



A NOVEL APPROACH OF SOCIAL MEDIA ANALYTICS FOR PREDICTING NATIONAL CONSUMER CONFIDENCE INDEX

MUHAMMAD ASHRAF¹, ARSLAN ALI RAZA², MUHAMMAD ISHAQ³

ABSTRACT

Consumer confidence captures the difference in consumer attitudes when making buying decisions. The consumer confidence index (CCI) is referenced by businesses, governments and other institutions when they make strategic decision. Analysis of Pakistani official consumer confidence index in comparison with the actual data reveals inconsistency between the official index and actual state of the economy in Pakistan. The data used in computing the official CCI has been collected using cross-sectional survey which has inherent limitations which spur research efforts to clarify and develop evidence on associations that form the core of the economic indicator of consumer confidence. It poses challenges to industry with regards to measuring the CCI as it is subjected to rigorous analysis. One aspect of such dynamic has to do with the emergence of new sources of digital data that could be used to measure consumer confidence. However, it is conjectured that using consumer sentiment data on social media might show greater marginal significance for consumer confidence, because the surveys might capture effects that will not appear in the data. Social media can offer a huge volume of data on consumer confidence, the analysis of which can be conducted at a more rapid time and can also refine the accuracy of the CCI using data from far larger populations. However, this study aims to propose a novel method of social media analytics for predicting national consumer confidence index (NACOP) that reflect true state of the economy. The proposed NACOP utilizes big data and data science to predict the national consumer confidence using large data sets of purchasing behaviour, jobs/employment, consumer price and personal finance from social media platforms. The study also explains the architecture of proposed NACOP and provides significant implications to academicians and practitioners.

Keywords: Big data, Consumer confidence index, NACOP, Sentiment analysis, Social media analytics,

JEL Codes: L82

I. INTRODUCTION

Global financial crisis of 2008 has been considered the most financial crisis after the Great Depression (Pendery, 2009). Researchers and professionals identified that erosion of consumer confidence causes longevity of the crisis (Cheng, 2019; Petev et al. 2011), indicating the importance of consumer confidence in economy. According to Behavioral Economics, consumers expenditures depend on not only their financial conditions but also their attitudes reflecting in consumer confidence. A notable indication of the economic prospects of an economy is the level of its consumers' confidence in the economy. In mainstream economic discussions, this is expressed as a quantified metric called the Consumer Confidence Index [Hereafter, it is called "CCI"]. Consumer confidence captures the difference in consumer attitudes when making buying decisions, while keeping the financial condition and surrounding business environment constant. It is a crucial data that it is referenced by businesses, governments and other institutions when they make announcements about the strategy direction their institution is taking or is about to pursue. Even in mainstream business media channels (e.g., Bloomberg), the CCI is referenced and makes headline news. The consumer confidence reflected in their ability to spend on goods and services is a cue of the direction of economic activities, whether they feel they have or will have the economic means to buy goods and services they require. These in turn influence the manufacturing of goods and delivery of services, affecting the levels of labor employment, the income tax generated for the government, profits or loss enjoyed by businesses and the commitment they go on to

¹ Department of Management Sciences, COMSATS University Islamabad, Vehari Campus

² Department of Computer Science, COMSATS University Islamabad, Vehari Campus

³ Department of Mathematics, COMSATS University Islamabad, Vehari Campus

make on capital investment. The CCI information influences business strategy decisions made by manufacturers, retailers, banks and governments, relative to other key economic indicators (Investopedia, 2014; Ali, 2015; Ali, 2018; Ali and Bibi, 2017; Ali and Ahmad, 2014; Ali and Audi, 2016; Ali and Audi, 2018; Ali and Rehman, 2015; Ali and Senturk, 2019; Ali and Zulfiqar, 2018; Ali et al., 2016; Ali et al., 2021; Ali et al., 2021; Ali et al., 2015; Arshad and Ali, 2016; Ashraf and Ali, 2018; Audi et al., 2022; Audi and Ali, 2017; Audi and Ali, 2017). A consistently declining trend in CCI would indicate that consumers have a negative economic outlook which implies further decline in economic activity. While a positive CCI trend would be taken as antecedent to growth in economic activity in the near future which is a good sign for business. By analyzing the correlation between different economic indicators, businesses could base their decision of either investing in production capacity or cut labor costs, based on the anticipated economic activity inferred from the level of consumer confidence. Similarly, banks can also take measures to curb the anticipated increase in borrowing by consumers and the use of credit cards. The government could come up with measures such as reduction of taxes in order to encourage spending. The opposite will apply in a case of an upward trend in the index (higher production/inventories and tax increase).

From 1921 to 1933, the United States government adopted a policy of laissez-faire which required them to be less involved in the affairs of the economy (however, there was no CCI at this time). They achieved this by keeping taxes low thereby encouraging businesses to invest more money for expansion and also creating more money for consumers to spend (BBC, 2014; Igboayaka, 2015). Therefore, it seems reasonable to conclude that CCI influences decisions made by pivotal economic entities such as governments. It could also be argued that the CCI as an economic indicator has been considered a 'reliable' way of predicting whether economic activity will grow or decline in the near future. For example, the Conference Board of Canada has been publishing a CCI since 1980. During the past 41 years, despite the creation of other metrics to measure the consumers' confidence, the CCI is still referred to each time it is published (Igboayaka, 2015). Nonetheless, the reliability of the CCI is directly linked to how accurately its methods of measurement capture the economic sentiment of the population of a country. In this regard, the methods of measurement start with a manual random survey of different demographics of the population. These surveys are known as Consumer Confidence Surveys [Hereafter, it is called "CCS"] and they are distributed by postal mail, electronic mail and telephone campaigns. Predicting Consumer Confidence is challenging. The industry challenge is that the official Consumer Confidence Index (CCI) published by State Bank of Pakistan may not reflect the true state of Pakistani economy. In Pakistan, official CCI covers households' perceptions relative to the last six months and their expectations for the next semester, including personal financial conditions, unemployment, consumption of durable goods and overall economy. The index is based on the stratified random telephonic survey of 1600 households across Pakistan (Economic Trending, 2019; State Bank of Pakistan, 2021). The survey is conducted by the Institute of Business Administration (IBA), Karachi and State Bank of Pakistan (SBP) every two weeks since January of 2012 (Economic Trending, 2019; State Bank of Pakistan, 2021). Analysis of the official CCI in comparison with the actual data reveals that as official CCI decreases (from 44.27 to 41.92 points), actual Consumer Spending (10626857 to 11101050 M PKR), Consumer Credit (680499 to 684676 M PKR), Gasoline Prices (0.73 to 0.74 USD/Litre) and Private Sector Credit (5973865 to 6003140 M PKR) increases, indicating inconsistency between official CCI and actual state of the economy in Pakistan. However, the official CCI seems to be incapable of reflecting true condition of the economy consequently it provides little value (or no information) to industry player and end users. It might be due to inherent limitations in the way its data is collected and computed. The data used in computing the official CCI has been collected using cross-sectional survey which has limitations. First, questionnaire-based survey may not reflect the real-life consumer behavior. Second, CCI Survey is performed bi-monthly, lack of timely information access for banks, retailers to react to volatility or changes in consumer behaviour. Third, sample size is limited to 1,600 household per survey, thus making it difficult to extrapolate and may not be representative to the overall population. Fourth, questions are speculative and subjective in nature. Consumers are easily influenced by myriads of national events that took place. In spite of the limitations, the official CCI figure by the State bank of Pakistan is the only accepted indicator that the industry and key players can rely on, while the official CCI method is based on consumer perception that is highly subjective. It would hardly to find that confidence surveys reflect the current state of consumer purchasing behavior and the overall economy (Shayaa et al., 2017; Audi et al., 2021; Audi and Ali, 2016; Audi et al., 2021; Audi et al., 2021; Audi et al., 2021; Haider and Ali, 2015; Kaseem et al., 2019; Roussel et al., 2021; Sajid and Ali, 2018; Senturk and Ali, 2021; Mehmood et al., 2022). However, it poses challenges to industry with regards to measuring the CCI and these challenges are as follows:

II. DATA FREQUENCY AND TIMELINESS

- Official CCI figures are published quarterly, which does not allow for end users/industry players to make accurate prediction or planning.

- There is a need to publish weekly or monthly CCI indicator for industry players and end users to facilitate their strategic business activities (e.g., marketing plans, financial forecasting, investments plans, etc.)

III. QUALITY OF RESPONDENTS

- Most questionnaire surveys are conducted via phone call or via face-to-face interviews by a third party or companies. And in most cases, the phone call lands to the similar demographics; i.e house wives, civil service, and students.
- Results are highly biased and many not represent the collective view of the consumer confidence at the national level.

IV. LIMITED SAMPLE SIZE

- Due to the laborious and time consuming approach via phone calls and face-to-face interviews, the sample size can only be limited to 1,000 – 2,000 people, or households.
- To extrapolate CCI values to a larger and diverse population in Pakistan, the result may not be representative or accurate.

V. LACK OF ALTERNATIVE DATA SOURCE

- There is no alternative consumer confidence metrics that truly reflect the condition of the economy.

Meanwhile consumers' activities in Pakistan provide exciting opportunities to leverage on publicly available data on social media. Pakistan has a population of over 212.7 million, 65.5 million labor workforce, \$256.66 billion consumer spending, 163 million mobile penetration, 65.13 million Internet users and 37 million social media users (87.68% Facebook, 4.63% twitter, 2.82% Pinterest, 2.19% Youtube & 1.96% Instagram). In digital era, data flow is growing faster than trade and finance as indicated by Hirt and Willmott (2014). However, it is conjectured that using consumer sentiment data on social media might show greater marginal significance for consumer confidence, because the surveys might capture effects that will not appear in the data. Social media can offer a huge volume of data on consumer confidence, the analysis of which can be conducted at a more rapid time and can also refine the accuracy of the CCI using data from far larger populations. This study aims to propose an alternative method of social media analytics tool for predicting national consumer confidence index that reflect true state of the economy. This method is named as "*National Consumer Confidence Predictor*" [Here after it is called "**NACOP**"]. The NACOP utilizes big data and data science to predict the national consumer confidence using large data sets of purchasing behaviour, jobs/employment, consumer price and personal finance from social media platforms. NACOP publishes consumer confidence data based on a set of consumer behavior parameters similar to the traditional consumer confidence index (CCI) approach. NACOP figures will act a leading indicator which reduces time-to-production, whilst waiting for the official CCI figures. However, the overall purpose of the study is to propose NACOP using social media analytics for measuring and predicting consumer confidence in Pakistan.

VI. LITERATURE REVIEW

VII. ESTABLISHED METHODS OF CONSUMER CONFIDENCE INDEX MEASUREMENT

The conventional means of determining consumer confidence are different from one country to the other. Also, the organizations that provide this information are different, being either a not-for-profit think tank or government departments and sometimes profitable private businesses. In Canada and the United States of America, telephone interviews and postal mail correspondences are used to conduct survey among a random selection of households. This could range from 2,000 to 5,000 correspondents (The Conference Board of Canada, 2014; Igboayaka, 2015). The Conference Board which publishes the CCI for the United States posits five questions on the CCS questionnaire sent to corresponding households (McKinsey 2012). They are:

- 1) What are the current business conditions?
- 2) What are the business conditions for the next six months?
- 3) What are the current employment conditions?
- 4) What are the employment conditions for the next six months?
- 5) What is the total family income for the next six months?

Survey participants are asked to answer each question as "positive," "negative" or "neutral?" The survey carried out bi-monthly starting in 1967 but changed to monthly tracking in 1977 (The Conference Board, 2020; Akhilesh Ganti, 2020). The publication of the CCI started in 1946. It was based on the University of Michigan's Survey of Consumers.

Since then, different aspects of its versions have evolved. For example, today’s version includes derivatives like the frequency of its measurement, its sample size and method of data collection, including the dependent variables and independent variables, the rules governing whether respondents from a previous data collection exercise could participate again, and what the maximum percentage of such respondents would be are added (Igboayaka, 2015). It has also been suggested that the CCI was established to produce an economic indicator that would “lead” imminent economic activity and thus be used to forecast consumer expenditure (Roberts & Simon, 2001; Igboayaka, 2015).

When it comes to how the CCI is determined, three of the oldest forms of measurement constitute the foundation upon which today’s concepts are built (Igboayaka, 2015). They are ABC News/Money, the Conference Board (of the USA) and the University of Michigan consumer confidence surveys. Pertinent information about their methods is shown in Table 1. All methods shown in Table 1 are based on the assumption that *confidence/sentiment* can be used to predict household spending, especially for durable products which are ‘discretionary’ in nature, given the ease with which their purchase could be postponed. As Mueller (1963, pp. 899) stated, “*confidence variables were close predictors of durable and non-durable household expenditures.*”

Table 1: Methodology of Consumer Confidence Index Surveys

	ABC News/Money	Conference Board	University of Michigan
Methods	Telephone	Mail	Conference
Sampling	RDD with random selection in household	Selection from a non random panel	RDD with random selection in household
Weighting	For probability of selection and to Census (region, age, race, sex and education)	Not disclosed	For probability of selection and to Census (age and income)
Sample Size	About 1,000 (250 per week x4 weeks)	About 2,500 for end-of month release; 3,500 for later revision	250-300 for mid-month release; 500 for end-of month revision
Field period	Wed-Sun each week; Results based on a four-week rolling average	Sent first of the month; Accepts returns through end of month	Around first of the month through a few days before the release
Release	Weekly, Tuesday evening	Prelim. Figures, last Tuesday of month; final figures with next month’s release	Preliminary figures at mid-month; final figures at end of the month
History	Started in December 1985	Started bimonthly in 1967; went to monthly in 1977	Started annually in 1946; quarterly in 1952 and monthly in 1978

Note: Differences between the three oldest forms of consumer confidence index measurement (Source: Reserve Bank of Australia 1 (Roberts & Simon, 2001))

Research has extended into components of the different methodological strategies used by these three consumer confidence indices. Most among these components are the sample size, the weightings used, the sampling and the method of data acquisition, such as telephone or mail. All of these components continue to drive the evolution of the consumer confidence index which originates from consumer confidence surveys. The majority of previous research focused on the economic indicators of consumer confidence, and refers to evaluating how the different components used to determine it are used, ignored, assumed to be significant. In particular, the choice of variables and their assignment as being dependent or independent for the application of statistical procedures like regression is considered. This continuous variation leads to a continuous revision of the consumer confidence index even after it has been published. Generating an index for consumer confidence is based on a method of data collection that has not changed much since 1946. Instead, it is the methods of processing the data, how the sample of the population that participates is rotated, the rules governing the participation of previous participants in new surveys and the data models used to develop an interpretation from the data that has gained widespread attention in the research community. Moreover, consumer confidence surveys are regularly conducted in many countries nowadays. For completeness, a review summary is given in Table 2, where consumer confidence indices from around the world are explained in terms of their methodology and approaches.

As shown in Table 2, consumer confidence indices are computed around the world and they can holistically be understood based on Organization Responsible, History, Methodology, Sample Size, Timing and Data Release, Operations and Questions, and Computation of Indices. Following organizations are responsible for publishing consumer confidence index in various countries; State Bank of Pakistan (in collaboration with IBA Karachi) in

Pakistan, Conference Board in Canada, Conference Board and university of Michigan in USA, GFK and MORI in United Kingdom, Commission in the Journal European Economy in France, and MasterCard worldwide index in Nigeria. Telephonic or mail survey is conducted for data collection from maximum 2000 respondents. Conference Board and university of Michigan carry survey first day of each month. The conference board releases its result at the end of the month whereas Michigan publishes its initial result on the second Friday of the month which is around 50 – 60% of the total number of responses collected. The final results are published at the end of the month (Curtin 2003). GFK conducts survey and published once every month. Commission in the Journal European Economy carry telephonic survey every first three weeks of the month except in August. State Bank of Pakistan conducts survey first week of every odd number month and results are published bimonthly frequency (State Bank of Pakistan, 2021).

MasterCard index in Nigeria is based on five questions related to the Economy as a whole, rate of employment in the country, the stock market, regular income of the consumer and quality of life experience or expected by the consumer. French consumer confidence index is based on five questions related to past and outlook position of personal finance, living standard and purchases. GFK survey in the United Kingdom is based on five specific questions related to last and next twelve months of their financial, economic and purchase position. The indices produced by both Michigan and consumer board based on questions on the current condition of the economy and questions about future expectations. Conference board survey is based on four specific questions related six months ago and from now of their financial, employment and buying conditions. State bank of Pakistan survey is based on survey six questions related to last and next six months of their financial, economic, employment and purchase positions (State Bank of Pakistan, 2021).

The consumer confidence survey measures do contain information about the current as well as future path of consumer spending but fall short of capturing real consumer purchasing behavior. The consumer confidence surveys might capture effects that will not appear in the data. There is one fundamental limitation of survey-based methods. It has to do with the non-participation of respondents and this is deemed as a higher risk due to inaccuracy compared to the risk of unrepresentative information that could result from the use of a small sample size (Roberts & Simon, 2001). This is the case of the Conference Board of Canada. On its website it is stated that the size of its sampled respondents is 2500 – 5000, but only about 1,500 respondents replied to the survey (The Conference Board, 2011; Bloomberg News, April 2003). Similarly, State Bank of Pakistan collect data via telephonic interview from 1600 respondents which is very small compared to the population of Pakistan. Additionally, the survey questions are speculative and published quarterly.

The limitations of cross-sectional nature of consumer confidence survey continues to spur research efforts to clarify and develop evidence on associations that form the core of the economic indicator of consumer confidence. Yet in another regard, the index of consumer confidence is subjected to rigorous analysis, especially when associations with economic activity are being discussed. One aspect of such dynamic has to do with the emergence of new sources of data that could be used to measure consumer confidence. However, this study conjecture that using consumer sentiment data on social media might show greater marginal significance for consumer confidence, because the social media can offer a huge volume of data on consumer confidence. Moreover, the analysis of such social media data can be conducted at a more rapid time and can also refine the accuracy of the CCI using data from far larger populations.

VII. SOCIAL MEDIA NETWORKS AS A SOURCE OF DATA FOR CONSUMER CONFIDENCE MEASUREMENT

Social Media is composed of online technology tools which enable communication between different people from around the world via the Internet. According to Margaret (2011), “*Social media is the collective of online communications channels dedicated to community-based input, interaction, content-sharing and collaboration*”. Social media is large and well known for its social network and content sharing abilities - videos and pictures (Igboayaka, 2015). It can also be seen as a setting for virtual discourse where people create content, share it, bookmark it and network at an extraordinary degree and speed. It is also used as a real-time snapshot of updates of interest, location, instant memories and so on. The use of social media networks has brought about one of the biggest cultural shifts since the industrial revolution. In 2020, over 3.6 billion people were using social media worldwide, a number projected to increase to almost 4.41 billion in 2025 (Statista, 2021b). It is engendering a whole new wave of human interaction, different from the structured methods that existed before it. The increased use of mobile Internet technology has also contributed to the growth in adoption of social media networks illustrated in Figure 1. Some examples of popular social media networks as seen in the figure include Facebook, Youtube, Whatsapp, Facebook Messenger, Instagram, Wechat, Tiktok, and Twitter.

Table 2: Summary of Consumer Confidence Indices from Around the World

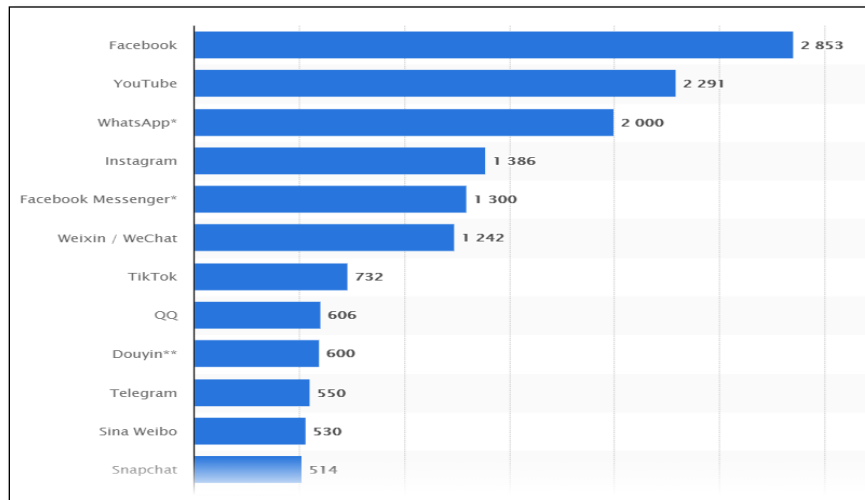
	Pakistan	Canada	United States of America	United Kingdom	France	Nigeria
Organization Responsible	State Bank of Pakistan (SBP) and IBA Karachi	Conference Board	Conference Board and university of Michigan	GFK and MORI.	Commission in the Journal European Economy	MasterCard worldwide index.
History	SBP started computing Consumer Confidence Index in January 2012 and having a bi-monthly frequency	The conference board started its operation in 1980 and has been independent research organization since inception. (Conference board Canada 2014).	The Conference board started in 1967 and Michigan’s Consumer index started in 1946.	GFK started across Europe as early as 1970 but officially started in UK in the year 1995.		
Method	Ttelephony survey of households based a stratified random sampling technique	Telephone	The conference board uses mail to acquire its own consumer confidence data while the Michigan uses Telephone as its method of acquiring the data.		Telephone.	Gathering information based on usage of master cards by the consumer.
Sample Size	Approximately 1600 households are used for the telephonic survey	Approximately 2,000 households are used for the survey.			The survey is carried out on approximately 2000 households.	
Timing and Data Release	The survey is being conducted since 2012 via telephone of households that are selected randomly across the country in the first week of every odd number month and results are published bimonthly frequency	In 1980 the consumer confidence data was producing once in four months, in 2002 the pattern was changed and the data was then produced monthly.	Both organizations start carrying out their survey first day of each month. The conference board releases its result at the end of the month whereas Michigan publishes its initial result on the second Friday of the month which is around 50 – 60% of the total number of responses collected. The final results are published at	The survey is carried out and published once every month. GFK focuses its questions on household finances and the general economic outlook.	This telephone numbers are selected randomly and dialled every first three weeks of the month except in August. July and September are used to calculate August.	This survey is carried out twice every year in the various geographical region in the country. (Vanguardngr 2011).

			the end of the month (Curtin 2003).			
Operations and Questions	<p>The consumer confidence survey is based on following six questions:</p> <ul style="list-style-type: none"> - How do you assess present financial position of your family compared to the last six months? - How do you assess present general economic condition of the country compared to the last six months? - In your opinion, compared to the last 6 months, how do you see the current time for buying durable goods such as furniture, refrigerator, television etc.? - How do you expect your financial position to change over the next six months from now? - How do you expect general economic conditions in the country to develop over the next six months from now? - What do you think unemployment over the next six months from now? 	<p>The consumer confidence survey is based on following four specific questions:</p> <ul style="list-style-type: none"> · Considering everything, would you say that your family is better or worse off financially than 6 months ago? · Again, consider everything, do you think that your family will be better off, the same or worse off financially 6 months from now? · How do you feel the job situation and overall employment will be in this community 6 months from now? · Do you think that right now is a good or bad time for the average person to make major outlay for items such as a home, car or other major item? 	<ul style="list-style-type: none"> · The indices produced by both Michigan and consumer board based on questions on the current condition of the economy and questions about future expectations. <p>For the measurement of the current condition the consumer board questions are only directed towards the view of economic conditions and not the personal experience of the consumer whereas Michigan questions are directed towards family finances and the peoples spending in general.</p> <ul style="list-style-type: none"> · In order to effectively measure future expectations both the conference board and Michigan assign two question directed specifically towards the economic condition and one question directed towards the consumer's personal finances. <p>One of the major differences is the structure of the question, Consumer board tends to ask</p>	<p>GFK survey in the United Kingdom is based on five specific questions. They include:</p> <ul style="list-style-type: none"> · How has the financial situation of your household changed over the last twelve months? · How do you expect the financial position of your household to change over the next twelve months? · How do you think the general economic situation in this country has changed over the last twelve months? · How do you expect the general economic situation in this country to develop over the next twelve months? · In view of the general economic situation, do you think now is the right time for people to make major purchases such as furniture or electrical goods? 	<p>The questions asked are based on five indicators, they include:</p> <ul style="list-style-type: none"> · Personal financial position –Past change · Personal financial position- outlook · Likelihood of major purchases · Living standards in France- past change · Living standards in France- outlook 	<p>MasterCard index worldwide survey questions in Nigeria are based on the following sectors</p> <ul style="list-style-type: none"> · The Economy as a whole · The rate of employment in the country · The stock market · Regular income of the consumer · Quality of life experience or expected by the consumer

			questions about the consumers view of the condition in their local area but Michigan aims at gathering data based on the customer views about the condition of the nation as a whole rather than their local environment. The approach of questions used by these bodies was widely criticized by Dominitz and Manski (2004).			
--	--	--	---	--	--	--

<p>Computation of Indices (Positive & Negative)</p>	<p>Result of consumer confidence surveys is reported in the form of Diffusion Index (DI), indicating the general tendency of respondents about a certain aspect of a particular survey. It is computed in the following two steps: Step 1: Net Response (NR) = (1.00*PP) + (0.50 × P) + (-0.50 × N) + (-1.00*NN). Step 2: Diffusion Index $DI = (100 + NR) / 2$ Whereas, PP = Increase/improve significantly; P = Increase/improve; E = Unchanged/neutral; N = decline/deteriorate; NN = decline/deteriorate significantly.</p>	<p>After collection of consumer confidence data, the positive and negatives are used to compute and analyse the data in order to provide a result to the general public on the consumer confidence of the country's economy. When the positive and negative has been sorted out the index is then computed for each question using the formula: <i>Percentage of positive responses/ (percentage of positive responses + percentage of negative response)</i></p>	<p>According to Daniel M., Gary E., Dalia S., <i>"The Conference Board index is computed by taking the positive percentage for each question divided by the sum of the positive and negative percentages. This number is then divided by the base year value from 1985 and multiplied by 100. The resulting values from each question are averaged to form the overall index"</i> <i>"Michigan computes its index by taking the difference between the positive and negative percentages for each question and then adding 100 to each. These are summed and then divided by a factor representing the base year, 1966. Finally, a small correction is made to account for a design change in the 1950s"</i>.</p>	<p>The indices are computed using the positive and negative values. According to Stuart B., Melissa D; <i>"The results are published as a net balance of positive less negative responses, with those who answered 'a lot' in either direction given twice the weight of those who answered 'a little'"</i>.</p>	<p>The consumer confidence is the calculated using positive and negative responses with formula; <i>% of positive response - % of negative responses.</i></p>	<p>The index is computed using positive, negative and neutral value. 0 values is assigned as the negative value which is known as the most pessimistic, 100 is seen as the most positive value which as represented as the most pessimistic and 50 is seen as the neutral value.</p>
--	---	--	--	--	--	--

Figure 1. World's most used social media platforms (Statista, 2021b)



Social media networks are a versatile source of information which could be harnessed for assessing the human perceptions that reflect their confidence in the economy. For example, Facebook and Twitter are popular networks used to express opinion, a state of mind or a perspective which could be mined for the sentiment of users relative to their economic situation. Youtube is not only providing video contents but also allow users to comment on the video, generating huge sentimental data about the video. Also, blogging sites including News sites, they generate a rich source of opinion towards a subject of discussion; e.g., consumer price, personal finance, unemployment, purchasing behavior, a new government policy, stock market fluctuations, a commercial investment deal and much more. Aggregation of such information can be used to 'know' the economic state of consumers. Furthermore, 60% of the world's population has access to the Internet and now interacts using social media networks (Statista, 2021b). Since its inception in 1996, social media has managed to infiltrate half of the 7.7 billion people in the world (Brian Dean, 2021). Social media platforms almost tripled their total user base in the last decade, from 970 million in 2010 to the number passing 4.48 billion users in July 2021 (Brian Dean, 2021). Therefore, tapping into this 'new' data source of social media could give a supplementary boost to the conventional methods of data gathered to measure the level of consumer confidence in the economy.

For use in the economic and business sector, social media involves the gathering of comments made in online forums, created with the sole purpose of educating people about products, brands, services and general issues in the economy (Blackshaw & Nazzaro, 2004). Mangold and Faulds (2009) defined social media as a wide range of word-of-mouth forums which allows users to be able to express their feeling about a topic or even create a topic for discussion. The communicating parties are able to connect and hold conversations using formats such as; text, audio, video, images and other multimedia. Social media has become one of the strongest and most powerful tools used online since it allows thousands of users to be reached within the shortest minimum time possible. According to statistics provided by socialnomics.net, on the speed of development of the social media, it took the radio 38 years to reach 50 million users, TV 13 years to reach 50 million users, but it took Facebook just 9 months to reach 100 million users (Igboayaka, 2015). The tremendous difference in outreach suggests that social media is an extremely fast way to get to customers or users about a particular product, service or issue. Social media has not only been used for personal use but it has been used for business as well. Using social media has become one of the new methods used by The Economist to gather data on the level of consumer confidence in the economy, the methods including: counting the number of likes and dislikes of a topic, how many comments have been given by consumers on relating topics and gathering how many re-tweets were done by consumers (Igboayaka, 2015). This method of gathering data shows the real consumer sentiments towards a brand or the economy based on the comment provided by the consumer. A study on social media effectiveness showed that 53% of the authenticated Twitter users recommended a company and its products by tweeting about it and 48% of the people actually delivered on their intentions to buy that particular product (Performics, 2010).

Social media networks are rich with expressions of sentiment which is important to the measurement of consumer confidence index. *Sentiment* is known as the feeling or reason for an expression made behind a comment and reference about a particular topic, news or product (Igboayaka, 2015). Various tools have been made which enable measuring

the consumers' sentiment. From a business perspective, the ability of a business owner or a brand to be able to measure or understand the consumer's sentiment behind a comment is very beneficial to the growth of the organization. This helps to determine the state of mind a consumer was in as at the time he/she made the comment. Asur and Huberman (2010) conducted a study on how social media can help predict the future outcome of a new movie before it is released. In this research, Twitter was used to forecast box-office revenues for movies. The result from the research was calculated by analyzing the number of tweets that are currently on the network based on a particular movie. These tweets were divided into positive and negative tweets, Positive tweets are interpreted as representative of customers looking forward to watching the movie based on the preview, while the negative tweets mean that the customers are not interested in watching the movie. A linear regression model for predicting box-office revenues of movies in advance of their release was constructed. The results showed that there is a strong correlation between the amount of attention or discussions consumers have about a movie about to be released and its subsequent ranking when the movie is eventually released (Asur & Huberman, 2010). Similarly, research conducted by Chris Barry, Rob Markey, Eric Almquist and Chris Brahm (2011) to examine how the social media has affected the consumers' confidence on a product or the organization as a whole. The questions asked on the social media about the product included "How likely would you recommend [this company or product] to a friend or colleague in social media?" After the analysis, they stated that: "Customers who engage with companies over social media are more loyal and they spend up to 40 percent more with those companies than other customers".

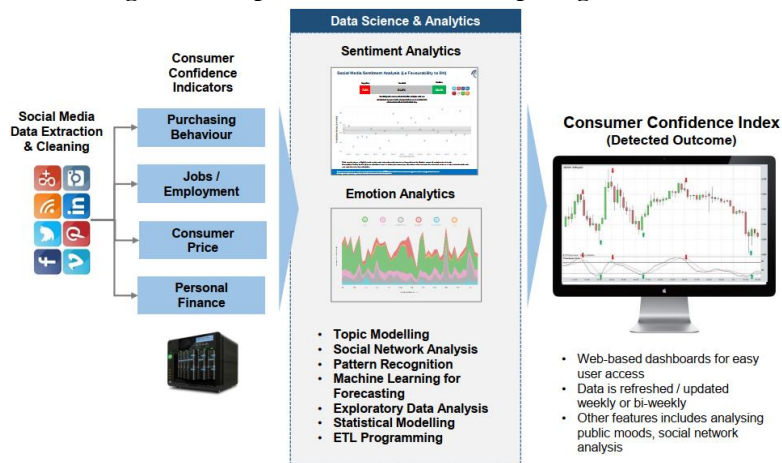
In 2012, Schweidel, Moe, and Boudreaux carried out an analysis of the potential to get brand sentiment from social media conversations. They used data which was collected from different social media domains such as Facebook and Twitter. They proposed the use of a hierarchical Bayesian regression model which was used to measure these sentiments effectively. Based on the research carried out, it has been discovered that social media is one of the fastest and easiest mediums that can be used to get large amounts of consumers reviews which can be used eventually to calculate the overall consumer confidence on a brand or the economy as a whole. Different approaches on what actually affects consumer confidence were also researched. Among them are unemployment, rising inflation and stock price, but the greatest question still remains "What actually causes a rise and fall in consumer confidence?" In conclusion, the social media itself can actually cause a rise and fall in consumer confidence in the world today as information about the economy can easily be known by consumers as quickly as it happens (using blogs, Facebook, Twitter). This affects how consumers feel about the economy unlike the days when such platform was not easily accessible. Using social media networks as a platform for social intelligence gathering is catching on as a data source in the market intelligence research industry. Companies such as Nielsen, Netbase, Frost & Sullivan, and Mintel enlist their use of content from social media networks, for the purpose of intelligence gathering and marketing (Netbase, 2010; Nielsen, 2013). It is no surprise that a lot of research continues to concentrate on the improvement of analytical methods and frameworks to translate such content into useful information. It is this translation that is pivotal to the usefulness and correctness of whatever information is synthesized from social media networks. For example, an area of study like sentiment analysis frameworks have originated concepts such as Natural Language Processing (NLP) that can be used as a tool to translate data from social media networks into meaningful insight. However, consumer confidence index and how it is measured have expanded their scope beyond the use of questionnaires alone, and data sources present opportunities to gain another perspective on consumer behavior in terms of economic habits (past), situation (present) or intentions (future). At this time, social media networks are a relevant sources and more effective compare to cross-sectional nature of survey questionnaires.

VI.II. PROPOSED METHOD FOR COMPUTING NACOP

An overview of the proposed method to build a National Consumer Confidence Predictor (NACOP) is presented in Figure 2. The NACOP function in four steps. First, NACOP extracts millions of social data from conversations of Pakistani users related to consumer purchases (or purchasing intent), jobs and employments, state of personal finances and other consumer indicators that are relevant to predict the state of the economy. Second, data will be cleaned and subjected to multiple data science approaches to ensure relevancy of the outcome in predicting the consumer sentiment and the public mood. Third, data will be stored online and made available to end users via a web-based portal or mobile application. Public and paid-subscribers will be informed via seminar, workshop, or other marketing ways. Fourth, data will be updated every week to provide almost-real time insights to reaction from the public related to policy changes or national events.

Initially, four data types consisting of Purchasing Behavior, Job/Employment, Consumer Price and Personal Finance are extracted from social media for the purpose of predicting the consumer confidence. A brief description of the data types is presented as follows:

Figure 2: Proposed Method for Computing NACOP



VI.III. PURCHASING BEHAVIOUR

Purchasing data will consist of following three dimensions:

- **What people will buy** - any posts related to what people plan to purchase (e.g., going for movies/dinner/entertainment, buying groceries, planning a holiday trip, upgrading vehicles, new house application) and any form of expenses.
- **What people have bought** – any posts related to the past; i.e., items which they have bought recently;
- **What people will not buy** – any posts that exhibit any form of comments related to expensive items, pricing, (which includes complaints)

VI.IV. JOBS/EMPLOYMENT

Jobs/employment data will contain following two dimensions:

- **Employment** – Posts related to new jobs offers, job offers received, salary revision, bonus, increment, new positions, career advancement
- **Unemployment** – Posts related to unemployment, difficulty in finding jobs, fresh graduates complaints on getting their first job, mutual separation and retrenchment, complaints on low wages, foreign labour, expatriates, etc.

VI.V. CONSUMER PRICE

Consumer price data will contain following dimension:

- **Consumer prices** – Posts related to consumer prices based on Consumer Price Index (CPI) products, price complaints, price increase or decrease.

VI.VI. PERSONAL FINANCE

Personal finance data will contain following two dimensions:

- **Savings Capacity** – posts on personal current financial situation, current debt position, current savings, dividends or returns from financial institutions or investments.
- **Savings Intention** – posts related to saving intentions in any form of investments; i.e., current or saving accounts, stocks, bonds, public mutual, etc.

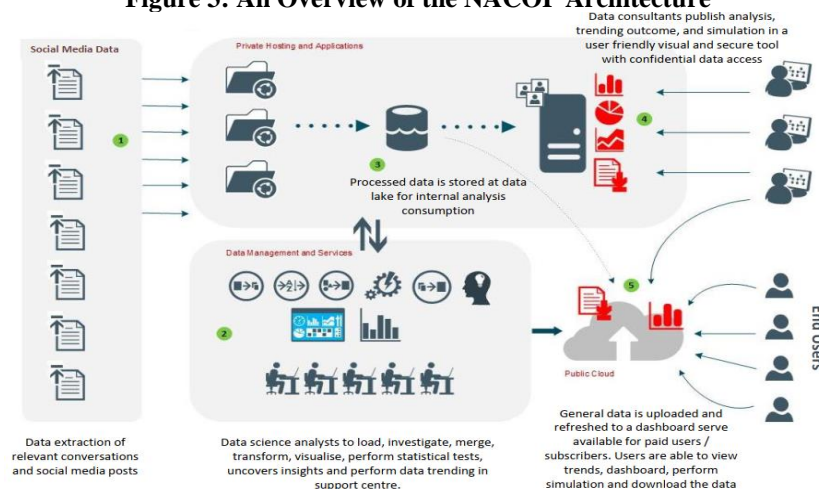
VI.VII. ARCHITECTURE OF THE PROPOSED NACOP

An overview of the proposed NACOP architecture is presented in Figure 3. After data extraction of relevant conversations and social media posts, it is stored at Data Lake for internal analysis consumption. Then, data science analysts to load, investigate, merge, transform, visualize, perform statistical tests, uncovers insights and perform data trending in support centre. Subsequently, data consultants publish analysis, trending outcome, and simulation in user friendly visual and secure tool with confidential data access. General data is uploaded and refreshed to a dashboard serve available for users. They will be able to view trends, dashboard, perform simulation and download the data.

VI.VIII. RESEARCH IMPLICATIONS AND FUTURE STUDIES

The study has both theoretical and practical implications. In spite of the limitations, the official CCI is the only accepted indicator in Pakistan that the industry and key players can rely on. We proposed an alternative method of social media analytics for predicting consumer confidence index. It can benefit academicians, retailers, financial institutions and Banks, industry players, and government. Future studies can empirically test the proposed NACOP technique with the big data extracted using social media data extractions. Moreover, the practitioners can also implement the proposed NACOP technique in web portal or mobile app that will provide various features regarding the consumer confidence index retrieved and analyzed via social media posts. For example, summary of insights and key events which took place on selected peaks and access to historical data to view macro indicators/high level figures by date. Additionally, it can show the public/consumer mood based on all key economic consumer indicators (e.g., purchasing intent, personal finance, employment & consumer price) and NACOP results can be represented in different colors in the bar chart. The practitioners can further provide the qualitative insights by describing the key events to benefit end users and decision makers. Moreover, the NACOP users can compare two or more consumer confidence indicators (e.g., Purchasing vs. Jobs/Employment) to observe trends and patterns triggered by potential events or changes in policies.

Figure 3: An Overview of the NACOP Architecture



VII. CONCLUSION

The industry challenge is to rely on the official CCI published by State Bank of Pakistan which is the only available indicator of consumer confidence in Pakistan. This study proposed a novel approach of social media analytics for predicting national consumer confidence index which is named as “NACOP” with the aim of rolling-out a viable commercial product/solution in the form of NACOP web-based portal. The proposed NACOP utilizes big data and data science to predict the national consumer confidence using large data sets of purchasing behaviour, jobs/employment, consumer price and personal finance from social media platforms. The study also explains the architecture of proposed NACOP technique for computing national consumer confidence index and provides significant implications to academicians and practitioners.

REFERENCES

- Akhilesh, G. (2020). *What Is the Consumer Confidence Index (CCI)?*
- Ali, A. (2015). *The impact of macroeconomic instability on social progress: an empirical analysis of Pakistan*. (Doctoral dissertation, National College of Business Administration & Economics Lahore).
- Ali, A. (2018). Issue of Income Inequality Under the Perceptive of Macroeconomic Instability: An Empirical Analysis of Pakistan. *Pakistan Economic and Social Review*, 56(1), 121-155.
- Ali, A. and Bibi, C. (2017). Determinants of Social Progress and its Scenarios under the role of Macroeconomic Instability: Empirics from Pakistan. *Pakistan Economic and Social Review* 55 (2), 505-540.
- Ali, A., & Ahmad, K. (2014). The Impact of Socio-Economic Factors on Life Expectancy in Sultanate of Oman: An Empirical Analysis. *Middle-East Journal of Scientific Research*, 22(2), 218-224.
- Ali, A., & Audi, M. (2016). The Impact of Income Inequality, Environmental Degradation and Globalization on Life Expectancy in Pakistan: An Empirical Analysis. *International Journal of Economics and Empirical Research*, 4 (4), 182-193.

- Ali, A., & Audi, M. (2018). Macroeconomic Environment and Taxes Revenues in Pakistan: An Application of ARDL Approach. *Bulletin of Business and Economics (BBE)*, 7(1), 30-39.
- Ali, A., & Rehman, H. U. (2015). Macroeconomic instability and its impact on the gross domestic product: an empirical analysis of Pakistan. *Pakistan Economic and Social Review*, 285-316.
- Ali, A., & Şenturk, I. (2019). Justifying the Impact of Economic Deprivation, Maternal Status and Health infrastructure on Under-Five Child Mortality in Pakistan: An Empirical Analysis. *Bulletin of Business and Economics*, 8(3), 140-154.
- Ali, A., & Zulfıqar, K. (2018). An Assessment of Association between Natural Resources Agglomeration and Unemployment in Pakistan. *Pakistan Vision*, 19(1), 110-126.
- Ali, A., Ahmed, F., & Rahman, F. U. (2016). Impact of Government Borrowing on Financial Development (A case study of Pakistan). *Bulletin of Business and Economics (BBE)*, 5(3), 135-143.
- Ali, A., Audi, M., & Roussel, Y. (2021). Natural Resources Depletion, Renewable Energy Consumption and Environmental Degradation: A Comparative Analysis of Developed and Developing World. *International Journal of Energy Economics and Policy*, 11(3), 251-260.
- Ali, A., Audi, M., Bibi, C., & Roussel, Y. (2021). The Impact of Gender Inequality and Environmental Degradation on Human Well-being in the Case of Pakistan: A Time Series Analysis. *International Journal of Economics and Financial Issues*, 11(2), 92-99.
- Ali, A., Mujahid, N., Rashid, Y., & Shahbaz, M. (2015). Human capital outflow and economic misery: Fresh evidence for Pakistan. *Social Indicators Research*, 124(3), 747-764.
- Arshad, S., & Ali, A. (2016). Trade-off between Inflation, Interest and Unemployment Rate of Pakistan: Revisited. *Bulletin of Business and Economics (BBE)*, 5(4), 193-209.
- Ashraf, I., & Ali, A. (2018). Socio-Economic Well-Being and Women Status in Pakistan: An Empirical Analysis. *Bulletin of Business and Economics (BBE)*, 7(2), 46-58.
- Asur, S. & Huberman, B. A., (2010). *Predicting the Future with Social Media*. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference. IEEE/WIC/ACM, 492 – 499.
- Audi, M & Ali, A. (2017). Socio-Economic Status and Life Expectancy in Lebanon: An Empirical Analysis. *Archives of Business Research*, 5(11), 159-170
- Audi, M. & Ali, A. (2017). Environmental Degradation, Energy consumption, Population Density and Economic Development in Lebanon: A time series Analysis (1971-2014). *Journal of International Finance and Economics*, 17(1), 7-20.
- Audi, M. Sadiq, A. Ali, A. and Roussel, Y. (2021). Performance Evaluation of Islamic and Non-Islamic Equity and Bonds Indices: Evidence from Selected Emerging and Developed Countries. *Journal of Applied Economic Sciences*, 16(73), 251– 269.
- Audi, M., & Ali, A. (2016). A Causality and Co-integration Analysis of Some Selected Socio-Economic Determinants of Fertility: Empirics from Tunisia. *Bulletin of Business and Economics (BBE)*, 5(1), 20-36.
- Audi, M., Ali, A., & Al-Masri, R. (2022). Determinants of Advancement in Information Communication Technologies and its Prospect under the role of Aggregate and Disaggregate Globalization. *Scientific Annals of Economics and Business*. <https://doi.org/10.47743/saeb-2022-0009>.
- Audi, M., Ali, A., & Roussel, Y. (2021). Aggregate and Disaggregate Natural Resources Agglomeration and Foreign Direct Investment in France. *International Journal of Economics and Financial Issues*, 11(1), 147-156.
- Audi, M., Ali, A., & Roussel, Y. (2021). Measuring the Tax Buoyancy: Empirics from South Asian Association for Regional Cooperation (SAARC). *Empirical Economics Letters*, 20(12).
- Audi, M., Ali, A., & Roussel, Y. (2021). The Advancement in Information and Communication Technologies (ICT) and Economic Development: A Panel Analysis. *International Journal of Innovation, Creativity and Change*, 15(4), 1013-1039.
- Barry, C. Markey, R. Almquist, E. B. & Chris, (2011). *Putting Social Media to Work*.
- BBC, (2014). *Boom and Bust*.
- Blackshaw, P. & Nazzaro, M., (2004). Consumer-Generated Media (CGM) 101: Word-of-Mouth in the Age of the Web-Fortified Consumer. *A Nielsen Buzz Metrics White Paper*, Spring.
- Bloomberg News, (April 2003). Consumer Confidence Shows a Substantial Gain. *The New York Times*, 8.
- Brian, D. (2021). *Social Network Usage & Growth Statistics: How Many People Use Social Media in 2021?*
- Cheng, Y. (2020). The Determinants, Implications and Interaction of Consumer Sentiment, Durham theses, Durham University. Available at Durham E-Theses Online:
- Curtin, R. T., (2002). *Surveys of Consumers: Theory, Methods, and Interpretation*. Washington DC, S.N.
- Haider, A., & Ali, A. (2015). Socio-economic determinants of crimes: a cross-sectional study of Punjab districts. *International Journal of Economics and Empirical Research*, 3(11), 550-560.

- Hirt, M., & Willmott, P. (2014). Strategic principles for competing in the digital age. *McKinsey Quarterly*, 5(1), 1-13.
- Igboayaka, J. V. C. E. (2015). *Using social media networks for measuring consumer confidence: Problems, issues and prospects* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Investopedia, 2014. *Capital Investment*.
- Kassem, M. Ali, A. & Audi, M. (2019). Unemployment Rate, Population Density and Crime Rate in Punjab (Pakistan): An Empirical Analysis. *Bulletin of Business and Economics (BBE)*, 8(2), 92-104.
- Mangold, G. W. & Faulds, D. J., (2009). Social media: The new hybrid element of the promotion mix. *Elsevier*, 52(4), 357–365.
- Margaret, R., (2011). *Capex (capital expenditure)*.
- McKinsey & Company, (2012). *The social economy: Unlocking value and productivity through social technologies*.
- Mehmood, A. Siddique, H. M. A., and Ali, A. (2022). Impact of Health on Worker Productivity: Evidence from South Asia. *Bulletin of Business and Economics*, 11(2), 1-8.
- Mueller, E., (1963). Ten Years of Consumer Attitude Surveys: Their Forecasting Record. *Journal of the American Statistical Association*, 58(304), 899-917.
- Netbase, (2010). *How Does Netbase Achieve the Best Accuracy for Understanding Consumers Online?*
- Nielsen, (2013). *Consumer Confidence Concerns and Spending Intentions around the World*.
- Pendery, D., (2009). Three top economists agree 2009 worst financial crisis since great depression; risks increase if right steps are not taken. *Business Wire News.*)
- Performics, (2010). *Social Networking Study: Facebook Use Continues to Rise; Brand Participation and Engagement Heavily Welcomed by Social Networkers*.
- Petev, I. D., Pistaferri, L., & Saporta-Eksten, I. (2011). Consumption decisions are crucial determinants of business cycles and growth. Personal consumer. *The Great Recession*, 161.
- Roberts, I. & Simon, J., (2001). *What Do Sentiment Surveys Measure?*
- Roussel, Y., Ali, A., & Audi, M. (2021). Measuring the Money Demand in Pakistan: A Time Series Analysis. *Bulletin of Business and Economics (BBE)*, 10(1), 27-41.
- Sajid, A. & Ali, A. (2018). Inclusive Growth and Macroeconomic Situations in South Asia: An Empirical Analysis. *Bulletin of Business and Economics (BBE)*, 7(3), 97-109.
- Schweidel, D. A., Moe, W. W. & Boudreaux, C., 2012. Social Media Intelligence: Measuring Brand Sentiment from Online Conversations. *Marketing Science Institute*, 12-100.
- Şentürk, İ., & Ali, A. (2021). Socioeconomic Determinants of Gender Specific Life Expectancy in Turkey: A Time Series Analysis. *Sosyoekonomi*, 29(49), 85-111.
- Shayaa, S., Al-Garadi, M. A., Piprani, A. Z., Ashraf, M., & Sulaiman, A. (2017). Social media sentiment analysis of consumer purchasing behavior vs consumer confidence index. In *Proceedings of the International Conference on Big Data and Internet of Thing* (pp. 32-35).
- State Bank of Pakistan (2021). Consumer Confidence Survey.
- Statista (2021a). Most popular social networks worldwide as of July 2021, ranked by number of active users.
- Statista (2021b). Number of global social network users 2017-2025.
- The Conference Board of Canada, (2014). *Consumer Confidence*.
- The Conference Board, (2011). *Consumer Confidence Survey Technical Note - February 2011*.