money is inversely related to an item's price (Chen, Monroe, and Lou 1998). Therefore for a company, they can either provide the absolute price discount, or the relative price discount in %. However, it is still unclear whether a \$20 savings or a 20% discount offer would be more appealing to a consumer on a \$100 jacket. The question then will be, in the presence of both of these information, would one of these lead to better prediction of customer demand of a product?

2.1.3 Free delivery

Besides using price discounts, consumers are likely to purchase a product online if free delivery is offered. (Doern and Fey 2006) in their studies on e-commerce developments and strategies in Russia, found that offering free delivery will result in better customer loyalty and trust. Yip and Law (2002) in examining users' preferences for web site attributes also found that besides special discounts, free delivery is a feature that will attract online users. However, not all websites that use free delivery as an incentive are successful. Smith and Rupp (2003) found that free delivery promotions resulted in the failure of online companies such as Kozmo.com and Urbanfetch.com. The main reason cited in their studies was that online websites which used free delivery to attract users and increase their user base, will find it difficult to retain their customers when they remove the free delivery promotion. With e-commerce becoming mature, offering free delivery may also be something that is expected by consumers, thus those who do not offer free delivery may lose out to their competitors. Based on the discussions, we will be using free delivery as a predictor of customer demand of products, in particular in the presence of other promotional strategies such as price discounts.

2.2. Online Reviews

With the growth of online media, contemporary users regularly and actively share their opinions on products and services with others on various online platforms such as product reviews, blogs, twitter and wikis (Tirunillai and Tellis 2012). This type of communication is presented in "reviews" and the contents are regarded as Word of Mouth. Word of mouth is defined as "all informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers" (Westbrook 1987). Compared with traditional advertising such as television and newspaper ads, eWOM is perceived by consumers as being more credible than private signals, and the information are much more easily accessible through social networks (Davis and Khazanachi 2008). Unlike traditional word of mouth, online word of mouth or electronic word of mouth (eWOM) has far greater reach to other users, and offer much more richness in contents. Users are able to share their online reviews using pictures and even videos. Furthermore, eWOM is able to aggregate both positive and negative information on an online review website from different sources, while traditional word of mouth is only able to capture a single piece of either positive or negative information (Lu et al. 2013).

Previous studies have shown that eWOM is able to affect the sales of products (Chevalier and Mayzlin 2006; Duan, Gu, and Whinston 2008). In these studies, some of the most commonly used attributes of eWOM include the valence, volume, and dispersion of reviews (Lu et al. 2013). Duan, Gu, and Whinston (2008) developed a dynamic simultaneous equation system to capture the relationships between eWOM and motion picture sales. Their studies found that although the valence of eWOM does not directly affect the sales of movie tickets, higher valence was able to generate higher eWOM volume, which in turn increase sales of movie tickets. Chevalier and Mayzlin (2006) examined the effects of eWOM on book sales, and found that customer online reviews are able to influence the sales of books. Lu et al. (2013) examined three year panel data set from an online restaurant review website, and found support between the relationships of valence and volume and product sales.

2.2.1 Online review valence (average rating)

Online review valence is defined as the evaluation score of a specific products or services in eWOM (Lu et al. 2013). Although researchers in the past have proposed that online review valences to have persuasive effects on consumers' purchasing decisions (Cheung and Thadani 2012), the findings on such relationships have been inconsistent. Liu (2006) in his study on the relationships between eWOM and box office revenue, found that although there are significant relationships between the two, most of the explanatory power were derived from the volume of eWOM and not from its valence. Duan, Gu, and Whinston (2008) extended the study by Liu (2006), and found that valence by itself was not able to influence the sales of movies. However, Duan, Gu, and Whinston (2008) found that valence has an indirect relationship with movie sales, as it was able to influence eWOM volume, which in turn influence movie sales. Similarly, Davis and Khazanchi (2008) in their study on multi-product categories online, found that valence of eWOM does not influence the product sales.

On the other hand, researchers such as Lu et al. (2013), Zhu and Zhang (2010) and Chevalier and Mayzlin (2006) found support for the relationship between online review valence and product sales. The ratings of a product and service is increasingly important in an eWOM environment as consumers today are more likely to use make decisions based on wisdoms of the crowds (Chen and Singh 2001). Dellarocas, Awad, and Zhang (2004) also found that valence is one of the strongest predictor of sales among all the other word of mouth attributes.

Besides the inconsistent in previous findings, most of these existing studies have examined the role of eWOM on experience products such as movies and books. Studies on search products however have remained sparse. Search products include electronics where a consumers can evaluate the specific attributes of the product before purchasing (Cui, Lui, and Guo 2012). When consumers purchase electronic products, they are more likely to apply a systematic decision making process by evaluating specific attributes of the product. In comparison, customer purchasing movie tickets or books are more likely to make their decisions based on extrinsic attributed related cues (Cui, Lui, and Guo 2012). Therefore when purchasing an electronic product, a customer will evaluate the product's technical aspects and performances. In an online information, such information are readily available, and the ratings of the products are prominently displayed to the customers (Cui, Lui, and Guo 2012). Therefore valence may be a predictor of electronic product sales, and not in experience products such as in the studies conducted by Liu (2006) and Duan, Gu, and Whinston (2008). Based on the discussions and inconsistent previous findings, this study include valence as a predictor of the electronic product sales.

2.2.2 Online review volume

Online review volume is the quantity of reviews on the product or service (Lu et al. 2013). Studies on why online review volume increase the product sales is relatively straight forward, i.e., more discussions about a product or service in eWOM will lead to an increased awareness among consumers, resulting in changes in sales according to Davis and Khazanchi (2008). In this study, online review volume refers to the number of comments from reviewers about the specific electronic product. Previous studies have provided very strong support between the relationships of eWOM volume and consumers' demand of the product (Duan, Gu, and Whinston 2008; Davis and Khazanchi 2008; Lu et al. 2013). Lu et al. (2013) for example, found that the volume of eWOM has led to an increase in the sales of a product in the context of restaurants. However, Cui, Lui, and Guo (2012) stated that the volume of online reviews play a more important role in experience goods when compared to search products such as electronics. In the case of experience goods, as users are not able to experience product attributes, they tend to focus on extrinsic cues such as product popularity which is reflected by the eWOM review volume. In the case of electronic products, Cui, Lui, and Guo (2012) believe that the valence of eWOM plays a more important role than its volume. Given that most studies have supported the importance of online review volume in predicting product sales, we found that it is important to include volume as one of the predictors in our present study. However, this study will also examine to confirm if valence will play a more important role than volume in our study which uses electronic products as suggested by Cui, Lui, and Guo (2012), or will our result support previous studies.

2.2.3 Percentage of negative online reviews

Studies in the past have found that negative eWOM can influence a customer's purchasing decision more than the positive eWOM (Cheung and Thadani 2012). Negative eWOM is also stated to spread much faster than positive eWOM (Cui, Lui, and Guo 2012). A positive comment on eWOM by consumers reflect a product's quality and reputation, while a negative comment shows that users lack confidence in a product and may result in poor product sales (Cui, Lui, and Guo 2012). Thus other than the ratings of the product, the proportion of positive and negative reviews can influence the purchasing decisions of consumers. Psychologists have found that negative information tends to influence evaluations much more strongly when compared to extreme positive information (Ito et al. 1998). Negative information are therefore considered as being more useful for decision-making purposes and is given greater weight when compared to positive information (Lee, Park, and Han 2008).

Chakravarty, Liu, and Mazumdar (2010) found that negative word of mouth has a strong effect on movie evaluations when compared to positive word of mouth. It was found that the negative word of mouth's influence on less frequent moviegoers was enduring even in presence of positive critics review. Zhang, Craciun, and Shin (2010) in examining software programs found that when associated with prevention consumption goals, negative reviews are more persuasive than positive reviews. Lu et al. (2013) also found that negative review percentage has a direct effect on restaurant sales. As negative reviews may strongly affect product sales, we included the percentage of negative reviews as predictors of customer demand of electronic products. Furthermore, we have included the percentage of positive reviews in order to compare to see if negative comments do indeed play a more important role than positive comments in determining product sales.

2.2.4 Helpful reviews and number of questions answered

Although online reviews are useful to consumers in product purchasing decisions, as the availability of online reviews become widespread, consumers are more likely to be influenced by the evaluation and use of the reviews (Mudambi and Schuff 2010). Helpfulness is being used by online marketplace to measure how consumers evaluate a review. In some websites for example, users will be asked the question “Was this review helpful to you?” after a customer review, and it also provides helpfulness information alongside the review (e.g. “20 out of 40 people found the following review helpful”) (Mudambi and Schuff 2010). Past studies have examined extensively on what makes a review helpful (Mudambi and Schuff 2010; Min and Park 2012). However, studies on whether helpful reviews can result in better sales have remained elusive.

Besides providing consumers with the helpfulness of reviews, some online marketplace helps its customers by answering customers’ questions. Computer mediated communication technologies such as instant messengers can allow consumers to enquire further information from online sellers, and this has been proven to be successful in websites such as Taobao.com (Ou, Pavlou, and Davison 2014). Whether having more answered questions will result in better product sales however, have not been examined by researchers.

2.3 Interactions between online reviews and online promotional marketing (discount rate)

A challenge faced by customers today is the amount of online information available to them. The vastness of information can result in difficult decision making when making purchases (Chong and Ngai 2013). Many companies are less likely to adopt just one marketing strategy when promoting their products online. As Lu et al. (2013)’s study shown, firms today are more likely to offer multiple information channels from vendors to users and from previous customers to users to reach out to them. However, would promotional marketing efforts from a firm be a better predictor of its product sales, or would online reviews offer better predictions?

Although Lu et al. (2013)’s study has examined if eWOM moderates the effects of online marketing promotions, it was focused on restaurant sales rather than electronics where the role of discounts and reviews may have different effects due to the different characteristics of the products. Consumers of electronic products are constantly on the lookout for the latest releases. Therefore when a newer product is on offer, would customers still purchase the older product, or would reviews on the older products still be relevant to the customer when newer products tend to have better performance and specifications? Furthermore, given that discounts have contrary findings in their impact on product sales whereby discounts may be viewed negatively on a product’s quality (Suri, Manchanda, and Kohli 2000), are discounts more effective and viewed more positively if there are more positive reviews? Lu et al. (2013) in their study on restaurant sales, found that the interactions between volume and coupon offerings have negative relationships, which suggests that when eWOM volumes are high, coupon promotions are not effective. This study extends on Lu et al. (2013)’s study, and examine if online reviews such as its volume, valence and percentage of positive comments’ interactions with price discounts can improves the prediction on product sales and customer demand of products.

3.0 Neural network

Neural network is a machine learning technique that is inspired by the human brain (Chiang, Zhang, and Zhou 2006), where the networks are presented as systems of interconnected neurons which can compute values from input information. A neural network can learn the intrinsic nature of patterns or processes from sample data (Sim et al., 2014). A neural network consists of a set of nodes that are distributed in hierarchic layers (Chiang, Zhang, and Zhou 2006). Most neural networks will contain three types of layers—input, hidden and output (Chong 2013). As the name suggests, the input node is the receiver which input data files. The output layer is the final information generated. The hidden layers are between the input and output layers. The hidden layers will receive inputs from neurons in the input layer, and knowledge is then stored by the interneuron connection strengths (i.e. synaptic weights) (Haykin 1994). Using supervised learning algorithm, the neural network will analyse the data set, and the synaptic weights of the neural weight will be adjusted to attain the desired design objective (Chong et al. 2013). They are then used to store knowledge and make it available for future use (Sim et al., 2014).

Figure 1 provides an example of neural network with five input neurons, four hidden layers and one output layer.

<<Figure 1 about here>>

In this study, neural network is applied to predict the factors influencing the customer demand of electronic products in an online environment. Previous studies on online reviews have mainly used explanatory statistical techniques to examine the research models (Shmueli and Koppius 2011; Lu et al. 2013). However, there is an increase in the need to apply predictive analytics into information systems research (Shmueli and Koppius 2011). One of the key advantage that predictive model such as neural network can offer is that they are able to create useful and practical model which can help researchers to develop new theory (Shmueli and Koppius 2011). Another challenge with statistical analysis relying on p-value instead of to establish significance testing may not be effective as larger amounts of data could result in every predictors being significant (George, Hass, and Pentland 2014). By using traditional statistical tools to analyse large data sets such as Big Data, there is a risk of obtaining false correlations in the result (George, Hass, and Pentland 2014).

Neural network can detect complex non-linear relations and all possible interactions that are not pre-defined, thus the studying of independent and dependent variables is not limited by hypothesized relationships (Sargent 2001; Moosmayer et al. 2013). Also, neural network is non-parametric model. Compared to parametric regression models, neural network does not assume about probability distributions of the variables (Aljahdali 2001). Previous works have shown that neural network generally outperformed or tied logistic regression (Dreiseitl and Ohno-Machado 2002), as well as multiple and discriminant regression analysis (Chong 2013). Neural network's 'black-box' nature is its main disadvantage. However, in this study, our intention is to identify the important variables rather than to interpret the model. Therefore, considering that neural network can be used to estimate model parameters and has several advantages, we applied neural network to identify the important predictors of sales.

4.0 Methodology

4.1 Research Context and data

Our research aims to demonstrate how manufacturers can use online shops as predictors of customer demands. In our research, a retailer hosted eWOM website Amazon.com is used as the source of our data. Amazon.com is successful, well known and highly reputable, it is a good source of data for our study on the influence of eWOM and online promotional strategies. We avoided sites which have a lesser ranking to prevent risk from skewed data due to low reputation and sales. Our risk is minimal on a site like Amazon.com.

This study will examine eWOM using Amazon.com, and the products include electronic devices such as camera, television, Hi-Fi, notebook etc. Unlike books, which have been examined by Chevalier and Mayzlin (2006), electronic products have shorter product shelf lives, and it would be interesting to see how the relationships between factors such as price, promotion and online reviews influence product sales (Chong and Ooi 2008). Similar to another study (Lu et al. 2013), we used Amazon.com solely as we are unable to include dispersion in our study. Dispersion of eWOM is defined as the extent in which conversation on a product or service is being carried out across broad range of communities. Instead, this research will examine the relationships between valence, volume, and percentage of positive, and percentage of negative of online reviews on the sales of electronic products. Table 1 provides the summary of variables used for this study and their descriptions. 15 variables, as well as 3 interactions effects (e.g. Positive Review X Discount Rate, Valence X Discount Rate and Volume X Discount Rate) are used as predictors of this study. Figure 2 shows these variables on Amazon.com.

<<Table 1 about here>>

<<Figure 2 about here>>

4.2 Big Data Technology

Our choice of technology sets a foundation for future studies, and serves its purpose in this research. Processing I/O (input/output) of hundreds of pages of web products is manageable on standard desktop with conventional network connections. However, managing tens to hundreds of thousands of web page processing, including cleaning data during runtime does require scalable technology. Big Data technology has become a necessity for research in the 21st century, and we lay a foundation here for extracting data for the purpose of informing our research.

Figure 3 describes our process. 1) Developmental workstation where our Node.JS agents are deployed for scraping the web using asynchronous I/O calls. 2) Physical server hosting Ubuntu 64bit virtual machines, and where data is stored and horizontally scaled. 3) Completed Web crawling and scraping datasets are converted into comma separated values for Neural Network analysis.

A series of asynchronous I/O algorithms were custom-developed for acquiring and pre-processing Amazon.com data. The system sits within a Web and Social Media Big Data client-server architecture, integrating various open-source server technologies used by large

corporations (e.g., LinkedIn, Yahoo, Microsoft, eBay, etc.) (Ch'ng, 2014). The system consists of 6x Linux Ubuntu 64bit Virtual Machines (VM) on 2x HP DL388p physical servers. The physical system is horizontally scalable as needs arises for additional VMs. Due to the asynchronous I/O nature of our algorithms, as data comes in, they are efficiently stored within MongoDB, a cross-platform NoSQL database that is horizontally scalable prior to synchronous extraction later into CSV (Comma Separated Value) file for our Neural Network analysis. Our asynchronous algorithms are coded in server-side JavaScript via Node.js. Node.js is an event-driven, non-blocking I/O model built on Google's V8 JavaScript engine. Due to the asynchronous nature of Node.js, it is capable of managing real-time, data-intensive applications such as what we have in the present research. Our algorithms are developed and deployed on a Dell T3600 Tower Workstation with 64GB of RAM, 6 cores 12 threads, with Quadro K4000 and Tesla K40c GPGPU. The Tesla K40c was prepared for parallel processing needs, however, it was not utilized as the data was not sufficiently large to require multicore processing.

We first identified product links before deploying our Web crawlers, once our crawlers have reached a sufficient population of product links, we deployed our asynchronous Web scraper agents, which parses the incoming HTML data and extract targeted elements via regular expressions in real-time. Incoming data are immediately stored within the MongoDB server, and a CSV file was generated after the scraper agents completed their jobs.

Our asynchronous code is capable of sending thousands of concurrent sockets where agents requests for Amazon.com pages, however, to prevent our IP from being blocked, a recursive mechanism were implemented so that we can control n requests per set. It takes on average 1.1392 seconds to call a HTTP request, get a HTML response and scrap the page of all required data. We extracted all electronics data there is on Amazon, otherwise we could have continued data scraping with our highly efficient system.

<<Figure 3 about here>>

The total number of records in our study is 35,203. Our sample includes 813 Audio and HiFi devices, 23716 Camera and Photo related devices, 9870 Computers and Accessories, 264 Television and Home Cinema, 92 outdoor and sports related electronic devices, and 448 other electronic devices.

4.3 Neural Network Analysis

The data for this research is analysed using back-propagation neural network. A three layer neural network is developed for this study. Values between 0 and 1 will be assigned to the initial weights and biases. The neural network will then be provided with training data with sets of inputs such as discount price, valence of reviews, volume of reviews and output which is the sales rank. The neural network model developed for this research is shown in Figure 4.

The difference between the actual output (e.g. sales rank) and the desired output will be calculated and back-propagated to the previous layers (Chong et al. 2013). The neural network applies the Delta rule to adjust the connection weight and reduces the output errors. This process is then back-propagated to the previous layer until it reaches the input layer (Chiang, Zhang, and Zhou 2006).

4.4 Validations of Neural Networks

We applied multilayer perceptron training algorithm to train the neural network in this study. Similar to existing studies (Chong et al. 2013), cross validations were conducted to avoid over-fitting of the model. In order to determine the ideal number of hidden nodes, we increase the hidden nodes starting from 1, and increase the number of hidden node by one and check this against the errors in the neural network. The ideal number of hidden network is one which does not increase the neural network's errors (Chong et al. 2013).

Networks with four hidden nodes were found to be complex enough to map the datasets without incurring additional errors to the neural network model. Our neural network therefore consists of 14 predictors, six hidden nodes, and one output variable.

The activation function for the hidden and output layers used in this study is the sigmoid function. The sigmoid function approaches the value of one for large positive numbers and 0.5 for zero and very close to zero for large negative numbers (Sim et al. 2014). As a result, it allows transition between the low and high output of the neurons. A ten-fold cross validation was performed whereby we used 90 percent of the data to train the neural network, while the remaining 10 percent was used to measure the prediction accuracy of the trained network. Root Mean Squared Error (RMSE) was computed to compare data from the training and testing sets to ensure that there are not much difference between the two tables. The RMSE of the validations are shown in Table 2.

<<Table 2 about here>>

From Table 3, the average RMSE for the training model is 0.00102 while the testing model is 0.0103. The RMSE values for the two models are relatively consistent and do not vary much, and we can therefore be confident that the network model is reliable in capturing the numeric relations between the predictors and outputs.

4.5 Sensitivity analysis

The importance for predictors in this study was calculated using sensitivity analysis. Sensitivity analysis performance was calculated by averaging the importance of the predictors over ten networks (Chong et al. 2013). The importance of the predictor variable is a measure of how much the network's model-predicted value changes for different values of the predictor variable (Chong et al. 2013). The importance was calculated by average the predictors' importance over ten networks and expressed as a percentage (Chong et al. 2013).

<<Table 3 about here>>

Table 3 shows that all 14 predictors are found to be relevant to all ten networks. The average result showed that the two main and most important predictors are volume*discount rate and number of answered questions. The result showed that in general, eWOM related variables such as positive reviews, negative reviews, valence, volume, rating of the most helpful favourable review, and rating of the most helpful critical review are better predictors of electronic product sales than online promotion strategies such as discounts and free deliveries. However, interaction effects between discount rate with volume and positive review are important predictors of electronic sales. We also conducted correlation analysis for the

predictors and sales, and found that besides negative reviews which have a negative relationship with sales rank, all other predictors have positive relationship with sales rank.

5.0 Discussions

Our results show that all predictors in our neural network are able to predict the sales of electronic products online, although some predictors play more important roles than others. In general, online review variables such as positive reviews, review helpfulness, negative reviews and so forth are important predictors of consumer demand. However, volume moderated by discount rate and number of answered questions are two of the most important predictors in our neural network model. This is a very interesting result as there are very few studies that examine how online volume can be moderated by other online variables such as discount rates.

Previous studies found that the volume of online reviews can increase the sales of products (Duan, Gu, and Whinston 2008; Davis and Khazanchi 2008), and hence serves as a predictor of consumer demand of products. However, this study found that although volume is an important predictor of product demands, its interactions with discount rate is the most important predictor of product sales, suggesting that products with high volume, and then subsequently given discount, can lead to an increase in the sales of electronic products in this study. Lu et al. (2013)'s study on online restaurants sales found that in a high volume eWOM environment, discount will not influence sales. However, our result contradicts Lu et al. (2013)'s findings, specifically in the context of electronic products. Thus consumers are unlikely to perceive electronic products' discount negatively which reduces their purchases, and they will also not be influenced by eWOM to the extent that discounts play no role in their demand of electronic products.

The number of answered questions is one of the most important predictors in our neural network model. Although there have been very limited studies on the role of answered questions, it shows that addressing customers' direct queries is an important factor that increase in product sales. Ou, Pavlou, and Davison (2014) in their study on Chinese online marketplace, shows that one key reasons that drives the customer re-purchasing intentions is the use of instant messenger allowing sellers to answer buyers' queries. Answering customers' questions and displaying them online can also show a sense of presence by the online marketplace, thus increasing the likelihood of purchases.

The number of positive reviews and negative reviews are both important predictors of electronic product sales online. Previous studies have mainly supported the notion that negative eWOM have stronger influence, and spread faster than positive eWOM (Cui, Lui, and Guo 2012; Cheung and Thadani 2012; Lee, Park, and Han 2008). However, our results show that positive eWOM have stronger prediction success on sales purchase than negative eWOM. Thus in the context of electronic products, consumers are influenced more by the positive online reviews, although it should be noted that the differences between the strengths of predictions of positive and negative reviews do not vary very much. Our study also shows that the interactions of positive reviews with discount rate are more likely to result in the sales of electronic products than just by offering discounts. This again confirms the importance of combining both online reviews with promotional strategies to drive sales of the electronic products. Surprisingly, although the interactions of discount rates with volume and positive reviews are both good predictors of product sales, the interactions between discount rate and valence is not a very strong predictor of sales compared to the other variables. Valence is also not a strong predictor in our neural network model. This shows that consumers are more likely

to be influenced by the numbers of positive or negative reviews, and in the case of Amazon.com, the numbers of ratings of 1 to 5. Although an overall review rating provides a general indication of the satisfaction level of consumers, consumers are more interested to know the specific breakdowns of the ratings than an overall figure.

Our study shows that offering discounts and free deliveries do not strongly influence the sales of products. This shows that in an online marketplace, online reviews play a pivotal role in driving sales. Free deliveries do not play an important role, and it provides similar incentives to consumers as reducing costs, but again, in the context of online marketplace, offering reduction in price, especially in the context of electronic products, are not significant enough to influence the purchasing decision of consumers.

6.0 Conclusions and Implications

This paper uses Big Data technology and neural network modelling to study the product demands of electronic products online. The results show that the variables used in the study are all useful predictors of product sales. The results also show that online reviews variables and online promotional marketing variables drawn from previous studies are applicable for developing a model that predicts the sales of electronic products online. Our model also showed that all predictors are important as none of them were pruned from during the sensitivity analysis. This study has several important implications.

First, this research demonstrates how research supported by Big Data technology can be used to efficiently gather large datasets, and as a result assist in predicting customer demand of products in online environments. The special set of asynchronous algorithms used and the Big Data architecture employed here demonstrates how data can be very quickly and efficiently extracted and pre-processed in real-time. Whilst we have demonstrated the data processing from a single site, the scalability of the integrated system can go beyond that of a single site in very large and connected datasets, such as that used in dispersion across social media and networks. As such, future longitudinal studies are possible. Furthermore, this research employs a neural network model to examine the prediction importance of the variables, instead of conducting traditional p-value testing which will not be effective in a big data environment (George, Haas, and Pentland 2014). Approach such as neural network to examine big data to examine business theories and models is recommended by George, Haas, and Pentland (2014), and subsequently being demonstrated in this paper. Practitioners in the future can also examine their products sold online, and use similar approach to predict customer demands.

Second, our research also examines the roles of online reviews and online marketing promotional strategies in influencing the sales of electronic products. Based on our findings, online sellers should pay important attentions to customers' online questions and display the answers online. A strategy to be considered by online marketplace such as Amazon.com is to provide live, chatting service that can answer customers' queries 24x7. Instant massagers have shown to be effective in increasing consumers' re-purchasing intentions in popular online marketplace such as Taobao.com, and more companies can employ this strategy.

Third, our results show that online reviews are important predictors of electronic product sales. Some online review attributes are more important than others. For example, the number of positive reviews and number of negative reviews are better predictors of sales when compared to the valence of the reviews. Therefore instead of just displaying an overall review figure,

when designing the website, managers or decision makers of the online companies should display the specific number of ratings given by online customers.

Fourth, our result shows that online marketing promotional strategies such as discounts should be employed together with online reviews to improve sales. Price reduction strategies such as discounts and free deliveries do not strongly influence customers' purchasing decisions. Decision makers can also consider an integrated strategies taking into considerations of online reviews and discounts will be more successful in increasing product sales.

Lastly, this study shows that manufacturers can predict their product demands via online marketplaces. E-commerce is now one of the main channels for selling products, and traditionally, companies focused on improving product forecasts and reducing bullwhip effects by examining data from their physical storefronts. With e-commerce increasingly becoming the main channels of product sales, manufacturers should also examine the forecast of their product demands through online marketplace. However, unlike traditional physical storefronts, online marketplace are also influenced by online reviews. As shown by this study, online reviews service as an important predictors of product sales online, and manufacturers should take account of this when planning for their productions.

7.0 Limitations and future directions

There are several limitations for this study. First, this study only examine electronic products. Future study should consider including more products and examine if our model is applicable to all product types. Secondly, our sample size is of around thirty thousand records is not very large. However, as mentioned in the paper, this study applied big data architecture, as well as employing neural network together with the architecture, and this can serves as a platform for future researchers who have larger sample size studies. Lastly, our study focused only on Amazon.com. Future researches can consider examine other online marketplaces for comparisons an generalizability.

Acknowledgement

"The authors acknowledges the financial support from the National Natural Science Foundation of China (NSFC), International Doctoral Innovation Centre, Ningbo Education Bureau, Ningbo Science and Technology Bureau, China's MoST and The University of Nottingham. The project is partially supported by NSFC no. 71402076 and NBSTB Project 2012B10055"

References

- Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market." Review of. *Journal of Computational Science* 2 (1):1-8. doi: <http://dx.doi.org/10.1016/j.jocs.2010.12.007>.
- Cao, Q., and M. J. Schniederjans. 2004. "Empirical study of the relationship between operations strategy and information systems strategic orientation in an e-commerce environment." Review of. *International Journal of Production Research* 42 (15):2915-39.
- Chakravarty, Anindita, Yong Liu, and Tridib Mazumdar. 2010. "The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation."

- Review of. *Journal of Interactive Marketing* 24 (3):185-97. doi: <http://dx.doi.org/10.1016/j.intmar.2010.04.001>.
- Chen, Mao, and Jaswinder Pal Singh. 2001. Computing and using reputations for internet ratings. Paper presented at the Proceedings of the 3rd ACM conference on Electronic Commerce.
- Chen, Shih-Fen S., Kent B. Monroe, and Yung-Chien Lou. 1998. "The effects of framing price promotion messages on consumers' perceptions and purchase intentions." Review of. *Journal of Retailing* 74 (3):353-72. doi: [http://dx.doi.org/10.1016/S0022-4359\(99\)80100-6](http://dx.doi.org/10.1016/S0022-4359(99)80100-6).
- Cheung, Christy M. K., and Dimple R. Thadani. 2012. "The impact of electronic word-of-mouth communication: A literature analysis and integrative model." Review of. *Decision Support Systems* 54 (1):461-70. doi: <http://dx.doi.org/10.1016/j.dss.2012.06.008>.
- Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Review of. *Journal of Marketing Research* 43 (3):345-54. doi: 10.1509/jmkr.43.3.345.
- Chiang, Wei-yu Kevin, Dongsong Zhang, and Lina Zhou. 2006. "Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression." Review of. *Decision Support Systems* 41 (2):514-31. doi: <http://dx.doi.org/10.1016/j.dss.2004.08.016>.
- Ch'ng, E. (2014) The Value of Using Big Data Technology in Computational Social Science, The 3rd ASE Big Data Science Conference, 4-7 August 2014, Tsinghua University Beijing China.
- Chong, Alain Yee-Loong. 2013. "Predicting m-commerce adoption determinants: A neural network approach." Review of. *Expert Systems with Applications* 40 (2):523-30. doi: <http://dx.doi.org/10.1016/j.eswa.2012.07.068>.
- Chong, Alain Yee-Loong, Felix TS Chan, Mark Goh, and MK Tiwari. 2013. "Do interorganisational relationships and knowledge-management practices enhance collaborative commerce adoption?" Review of. *International Journal of Production Research* 51 (7):2006-18.
- Chong, Alain Yee-Loong, and Eric TW Ngai. 2013. "What Influences Travellers' Adoption of a Location-based Social Media Service for Their Travel Planning?" Review of. *PACIS 2013 Proceedings* 210.
- Chong, Alain Yee-Loong, Keng-Boon Ooi, and Amrik Sohal. 2009. "The relationship between supply chain factors and adoption of e-Collaboration tools: An empirical examination." Review of. *International Journal of Production Economics* 122 (1):150-60. doi: <http://dx.doi.org/10.1016/j.ijpe.2009.05.012>.
- Chong, Alain Yee-Loong, and Keng-Boon Ooi. 2008. "Adoption of interorganizational system standards in supply chains." Review of. *Industrial Management & Data Systems* 108 (4):529-47. doi: doi:10.1108/02635570810868371.
- Cui, Geng, Hon-Kwong Lui, and Xiaoning Guo. 2012. "The effect of online consumer reviews on new product sales." Review of. *International Journal of Electronic Commerce* 17 (1):39-58.
- Davis, Alanah, and Deepak Khazanchi. 2008. "An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales." Review of. *Electronic Markets* 18 (2):130-41.
- Dellarocas, Chrysanthos, Neveen Awad, and Xiaoquan Zhang. 2004. "Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning." Review of. *ICIS 2004 Proceedings*:30.

- Dobrzykowski, David, Vafa Saboori Deilami, Paul Hong, and Seung-Chul Kim. 2014. "A structured analysis of operations and supply chain management research in healthcare (1982–2011)." Review of. *International Journal of Production Economics* 147, Part B (0):514-30. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.04.055>.
- Doern, Rachel R., and Carl F. Fey. 2006. "E-commerce developments and strategies for value creation: The case of Russia." Review of. *Journal of World Business* 41 (4):315-27. doi: <http://dx.doi.org/10.1016/j.jwb.2006.08.001>.
- Drozdenko, Ronald, and Marlene Jensen. 2005. "Risk and maximum acceptable discount levels." Review of. *Journal of Product & Brand Management* 14 (4):264-70. doi: doi:10.1108/10610420510609285.
- Duan, Wenjing, Bin Gu, and Andrew B. Whinston. 2008. "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry." Review of. *Journal of Retailing* 84 (2):233-42. doi: <http://dx.doi.org/10.1016/j.jretai.2008.04.005>.
- Flannagan, Ryan. 2014. "How important are online customer reviews?", Accessed 2014-10-30. <http://nuancedmedia.com/how-important-are-online-customer-reviews/>.
- Gendall, Philip, Janet Hoek, Tracy Pope, and Karen Young. 2006. "Message framing effects on price discounting." Review of. *Journal of Product & Brand Management* 15 (7):458-65. doi: doi:10.1108/10610420610712847.
- George, Gerard, Martine R Haas, and Alex Pentland. 2014. "Big Data and Management." Review of. *Academy of Management Journal* 57 (2):321-6.
- Haykin, Simon. 1994. *Neural networks: A comprehensive foundation*: NY: Macmillan College Publishing.
- Ito, Tiffany A, Jeff T Larsen, N Kyle Smith, and John T Cacioppo. 1998. "Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations." Review of. *Journal of personality and social psychology* 75 (4):887.
- Jensen, Thomas, Jeremy Kees, Scot Burton, and Fernanda Lucarelli Turnipseed. 2003. "Advertised reference prices in an Internet environment: Effects on consumer price perceptions and channel search intentions." Review of. *Journal of Interactive Marketing* 17 (2):20-33. doi: 10.1002/dir.10052.
- Lee, Jumin, Do-Hyung Park, and Ingoo Han. 2008. "The effect of negative online consumer reviews on product attitude: An information processing view." Review of. *Electronic Commerce Research and Applications* 7 (3):341-52. doi: <http://dx.doi.org/10.1016/j.elerap.2007.05.004>.
- Lichtenstein, Donald R., Richard G. Netemeyer, and Scot Burton. 1990. "Distinguishing Coupon Proneness from Value Consciousness: An Acquisition-Transaction Utility Theory Perspective." Review of. *Journal of Marketing* 54 (3):54-67. doi: 10.2307/1251816.
- Liu, Martin J., Natalia Yannopoulou, Xuemei Bian and Richard Elliot. 2015. "Perceptions of Authenticity within the Chinese Marketplace." *Journal of Business Research* 68(1), pp. 27 - 33.
- Liu, Yong. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue." Review of. *Journal of Marketing* 70 (3):74-89. doi: 10.1509/jmkg.70.3.74.
- Lu, X., S. Ba, L. Huang, and Y. Feng. 2013. "Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews." Review of. *Information Systems Research* 24 (3):596-612. doi: doi:10.1287/isre.1120.0454.
- Marshall, Roger, and Seow Bee Leng. 2002. "Price threshold and discount saturation point in Singapore." Review of. *Journal of Product & Brand Management* 11 (3):147-59. doi: doi:10.1108/10610420210430042.

- Min, Hye-Jin, and Jong C. Park. 2012. "Identifying helpful reviews based on customer's mentions about experiences." Review of. *Expert Systems with Applications* 39 (15):11830-8. doi: <http://dx.doi.org/10.1016/j.eswa.2012.01.116>.
- Mudambi, Susan M, and David Schuff. 2010. "What makes a helpful online review? A study of customer reviews on Amazon. com." Review of. *Management Information Systems Quarterly* 34 (1):11.
- Ou, Carol Xiaojuan, Paul A Pavlou, and Robert M Davison. 2014. "Swift Guanxi in online marketplaces: the role of computer-mediated communication technologies." Review of. *MIS Quarterly* 38 (1):209-30.
- Shmueli, Galit, and Otto R Koppius. 2011. "Predictive analytics in information systems research." Review of. *MIS Quarterly* 35 (3):553-72.
- Sim, Jia-Jia, Garry Wei-Han Tan, Jessica C. J. Wong, Keng-Boon Ooi, and Teck-Soon Hew. 2014. "Understanding and predicting the motivators of mobile music acceptance – A multi-stage MRA-artificial neural network approach." Review of. *Telematics and Informatics* 31 (4):569-84. doi: <http://dx.doi.org/10.1016/j.tele.2013.11.005>.
- Smith, Alan D., and William T. Rupp. 2003. "Strategic online customer decision making: leveraging the transformational power of the Internet." Review of. *Online Information Review* 27 (6):418-32. doi: doi:10.1108/14684520310510055.
- Suri, Rajneesh, Rajesh V. Manchanda, and Chiranjeev S. Kohli. 2000. "Brand evaluations: a comparison of fixed price and discounted price offers." Review of. *Journal of Product & Brand Management* 9 (3):193-207. doi: doi:10.1108/10610420010332458.
- Tirunillai, S., and G. J. Tellis. 2012. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance." Review of. *Marketing Science* 31 (2):198-215. doi: doi:10.1287/mksc.1110.0682.
- Westbrook, Robert A. 1987. "Product/Consumption-Based Affective Responses and Postpurchase Processes." Review of. *Journal of Marketing Research* 24 (3):258-70. doi: 10.2307/3151636.
- Yip, Lily, and Rob Law. 2002. "User preferences for web site attributes: A study of the Hong Kong Disneyland." Review of. *Asia Pacific Journal of Tourism Research* 7 (1):33-44. doi: 10.1080/10941660208722108.
- Zhang, Jason Q., Georgiana Craciun, and Dongwoo Shin. 2010. "When does electronic word-of-mouth matter? A study of consumer product reviews." Review of. *Journal of Business Research* 63 (12):1336-41. doi: <http://dx.doi.org/10.1016/j.jbusres.2009.12.011>.
- Zhu, Feng, and Xiaoquan Zhang. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics." Review of. *Journal of Marketing* 74 (2):133-48. doi: 10.1509/jmkg.74.2.133.

Figure 1 Neural Network Example

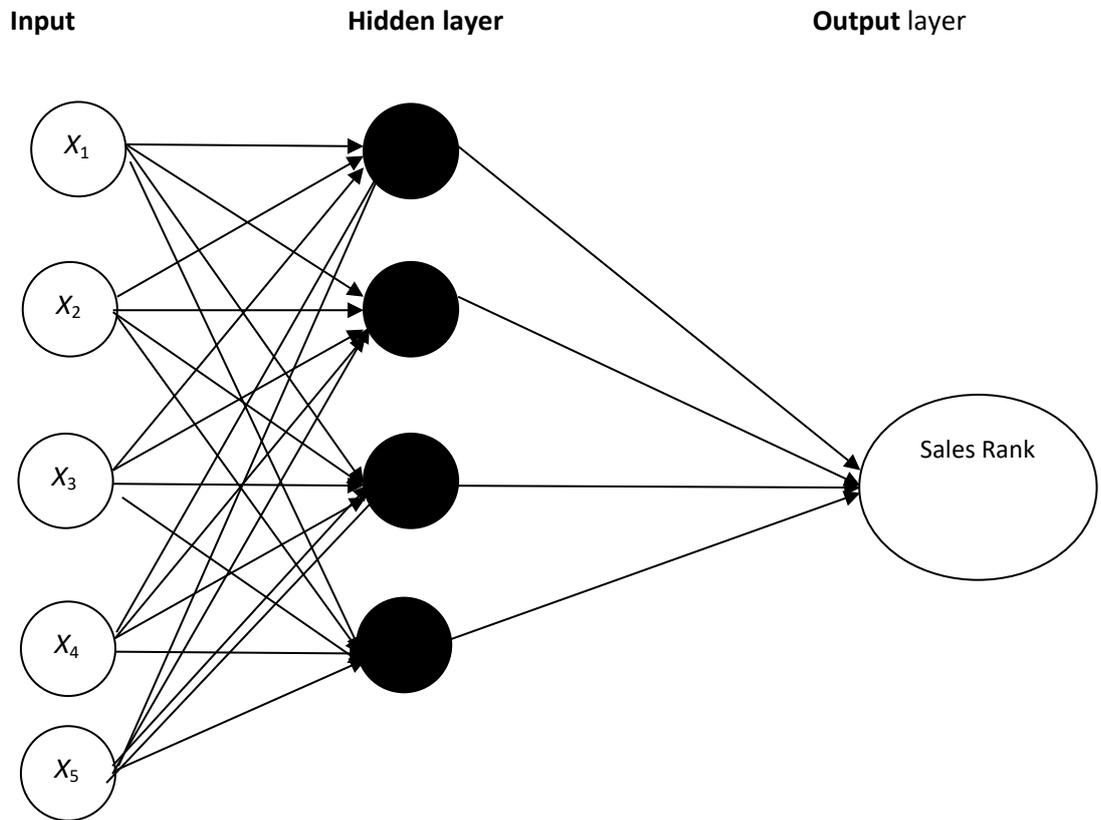


Figure 2 Variables in Amazon.com



Canon EOS 1200D Digital SLR Camera with EF-S 18-55mm f/3.5-5.6 III Lens

Part of Canon's [Summer Cashback Offer](#)

by Canon (B)

(A)  151 customer reviews | 138 answered questions (C)

RRP: £399.99 (D)

Price: £277.00 & **FREE Delivery** in the UK. [Details](#)

You Save: **£122.99 (31%)**

Only 10 (E) (F)

Sold by [L&J Photo](#) and Fulfilled by Amazon. Gift-wrap available.

Want it tomorrow, 17 June? Order it within **4 hrs 56 mins** and choose **One-Day Delivery** at checkout. [Details](#)

14 new from **£236.95** 3 used from **£275.00**

Style Name: **EF-S 18-55mm f/3.5-5.6 III Lens**

Roll over image to zoom in

Product details

Style Name: EF-S 18-55mm f/3.5-5.6 III Lens

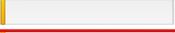
- Boxed-product Weight:** 1.4 Kg
- Batteries** 1 Lithium ion batteries required. (included)
- Item model number:** 9127B032AA
- ASIN:** B00IE3UR08
- Date first available at Amazon.co.uk:** 11 Feb. 2014
- Average Customer Review:**  (151 customer reviews)
- Amazon Bestsellers Rank:** **405 in Electronics** (G); Top 100 in Electronics)

Would you like to [update product info](#), [give feedback on images](#), or [tell us about a lower price?](#)

Customer Reviews

 151

4.6 out of 5 stars

5 star		115
4 star		24
3 star		4
2 star		4
1 star		4

(H) Share your thoughts with other customers

[See all 151 customer reviews](#)

(J) (K) Helpful Customer Reviews

(L) (M) 252 of 261 people found the following review helpful

 **A modest update to the EOS 1100D, but excellent value**

By (L) [REVIEWER](#) #1 HALL OF FAME on 15 Mar. 2014

Style Name: Body Only

I quite liked the EOS 1100D (but felt it was a bit dated despite being a bargain price) so was quite keen to try this to see what improvements Canon had made.

The answer is some, but probably not as many as I expected.

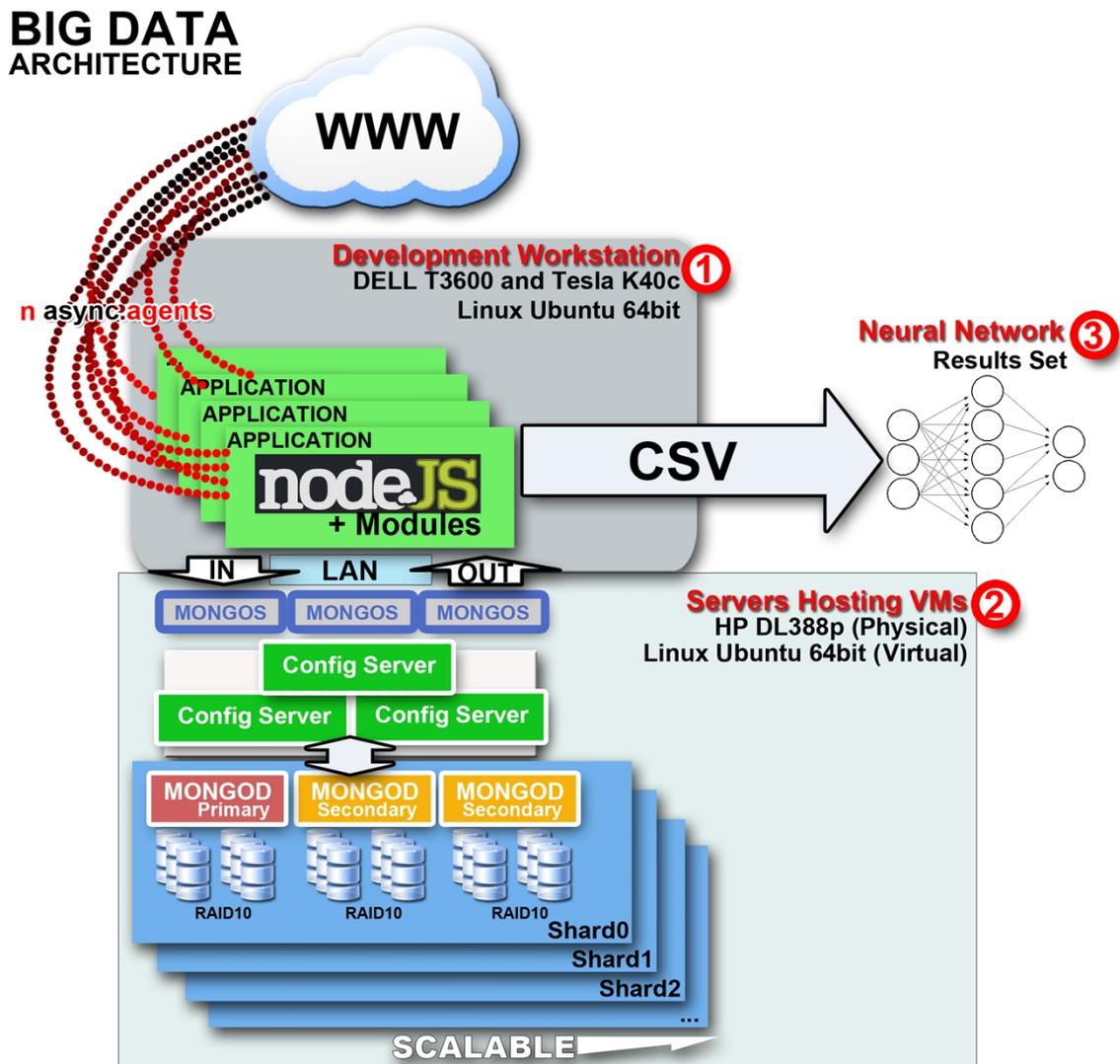
I'll list out the key areas where the EOS 1200D is improved over the 1100D

- + 18mp CMOS sensor (same as many other Canon's) up from the 12mp CMOS
- + Build does feel a bit better in the hand, the grip is now textured v the smooth one on the older model
- + The rear LCD is now 3" 460k dots v the 2.7" 230k EOS 1100D, quite a bit better, and sharper but not up

Note:

(A) = Customer Review Rating; (B) = Number of Customer Review; (C)= Number of Answered Questions; (D)= Free Delivery; (E) = Discount Value; (F) = Discount Rate; (G)= Sales Rank; (H) = Positive Reviews; (I) = Negative Reviews; (J) = Number of People Who Found the Most Helpful Review Useful; (K)= Total Number of People That Gave Feedback to the Most Helpful Review; (L) = Rating of the Most Helpful Review

Figure 3. Big Data Architecture



Note: Numbering describes the three part process. 1) Developmental workstation where our Node.JS agents are deployed for scraping the web using asynchronous I/O calls. 2) Physical server hosting Ubuntu 64bit virtual machines, and where data is stored and horizontally scaled. 3) Completed Web crawling and scraping datasets are converted into comma separated values for Neural Network analysis.

Table 1 Summary of variables used in this study

Variables	Description	Operationalization	Source
Free Delivery	If the product offers free delivery	Binary variable: 1 means the product offers free delivery in UK as listed in Amazon.com, and 0 otherwise	Doern and Fey 2006; Yip and Law 2002
Discount Value	The monetary value of the discount of the product	The fixed deductible amount of the product listed in Amazon.com, all in GBP	Gendall et al. 2006; Liu et al. 2015; Suri et al. 2000
Discount Rate	The percentage discount of the product	The percentage of the discount on the original price of the product listed in Amazon.com, all in %	Gupta and Cooper 1992; Madan and Suri 2001; Marshall and Leng 2002
Customer Review Rating (Valence)	The valence of customer review of the product	The average of all customer review ratings for the product, listed on Amazon.com	Chevalier and Mayzlin 2006; Chen et al. 2008; Cui et al. 2012; Duan et al. 2008; Forman et al. 2008; Lu et al. 2013; Zhu and Zhang 2010
Number of Customer Reviews (Volume)	The number of total customer reviews of the product	The listed total number of customer reviews of the product on Amazon.com	Chevalier and Mayzlin 2006; Chen et al. 2008; Cui et al. 2012; Duan et al. 2008; Davis and Khazanchi 2008; Forman et al. 2008; Liu 2006; Lu et al. 2013
Number of Answered Questions	The number of total answered questions of the product	The listed total number of answered questions of the product on Amazon.com	Ou et al. 2014
Positive Reviews	The percentage of 5 and 4 stars customer review	The percentage of 5 and 4 star customer reviews (number of 5 and 4 star customer reviews divided by the total number of customer reviews) listed on Amazon.com.	Chakravarty et al. 2010; Chevalier and Mayzlin 2006; Lu et al. 2013; Sonnier et al. 2011; Zhang et al. 2010
Negative Reviews	The percentage of 2 and 1 star customer review	The percentage of 2 and 1 star customer reviews (number of 2 and 1 star customer reviews divided by the total number of customer reviews) listed on Amazon.com. This measures the negative review.	Chakravarty et al. 2010; Chevalier and Mayzlin 2006; Lu et al. 2013; Sonnier et al. 2011; Zhang et al. 2010
Rating of the Most Helpful Review	The rating of the review that has the highest agreement on helpfulness.	The rating of the review that rates the product as 4 or 5 stars and has the highest agreement on helpfulness. In Amazon.com, people can choose if they think the review was helpful to	Mudambi and Schuff, 2010; Korfiatis et al. 2012; Chevalier and Mayzlin 2006;

them. The most helpful reviews are those with the smallest difference between the number of people who think it helpful and those who think it is not. Using the same criteria as Amazon.com, we define a review as favourable if its rating is 4 or 5 stars, and categorise a review as critical if its rating is between 1 and 3 stars. The most helpful favourable review is thus the favourable review that has the highest agreement on helpfulness, and this variable measures its rating.

Number of people who found the most helpful review useful	The total number of most helpful review	This variable measures the number of people who found the most helpful review useful to them.	Mudambi and Schuff 2010; Min and Park 2012
Total number of people that gave feedback to the most helpful review	The total number of people that expressed their opinion on the usefulness of the most helpful review	This variable measures the total number of people who gave feedback to the most helpful review (i.e. the total number of the people who think the review helpful and those who think the review not helpful).	Mudambi and Schuff, 2010; Korfiatis et al. 2012
Sales Rank	The Best Sellers Rank of the product	The ranking of the product provided by Amazon.com and this serves as a proxy of sales for this study.	Chevalier and Mayzlin 2006; Chen et al. 2008; Cui et al. 2012; Forman et al.2008; Sonnier et al. 2011

Table 2 Full validation results of neural network model

Network	Training	Testing
1	0.0103	0.0103
2	0.0102	0.0104
3	0.0103	0.0103
4	0.0103	0.0103
5	0.0102	0.0102
6	0.0103	0.0103
7	0.0104	0.0103
8	0.0100	0.0101
9	0.0102	0.0101
10	0.0102	0.0103
Mean	0.0102	0.0103

Table 3 Sensitivity Analysis

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	Importance
Volume*Discount Rate	0.22	0.10	0.06	0.17	0.22	0.11	0.12	0.19	0.15	0.17	15%
Number of Answered Questions	0.13	0.12	0.27	0.12	0.24	0.12	0.11	0.11	0.05	0.18	15%
Positive Reviews	0.06	0.06	0.13	0.04	0.20	0.10	0.08	0.06	0.06	0.08	9%
Total number of people that gave feedback to the most helpful review	0.07	0.08	0.14	0.07	0.11	0.08	0.11	0.08	0.06	0.04	8%
Negative Reviews	0.11	0.09	0.02	0.08	0.06	0.06	0.07	0.11	0.06	0.09	7%
Number of people who found the most helpful review useful	0.03	0.09	0.05	0.10	0.01	0.07	0.08	0.03	0.12	0.06	6%
Positive Review*Discount Rate	0.03	0.12	0.05	0.03	0.02	0.06	0.08	0.11	0.11	0.02	6%
Number of Customer Review (Volume)	0.06	0.06	0.06	0.04	0.03	0.11	0.07	0.07	0.09	0.04	6%
Customer Review Rating (Valence)	0.05	0.07	0.03	0.09	0.03	0.09	0.07	0.05	0.06	0.08	6%
Discount Rate	0.12	0.05	0.06	0.06	0.02	0.08	0.05	0.05	0.06	0.04	6%
Valence*Discount Rate	0.03	0.05	0.02	0.10	0.02	0.04	0.04	0.05	0.07	0.04	5%
Discount Value	0.04	0.04	0.07	0.04	0.01	0.03	0.05	0.05	0.04	0.08	4%
Rating of the Most Helpful Review	0.03	0.04	0.04	0.02	0.02	0.02	0.04	0.02	0.03	0.03	3%
Free Delivery	0.02	0.03	0.01	0.04	0.01	0.03	0.03	0.03	0.04	0.03	3%