International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 15, Issue 5, 2021

Sentiment Analysis through application of Big Data in online Retail Industry: A Conceptual Quantitative Study from the Perspectives of IT experts

Dr. Sadia Shaikh[a], Muhammad Faisal Sultan[b], Dr. Farhan Zeb Khaskhelly[c], Nida Zehra[d], [a]Assistant Professor, Benazir School of Business, Benazir Bhutto Shaheed University, Karachi, Pakistan. [b]Assistant Professor, Business Administration Department, KASBIT, Karachi, Pakistan, [c]Assistant Professor, IBA, University of Sindh, Jamshoro, Pakistan, [d]PhD Scholar, IBA, University of Sindh, Jamshoro, Pakistan, Email: [a]sadiakhurram@live.com

Big-Data is perceived as the source which may provide enormous to research as well as to business through influencing organizational performance. However, there are very few studies which highlighted the role of this evolving technology with the reference of Pakistan as low level of understanding of technology which is even highlighted by Gallop Pakistan. According to research the technology is significantly important for retail sector and might produce effective result for optimization of sales and growth. Hence there is a need to verify these postulates through relating the use of technology with Pakistan. Therefore, this study tried to conceptualize the impact of big-data application on the pricing mechanism of online retail stores of Pakistan through the opinion of IT experts. SMART-PLS has been used to analyze results of perception regarding the use of technology which significantly indicated importance of big-data for online retailers.

Key Words: *Big-Data, Sentiment Analysis, Online retailers & IT experts.*

INTRODUCTION

The dawn of Information Technology (IT) and data sciences is forcing companies to adapt quantitative methods for forecasting customers' needs in better manner. For e.g. it's much easy to understand demand through Radio Frequency Identification (Chong, Ch'ng, Liu & Li 2017). Similarly, Bollen Mao and Zeng (2011) quoted the example of big-data which is getting in lime light due to its significance and consideration given by executives and researchers.

Big Data can simply be defined as "A cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology" (Boyd & Crawford, 2012, p.663). Similarly, Thakuriah Tilahun and Zellner (2017) defined bigdata as "Structured and unstructured data generated naturally as a part of transactional, operational, planning and social activities, or the linkage of such data to purposefully designed data". (p. 14).

Through this companies have the liberty of information which was not accessible before although application of data in the field of business is not common. However, companies like Wal-Mart & Kohl etc are linking sales and economic data with demographic characteristics to predict consumer behavior (Chong et al., 2017). However, since last few decades' business are generating more data than ever before and thus, there is a requirement of way to use this extra ordinary amount of data in effective way (Fayyad, Piatetsky-Shapiro and Smyth, 1996 & Friedrich, Stoler, Moritz & Nash, 1983).

This can be better explained through quoting Mukherjee and Shaw (2016) that only YouTube has a viewership of around 4 Billion/day and approximately 48 minutes of video has been uploaded every. Similarly, 140 million active users of Twitter yield around 1 billion tweets. Need of electronic presence also resulted in generation of fifty plus website every minute & this enormous amount of data is termed as big-data. Big-data also has the capacity to use previous untouched data to develop insights that may aids in business and economic progression; thus, it is optimal to state that consumer analytics are actually is based on the combination of big-data & consumer behavior

(Sunil, Nobuyuki & Linda, 2015). However, the technology has negligible presence in developing side of the world (Kshetri, 2016), and one who wishes to explore need to take reference of developed sides of the world (Hajirahimova and Aliyeya, 2017b).

STATEMENT OF PROBLEM

Hybrid of Internet and big data result in massive convenience for effective analysis of data Chong et al., 2017) and technology is equally beneficial for all countries who wishes to increase their competitiveness (Hajirahimova and Aliyeya, 2017a). In fact, several countries are taking initiatives to improve national security, health facilities, economic and social conditions etc through big data (Makrufa & Hajirahimova, 2017a).

Upcoming economic and political contest between countries will relaying heavily on big data (Hajirahimova & Aliyeya, 2017b). However, in Pakistan there is severe lacking of digital format and only NADRA is equipped with advanced mechanism of bigdata but overall, they are also facing problem due to lack of data related to crime, education and healthcare (Ashraf, 2013).

Statements seems to be valid as the culture of Pakistan is much different in comparison to the developed sides of the world (Latif, Tunio, Pathan, Jianqiu, Ximei & Sadozai, 2018) and level of understanding regarding the technology is also much low (Gallup Pakistan, 2018). Though massive use of information technology is transforming an average user into a persistent generator of data hence increases the probability for the increase in overall data size which might be expand by 50% in coming times (IBM, 2012, Lycett, 2013 & Oracle, 2012).

Hence appropriate to treat big-data as basic feature of society and learning and analyzing through big-data is significantly associated with competition, growth, innovation and productivity (Jun & Park, 2017). However, there is significant lacking of studies associated with this field (Danah & Kate, 2012), especially in marketing lacking of quantities studies on application of big-data is resulting in formulation of inadequate and ineffective results (Aktas & Meng 2017).

However, effective planning might be made through capitalizing on enormous stream of data associated with online reviews and marketing activities. Hence through considering Aktas and Meng (2017) a quantitaive study on application of big-data on marketing activities might be massively significant especially to foster level of understanding regarding the technology in Pakistan (Gallup Pakistan, 2018).

THEORATICAL FRAMEWORK AND DELIMITATIONS

One of the top advantages of technology is accurate knowledge about the customers. In fact during current era even institutive decisions are based on big-data (Maheshwari, 2013). Legitimate to believe these statements as big-data application is not only helpful in online marketing but also aids in basic decision making e.g. consideration of cost, competitors price & value provided to customer (Baker, Kiewell & Winkler, 2014). However, Glass and Callahan (2014) indicated that steady and gradual progress must be made while linking data with company's decision making & strategy development etc. Although in Pakistan there is lacking of availability of data (Ashraf, 2013) as well as understanding of the concept of big-data (Gallup Pakistan, 2018).
However, it has been perceived that extensive use of IT applications would improve the supply chain performance as well as profitability of retail sector of Pakistan (Ali Subzwari and Tariq, 2016). Statements seems to be important as country is recently experiencing exponential growth in retail sector (Fazl-e-Haider 2018), which may use big-data to optimize their infrastructure, data sources and analytics etc (IBM, 2018). However, Ali et al. (2016) postulated that online retailing is still infancy in Pakistan and most of the retailers preferred deal in (FMCG) fast moving consumer goods (IBM, 2018).

Statement of Glass and Callahan (2014) added further complexities to the literature that adequate balance between creativity and analytical skills of data scientist is mandatory for effective decision making through big-data in the field of marketing. However, for effective decision making there is a need of quantitative analysis which may forecast these measures in more adequate manner (Aktas & Meng, 2017). Thus study is based on Pakistan through data of online retail industry to explore this infancy business (Ali et al., 2016), & uses pricing as the major resultant of big-data as it is one of the prime variable for retail sector (Aktas & Meng, 2017 & Grewal & Levy, 2007). Although to address Glass and Callahan (2014), study will also use availability of skilled data scientists as the moderating variable so support knowledge optimization as well as theory building.

SIGNIFICANCE

Schultz (2017) indicated that there will be an increase of 1.4 Billion in internet users till 2014 and total figures will be reaching 3.8 Billion with extensive increase of data generation from emerging markets (IDC, 2014). Although there is severe lacking of understanding regarding big-data in Pakistan (Gallup Pakistan, 2018) and conducting qualitative studies might not produces worthwhile results. Therefore, conduction of quantitative research work on big-data in emerging markets like Pakistan (New Desk, 2020) will be immensely beneficial for increase in understanding of researchers, academicians and professionals. Study also has a fold of significance towards entrepreneurship as the study is inclined towards online retailing which is also been preferred by entrepreneurs and new startups.

Hence this study has great deal of significance and must be treated as pervasive with respect to significance as it will address several lacking indicated by Aktas and Meng (2017) and Gallup Pakistan (2018).

LITERATURE REVIEW

The probability of increase in organizational data in upcoming times is getting higher and higher (Zanini & Dhawan, 2015).. Therefore, recent era is accompanied with recognition of the role big-data is playing in improving marketing and supply chain etc and also in support to the process of decision making (Rajeb, Rajeb & Keogh, 2020). There are some other benefits of deploying big-data i.e. betterment of internal processes & optimization of operational agility and flexibility (Santoro, Fiano, Bertoldi & Ciampi, 2019).

This form of data is also applicable social media to judge perception and views of customers and viewers while the field is termed as sentiment analysis which is also fruitful devising effective marketing strategies (Zanini & Dhawan, 2015). Statement seems to be valid as big-data is effective not only in capturing demographic details of online customers but also in designing of variouis functions like customization, pricing strategies and implementation of new point of sales (Santoro et al., 2019). Though this form of data is unstructured as well as complex in comparison to the others forms of data therefore firms must need to install sophisticated technological resources so to assess and analysis of data effectively (Rajeb, Rajeb & Keogh, 2020). Big-data is also found to be tempting for retail sector (Bradlow, Gangwar, Kopalle & Voleti, 2017).

Therefore, upcoming paragraphs will present impact of big-data on pricing mechanism of online retail business as this form of retailing is still infancy in Pakistan (IBM, 2018). Big-data technology will make retailers understand reasons behind the sales lift and aids them in optimization of their promotional campaigns more effectively. (Aktas & Meng, 2017). However, pricing is perceived as one of the most technical variable in retail business as every stock keeping unit (SKU) has its own demand and competition (Grewal & Levy, 2007)

However, use of big-data might make retailers understand level of elasticity in demand (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011) & may leads to dynamic pricing on the bases of demand and competition (Aktas & Meng, 2017 & Clark & Vincent, 2012). Adding further Aktas and Meng (2017) quoted example of well-known retailers e.g. Sainsbury's, Tesco, ASDA & Morrison's etc which are attaining competitive edge through big-data technology Similarly, Thau (2016), indicated that bigdata is fruitful for customizing sales promotion campaigns from retailers and may highlight most suited products to be placed on sales. In fact, big-data make retailer understand different market segments which is the key for retailers in selecting most suited target markets and pricing for different products. On the other side study of Iqbal Kazmi Manzoor Soomrani Butt and Shaikh (2018) indicated severe lacking of in-house data management specialists & the prime reason for shortage is uncertain return on investment from big-data analytics. However, lacking of qualified data scientist and high cost of staffing are also included in the list of potent reasons for the shortage of data management specialists. On the other side Dolezel and McLeod (2019) posited that have adequate inventory of skills from data scientist is also mandatory for adequate extraction of knowledge. In large size firms this activity has been done through job designing in a way which may accomplish the desired goal through breaking it into group tasks for IT personnel Though small and medium sized enterprises (SME) faces lot of difficulties in managing the same as they need cross sectional experts to deal with prevailing issues in the domain of business and IT (Iqbal et al., 2018).

RESEARCH METHODOLOGY

Research Design

Research design is a technique that aids in achieving answers to important research questions, especially to those in which researchers have keen interest (Oso & Onen, 2009). However, linking research design with research onion highlighted epistemology as the best sort of philosophy for this form of studies (Saunders, Lewis & Thornhill, 2007),

The philosophy was initially postulated by Guba and Lincon (1994) as logic which indicate what would be knower and what can be known? Postulated was support by Cohen Manion and Morrision (2007) that epistemology is associated with forms of knowledge building. However, to proceed towards answers there is a requirement of philosophical stance that to be accompanied with research philosophy as these stances are required to determine most effective method data for collection and analysis (Zukauskas & Vveinhardt & Andriukaiteiene, 2018). For this study stance is realism as it is adjustable with qualitative as well as quantitative designs (Saunders, Lewis & Thornhill 2015). Moreover, according to Saunders et al. (2007) strategy used in conduction of this study is experiment more specifically field experiment (Sekaran & Bougie, 2016). Method of data collection is mono method (Saunders et al., 2007) and unit of analysis is individual while time horizon for the collection of data is cross sectional (Sekaran & Bougie, 2016).

Sampling Design

Mugenda (2003) postulated that sampling design is the strategy by which one can understand the reason behind the selection of specific units in the study. These postulates are further clarified by Leedy and Ormrod (2005) that sample for the study must be coherent with the intended objectives and also must decrease overall cost of data collection. However, application of IT in retail sector is still in infancy stage (Ali et al., 2016) therefore, non-probability sampling has been used order to select respondents on their proximity to research (Jager, Putnick & Bornstein, 2017).

The method of sampling employed is quota sampling as its best when response rate is slow and excessive sampling can increase cost of sampling (Yang & Banamah, 2014). Although the online trade is almost 5% of the total retail trade in Pakistan (IBM, 2018) and concept of big-data is also in initial phases (Gallop Pakistan, 2018). Therefore, in accordance with Pathirage, Amaratunga and Haigh (2008) study is associated with theory building approach and the total sample size of the study is of fifty (50) respondents. All the respondents of the study belong to IT departments' of online retail sector & as mentioned earlier the sampling is quota sampling which is based on association with IT department of online retailer.

The use of IT representative as quota sampling has also been legitimize by Czaja Charness Fisk Hertzog Nair Rogers and Sharit (2006), that successful adoption of technology is significantly important for independence of any function. Hence the IT personnel are the most suited for understanding regarding the technology and its use.

Questionnaire

Study is done through the use of closed ended questionnaire which is adapted from Ducange et al (2017) and Le and Liaw (2017) however elements (questions) are refined in order to make these coherent & applicable. Moreover, measures indicated by Aktas and Meng (2017) and Valchanov (2017) are also been linked with study. Thus, through these including elements from quantities studies and parameters from qualitative study a systematic closed-ended questionnaire was formulated for the conduction of this study.

STATISTICAL TESTING AND ANALYSIS

In modern times SMART-PLS has been preferred by several researchers to device studies in the domain of management sciences (Benitez, Henseler, Castillo & Schuberth, 2020). Software uses two forms of models i.e. structural model and measurement model (Afthanorhan, 2014), while measurement models are further bifurcated into reflective and formative models (Benitez et al., 2020). Software also has the ability to link descriptive as well as inferential statistical measures to the both the bifurcations of measurement models (Hair, Risher, Sarstedt & Ringle, 2019). For this study the reflective measurement model has been developed after thorough literature review and incorporation of elements indicated by prior research work and therefore will also follow indications of Afthanorhan (2014) and Benitez et al. (2020).

Table 1 is placed to indicate the authenticity of outer loadings for each element associated with the construct of big-data on pricing strategies of online retailers. On the other hand, Hair Sarstedt Ringle and Mena (2012) indicated that in order to be valid with respect to outer loading each element must have 0.708 or greater values. Though minimum acceptable value of outer loading is 0.60 as indicated by Afthanorhan (2014) while table 1 indicated 0.687 as the least value therefore in accordance with Afthanorhan (2014) there is no need to delete any of the element.

Outer Loadings

	Big-Data	Moderating Effect 1	Pricing Strategies	Skilled Data Scientists
BD1	**0.864**			
BD2	**0.878**			
BD3	**0.935**			
BD4	**0.940**			
BD5	**0.917**			
PS1			**0.745**	
PS2			**0.890**	
PS3			**0.828**	
PS4			**0.892**	
PS5			**0.705**	
SD1				**0.687**
SD2				**0.875**
SD3				**0.776**
SD4				**0.871**
SD5				**0.716**
Skilled Data Scientists * Big-Data		**1.129**		

Table 1: Outer loading

R Square

	R Square	R Square Adjusted
Pricing Strategies	0.695	0.647

Table 2: Predictive Accuracy (Quality Criteria)

Table 2 highlights extent of variance caused by independent variable. The table is termed as predictive accuracy and is used to indicate variance caused by independent variable through applying the logic of ordinary least square (Benitez et al., 2020). Analysis is similar as of analysis for regression (Andreev, Heart, Moaz & Pliskin ,2009) and minimum acceptable value to assure quality criteria is 0.26 for R^2 while 0.50 and 0.75 or above are treated as moderate and extensive model fit (Cheah, Memon, Chuah, Ting & Ramayah, 2018). Though table 2 highlighted that the value of R^2 is 0.695 which is moderate fit in terms of predictive accuracy (Henseler Ringle & Sinkovics, 2009 & Hair Ringle & Sarstedt, 2013).

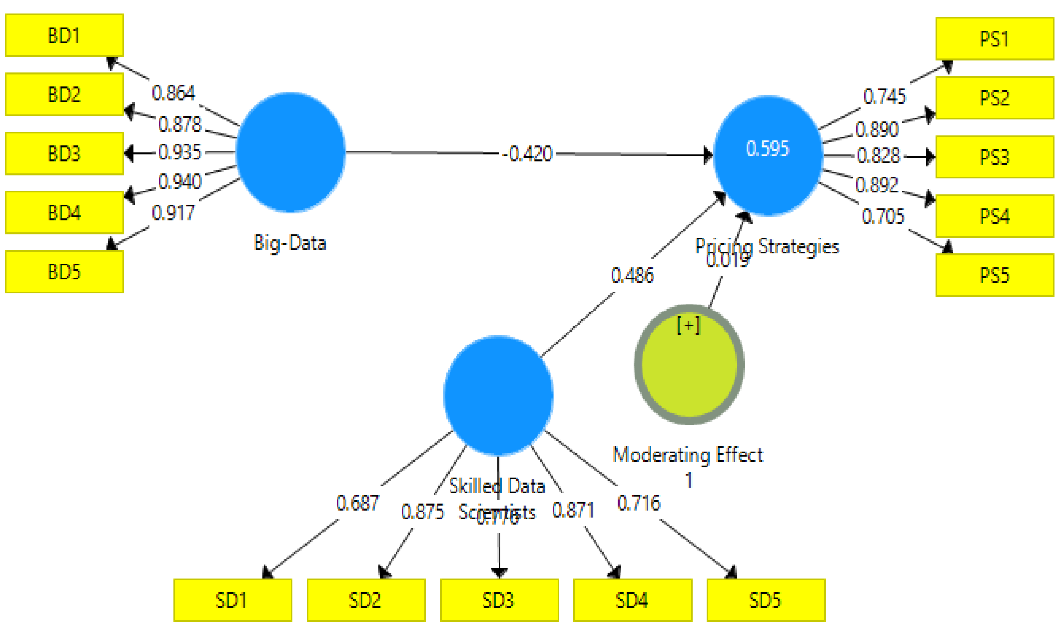

Figure 1: Outer Loadings & confirmatory factor analysis (CFA)

Construct Reliability and Validity

	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
Big-Data	0.946	0.949	0.959	0.823
Moderating Effect 1	1.000	1.000	1.000	1.000
Pricing Strategies	0.871	0.873	0.908	0.665
Skilled Data Scientists	0.846	0.850	0.891	0.622

Table 3: Construct reliability & convergent validity

Table 3 is highlighting construct reliability as well as convergent validity (Ab Hamid, Sami & Sidek, 2017 & Sijtsma, 2009 a&b).

Table also highlights reliability measures i.e. Cronbach's Alpha (α) and Goldstein rho, the table is also indicating convergent validity which is a measure of association which indicates the extent to which variables are measuring the same construct (Benitez et al., 2020). However, to show convergent validity researchers must indicate outer loadings, composite reliability and (AVE) average variance extracted (Sijtsma, 2009a&b). Outer loadings are already mentioned in table 1 and composite reliability and AVE are treated as sufficient tools to pose convergent validity (Sijtsma, 2009a&b). Table also indicated 0.7 or above values for all the forms of reliability measures and AVE for all the measures is more than 0.5, which ensures construct reliability as well as convergent validity.

Heterotrait-Monotrait Ratio (HTMT)

	Big-Data	Moderating Effect 1	Pricing Strategies	Skilled Data Scientists
Big-Data				
Moderating Effect 1	0.193			
Pricing Strategies	0.685	0.107		
Skilled Data Scientists	0.472	0.364	0.776	

Table 4: Discriminant Validity Via Hetrotrait-Monotrait Ratio

Table 4 is used to indicate discriminant validity through heterotrait-monotrait ratio (HTMT). HTMT is the ratio which highlights lack of correlation among the variables associated with same construct (Cheung & Lee, 2010). HTMT is also perceived as most preferred tool to highlight discriminant validity (Benitez et al., 2020). 0.85 is the peak acceptable value for the ratio as indicated by Hair Jr Sarstedt Ringle and Gudergan (2017), while in table 4 the top most value for the ratio is 0.776. Thus, in the light of these measures HTMT is appropriate enough for the study.

Path-Coefficient

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	P Values
Big-Data -> Pricing Strategies	-0.420	-0.416	0.053	7.865	**0.000**
Moderating Effect 1 -> Pricing Strategies	0.019	0.020	0.029	0.640	**0.523**
Skilled Data Scientists -> Pricing Strategies	0.486	0.491	0.053	9.190	**0.000**

Table 5: Path Coefficient (Total Effect)

Table 5 indicating path-coefficient to highlight impact of big-data on the pricing strategies of online retailers. The table is a tool highlight inferential statistics to regress hypotheses testing and analysis of relationship between variables. Hair et al. (2019) indicated that inferential statistics is one of the major criteria associated with the measurement model in SMART-PLS. Software uses t-statistics (Durate & Amaro, 2018) and p-values (Kock & Hadaya, 2018) to reflect inferential statistics in measurement models.

The minimum t-value required to indicate relationship between variables is 1.97 (Hair et al., 2011) and the maximum p-value which is acceptable to indicate relationship is 0.05 (Kock & Hadaya, 2018). Table 5 indicating that the t-value for moderation effect is (0.640) and p-values is (0.523), which are not adequate to regress the moderation effect. Thus, in the light of parameters of Hair et al (2011) and Kock and Hadaya (2018) it is optimal to believe that there is no moderation of skilled data scientist. However, big-data is still perceived as the potent tool for devising pricing strategies of online retailers.

DISCUSION AND MANAGERIAL IMPLICATIONS:

It has been mentioned earlier that uses of IT application by retailers in Pakistan is in infancy stage as highlighted by the Ali et al (2016) Therefore there is a severe lacking of studies on use of big-data for IT sector which resulted in severe lacking of understand regarding application of big-data. Thus most of the retailers are applying bricks and mortars the strategy of business (IBM, 2018 & Gallop Pakistan, 2018) hence this study uses the reference of studies conducted in west for e.g. Aktas & Meng (2017); Bradlow et al (2017); Ducange et al (2017); Le and Liaw (2017) and Seetharaman Niranjan Tandon and Saravanan (2017) to formulate research model. The model indicated that big-data is perceived as the potent tool for the optimization of pricing strategies as indicated by Aktas & Meng (2017), Clark and Vincent (2012) and Thau (2016). Similarly model also indicating the significant impact of availability of skilled data scientist in gaining insight regarding pricing strategies and pricing mechanism. Findings are found to be consistent with Glass and Callahan (2014).

Although the moderation of availability of skilled data scientist is not found valid according to the statistical testing which indicated the lack of importance of skilled data scientist for online retailers? Therefore, the findings are found to be inconsistent with Dolezel and McLeod (2019), which might mean online retail sector is still infancy (Gallop Pakistan, 2018) thus might not have large customer base & massive sales. Legitimized by the concept of product life cycle presented in one of the initial studies on the concept by Rink and Swan (1979), hence business activities might not be extensively requiring intensive skills of data scientists.

AREA FOR FUTURE RESEARCH

This research is one of the primaries which explore application of big-data in the context of marketing and sentiment analysis through online retail sector of Pakistan. However, the online retailing is not the major form of retailing and hence future studies might be linked application in the context of retailers having bricks and mortars as the business philosophy.

However, the data collection from stores using bricks and mortars as the philosophy is much difficult as these retailers are not much inclined towards sentiment analysis. The statement is appropriate to believe as Salvador and Ikeda (2014) indicated that effective information system is a prerequisite of effective decision making. Thus, quantitative studies became much difficult when is not available regarding preference, purchase intensity and preferred location. Although qualitative work through emphasizing on the role POS terminals and dumb terminals might provide reasonable knowledge in this area. However, if one wishes to work quantitatively then effective information system might be used as mediating variable to check impact of big-data on other marketing implications e.g. store layout and assortment. Last but not the least more work might be done through checking role of bigdata on other industries like insurance and banking to gauge broader impact of big-data.

REFERENCES

Ab Hamid, M. R., Sami, W., & Sidek, M. M. (2017, September). Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012163). IOP Publishing

Afthanorhan, W. M. A. B. W. (2014). Hierarchical component using reflective-formative measurement model in partial least square structural equation modeling (Pls-Sem). *International Journal of Mathematics*, *2*(2), 33-49

Aktas, E., & Meng, Y. (2017). An Exploration of Big Data Practices in Retail Sector. *Logistics*, *1*(2), 12

Ali, R., Subzwari, M., & Tariq, S. (2016). Impact of Information Technology on Retail Sector in Pakistan. *KASBIT Journal of Management & Social Science*, *9*(1), 63-93.

Andreev, P., Heart, T., Maoz, H., & Pliskin, N. (2009). Validating formative partial least squares (PLS) models: methodological review and empirical illustration. *ICIS 2009 proceedings*, 193

Ashraf, S., (2013), "Can we Optimize Pakistan via Big Data?", Tech Juice, https://www.techjuice.pk/can-optimize-pakistan-via-big-data/

Baker, W., Kiewell, D., & Winkler, G. (2014). Using big data to make better pricing decisions. *McKinsey Analysis. McKinsey & Company, https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions#*

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, *57*(2), 103168

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, *93*(1), 79-95.

Brown, B. Bughin, J., Byers, A.H.; Chui, M.; Dobbs, R.; Manyika, J. and Roxburgh, C., (2011), "Big Data: The Next Frontier for Innovation, Competition, and Productivity", *Technical Report for Mckinsey & Company*, Washington, DC, USA

Cheah, J. H., Memon, M. A., Chuah, F., Ting, H., & Ramayah, T. (2018). Assessing reflective models in marketing research: A comparison between PLS and PLSC estimates. International Journal of Business and Society, 19(1), 139-163

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, *55*(17), 5142-5156, doi: 10.1080/00207543.2015.1066519

Clark, R., & Vincent, N. (2012). Capacity-contingent pricing and competition in the airline industry. *Journal of Air Transport Management*, *24*, 7-11.

Cohen, L. M., & Manion, L. (2001). l. & Morrison, K. (2007). *Research methods in education*, *6*

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and aging*, *21*(2), 333.

Danah, B., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662-679.

Dolezel, D., & McLeod, A. (2019). Big Data Analytics in Healthcare: Investigating the Diffusion of Innovation. *Perspectives in health information management*, *16*

Ducange, P., Pecori, R., & Mezzina, P. (2018). A glimpse on big data analytics in the framework of marketing strategies. *Soft Computing*, *22*(1), 325-342.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37.

Fazl-e-Haider, S. (2018 February 19). Booming Growth in Pakistan's Retail Sector. Economist, http://www.pakistaneconomist.com/2018/02/19/booming-growth-pakistans-retail-sector/

Friedrich, O. Stoler, P. Moritz, M. & Nash, J. N. (1983). Machine of the year: The computer moves**.** *Time Magazine, 121(1).* 14–28

Gallup Pakistan, (2018), "Big Data Analysis Reports". Retrieved from http://gallup.com.pk/polls/gallup-history-project/big-data-analysis/

Glass, R., & Callahan, S. (2014). *The Big Data-driven business: How to use big data to win customers, beat competitors, and boost profits*. John Wiley & Sons.

Grewal, D., & Levy, M. (2007). Retailing research: Past, present, and future. *Journal of retailing*, *83*(4), 447-464.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications

Hair Jr, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2017). *Advanced issues in partial least squares structural equation modeling*. sage publications

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning*, *46*(1-2), 1-12

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31(1), 2-24*

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the academy of marketing science*, *40*(3), 414-433

Hajirahimova, M. S., & Aliyeva, A. S. (2017a). About Big Data Measurement Methodologies and Indicators. *International Journal of Modern Education and Computer Science*, *9*(10), 1

Hajirahimova, M. S., & Aliyeva, A. S. (2017b). Big Data Initiatives to Developed Countries. *Problems of information* society, 1, 10-19, doi: 10.25045/jpis.v08.i1.02

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing*. Emerald Group Publishing Limited.

IBM Institute of Business Value IBM, (2018), "Analytics: The real-world use of big data in retail**"**, *How innovative retailers extract value from uncertain data,* https://www-935.ibm.com/services/us/gbs/thoughtleadership/big-data-retail/

IDC, (2014), "Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives", Retrieved from https://www.emc.com/leadership/digital-universe/2014iview/executivesummary.html

Integreon Insight (2012). "Big just got bigger". Grail Research, 1–17 (Retrieved from http://www.integreon.com/pdf/Blog/Grail-Research-Big-Data-Just-Got-Bigger_ 232.pdf)

Iqbal, M., Kazmi, S. H. A., Manzoor, A., Soomrani, A. R., Butt, S. H., & Shaikh, K. A. (2018, March). A study of big data for business growth in SMEs: Opportunities & challenges. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-7). IEEE

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development*, *82*(2), 13-30

Jun, S. P., & Park, D. H. (2017). Visualization of brand positioning based on consumer web search information: Using social network analysis. *Internet Research*, *27*(2), 381-407.

Khan, A., (2017, July 31), "Trading data", Retrieved from https://www.thenews.com.pk/magazine/money-matters/194614-Trading-data

Kshetri, N. (2016). *Big data's big potential in developing economies: impact on agriculture, health and environmental security*. CABI. Wallingford, doi 10.1079/9781780648682.0000, ISBN 9781780648682

Latif, Z., Tunio, M. Z., Pathan, Z. H., Jianqiu, Z., Ximei, L., & Sadozai, S. K. (2018, March). A review of policies concerning development of big data industry in Pakistan: Subtitle: Development of big data industry in Pakistan. In *Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on* (pp. 1-5). IEEE.

Le, T. M., & Liaw, S. Y. (2017). Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-Commerce Context. *Sustainability, 9*(5), 798.

Leedy, P. D., & Ormrod, J. E. (2005). Practical Research Planning and Design . New Jersey: Pearson Merrill Prentice Hall.

Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems, 22(4)*, 381–386

Maheshwari, A. (2014). *Business Intelligence and Data Mining*. Business Expert Press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *Mckinsey & Company*, Washington, USA,

Mugenda, A. (2003). Research methods Quantitative and qualitative approaches by Mugenda. *Nairobi, Kenya*.

Mukherjee, S., & Shaw, R. (2016). Big Data–Concepts, Applications, Challenges and Future Scope. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(2), 66-74.

New Desk. (2020, May 13). Profit, https://profit.pakistantoday.com.pk/2020/05/13/pakistan-retains-its-place-in-msci-emerging-markets-index/

Oracle (2012). Big data for the enterprise. Oracle White Paper, 1–14, http://www.oracle.com/us/products/database/big-data-forenterprise- 519135.pdf.

Oso, W. Y., & Onen, D. (2009). *A general guide to writing research proposal and report*. Jomo Kenyatta Foundation.

Pathirage, C. P., Amaratunga, R. D. G., & Haigh, R. P. (2008). The role of philosophical context in the development of research methodology and theory. *The Built and Human Environment Review*, *1*(1), 1-10

Pecori, R. (2016). S-Kademlia: A trust and reputation method to mitigate a Sybil attack in Kademlia. *Computer Networks*, *94*, 205-218.

Rink, D. R., & Swan, J. E. (1979). Product life cycle research: A literature review. *Journal of business Research*, *7*(3), 219-24

Salvador, A. B., & Ikeda, A. A. (2014). Big data usage in the marketing information system. *Journal of Data Analysis and Information Processing*

Santoro, G., Fiano, F., Bertoldi, B., & Ciampi, F. (2019). Big data for business management in the retail industry. *Management Decision, 57(8), 980-1992*

Saunders, M. N. K., Lewis, P., Thornhill, A., & Bristow, A. (2015). Understanding research philosophies and approaches: Research methods for business students

Saunders, M., Lewis, P. & Thornhill, A. (2007). Research methods. *Business Students 4th edition Pearson Education Limited, England*

Schultz, J., (2017, October 10), "How Much Data is Created on the Internet Each Day?" https://blog.microfocus.com/how-much-data-is-created-onthe-internet-each-day

Seetharaman, A., Niranjan, I., Tandon, V. and Saravanan, , A. S., (2016), "Impact of Big Data on the Retail Industry" *Journal of Corporate Ownership & Control*, 14(1), 506-518

Sekaran, U. and Bougie, R., (2016), "Research Methods For Business: A Skill Building Approach", *John Wiley & Sons*, 1-448, ISBN 1119165555, 9781119165552

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107

Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *Psycholoog*, *44*(11), 561

Thakuriah, P. V., Tilahun, N. Y., & Zellner, M. (2017). Big data and urban informatics: innovations and challenges to urban planning and knowledge discovery. In *Seeing cities through big data* (pp. 11-45). Springer, Cham.

Thau, B., (2016, March 2), "Retail pricing strategies getting a makeover from data analytics", IBM Big Data and Analytics Hub, https://www.ibmbigdatahub.com/blog/retail-pricing-strategies-getting-makeover-data-analytics

Valchanov, I. (2017). Is data science really a rising career? https://www.quora.com/Is-data-science-really-a-rising-career/answer/IliyaValchanov

Voleti, S., Kopalle, P. K., & Ghosh, P. (2015). An inter-product competition model incorporating branding hierarchy and product similarities using store-level data. *Management Science*, *61*(11), 2720-2738.

Zanini, N., & Dhawan, V. (2015). Text Mining: An introduction to theory and some applications. *Research Matters*, *19*, 38-4

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

Žukauskas, P., Vveinhardt, J., & Andriukaitienė, R. (2018). Philosophy and paradigm of scientific research. *Management Culture and Corporate Social Responsibility*, 121.