ISSN 1800 – 4989 (Print)
ISSN 2719 – 2598 (Online) SLJM Vol 8 No 3: September-December 2022

Sri Lanka Journal of Marketing
University of Kelaniya - Sri Lanka

Application of Big Data Analytics in Telecommunication Sector Company in Sri Lanka: with Reference to Marketing Perspective

R.T. Senavirathne
seneviratneraveen@gmail.com

DOI: http://doi.org/10.4038/sljmuok.v8i3.114

ABSTRACT

The business world is getting competitive and complex day by day. The customer needs and expectations also changing rapidly. Therefore, making effective strategic marketing decisions in a short time is essential to hold a competitive advantage. Especially in the telecommunication sector, there are lots of data to be handled every day. By analyzing all those data and extracting the most critical insights and use those insights for strategic marketing decisions is essential to remain as the market leader. Today, Big Data Analytics is used to analyses massive databases and find patterns and extract valuable information and insights. Therefore, this research investigates on strategic marketing decisions, Big Data Analytics and how Big Data Analytics use in telecommunication sector to make strategic marketing decisions. Moreover, this research examines how Big Data Analytics can be applied to perform customer segmentation based on Psychographic segmentation by analyzing URLs using artificial neural network techniques. This study will provide suitable recommendations for the telecommunication companies to apply big data analytics to improve the revenue, and finally it points out the possible future study areas.

Keywords: *Big Data, Big Data Analytics, Customer Segmentation, Strategic Marketing Decisions, Telecommunication, URL Data*

1. BACKGROUND

The modern business world is heavily dependent on data-driven strategic decisions since every business and organisation collects massive amounts of data daily basis. Even though organisations used to analyse data with traditional analysis methods, due to the rapid increase of data types, amounts, and data collection speed, today, analysts tend to use modern techniques for data analysis. When data collection has primary characteristics of changing its varieties and increasing the volume of data, it commonly identified as 'Big Data'. Those methods used to analyse Big data, called the 'Big Data Analytics' (McAfee, Davenport & Patil, 2012).

'Big Data' can identify as a set of raw data and information characterised with high volume, variety and velocity where it required specific analytical methods and technologies to transform into value (Beyer, & Laney, 2012; Mauro, Greco & Grimaldi, 2016; Microsoft Research, 2013; Schroeck, Shockley, Smart, Romero-Morales & Tufano, 2012). 'Big data analytics' refers to a set of methods, techniques and technologies used to analysis 'Big data' in order to expose substantial hidden patterns and values (Agrawal, Pate, Patel, & Verma, 2016; Armour, Espinosa, Kaisler & Money, 2013; Chen & Zhang, 2014; Riahi & Riahi, 2018).

Linus Osuagwu (2016) defines 'strategic marketing decisions' as marketing decisions which require a commitment of substantial resources, which can be sensitive to the environment and oriented towards achieving objectives and long-term goals. Today, many large-scale companies, as well as small businesses, use Big data analytics in order to make strategic marketing decisions and get the benefit of hidden insights. Wal-Mart uses the product to redesign their arrangements and placement after understanding the buying patterns of their customers (Shaw, Subramaniam, Tan, & Welge, 2001).

According to a report in 2019 published by the Telecommunications Regulatory Commission (TRC) of Sri Lanka, 29 licenses have been granted for telecommunication purposes. Five licenses are granted for Data Communications

(facility-based and non-facility based). The number of Cellular Mobile Telephone Subscriptions are around 32,884,000, and also the Number of Fixed Access Telephone Subscriptions are around 2,299,700. Further, there are more than 1,654,170 Fixed Broadband Subscriptions and 11,754,160 Mobile Broadband Subscriptions in Sri Lanka where those numbers are increasing rapidly. The report of Economic and Social Statistics of Sri Lanka states that the gross national income (at the current market prices 2013) of the telecommunication service industry is over 75 billion Sri Lanka rupees by 2018. Further, it can see that it has a rapid increase from 41 billion in 2013 and according to the statistics, it will increase further (CBSL, 2019). Therefore, the telecommunication industry in Sri Lanka is highly essential for the citizens and to the economy of the country.

Today most of the telecommunication companies are involved in numbers of products other than voice. Data Communications, Fixed Access Telephone service, Cellular Mobile Telephones, Cable TV Satellite Broadcasting, and there are other many value-added products powered by the telecommunication sector companies (Annual Report Dialog (PLC), 2018; Annual Report Hutch (PVT), 2018; Annual Report SLT (PLC), 2018). The telecommunication sector companies provide numbers of products, as well as all the products and equipment, rely on digital platforms, the system generates massive amounts of various types of data. Due to the size of data and data collection velocity, use of the traditional analysing methods is limited. Therefore, the use of Big Data analytics is significant in the telecommunication industry (Sun, Wang, Wei & Zhan, 2017).

When it comes to cellular networks and BDA, there are researchers available such as optimising network performance by analysing time-series data and cell phone activity data in order to enhance mobile user experience (Sun and Zhu, 2020), and train a machine learning model to detect and predict traffic in cellular networks (Sultan and Zhang, 2018). However, many of these BDA researches are focused on optimising and enhancing operations of the telecommunications sector. Further, several journal articles examine applications of BDA on marketing strategies. However, those

researches were not focused on telecommunication industry (Medouri, Rahich, Saidali & Tabaa, 2019).

The telecommunication companies provide their services to large amounts of people with an enormous variety of expectations. However, most of the telco companies treat their customer base with cooperate and non-cooperate based by considering the usage. It has poor performance when it comes to marketing new products and services. Therefore, identifying each distinguishes customer segment is essential to provide quality service for each group of customers. Moreover, this identification process is highly involved with traditional data analysis methods. This research examines available literature on BDA, and how modern, sophisticated analytical methods of Big Data Analytics can apply to make effective strategic marketing decisions in the telecommunication sector in Sri Lanka.

2. LITERATURE REVIEW

In the 21st century, data and information have become the most valuable asset. Many companies and organisations collect data about their business and maintain databases. Today the internet users produce around 2.5 quintillion bytes in every single day, and it has forecast that by 2020 there will be approximately 40 zettabytes (40 trillion gigabytes) amount of data. However, processing that much data and extracting valuable information and insights is not that much more comfortable with traditional data analytic methods (Petrov, 2019). To address the problem of analysing a large number of data immediately, Big Data Analytics (BDA) comes to the play. For a better understanding of markets and business in an organisation, qualitative analysis and quantitative analysis on data is useful. It will help to business decisions according to the need of time (Chen & Zhang, 2015). Today, most of the companies and organisations use BDA to analyse the data they have and get proper information and insights about the market and the customers in order to make suitable strategic decisions and provide the products that the market request. Since many organisations

use BDA, the most critical fact is which organisation use the most innovative and advanced techniques suitably with BDA to their business (Gandomi & Haider, 2015).

2.1. Big Data Analytics

Big Data

There is empirical evidence that the definition of big data is changing with the advancing of the technology. A study in 2014 reflects the evolution of definitions of big data in the notation of V's; In 2011 Russom defined in terms of 3V's with Volume, Velocity and Variety; in 2012 Oracle and Forrester added 4th V Value and defined in terms of 4V's; finally, White added 5th dimension of Veracity (Akter, Chopin, Edwards, Fosso, & Gnanzou, 2015). According to White (2012), big data is defined in terms of 5Vs.

- Volume: Data can be found in ample storage or as in a large number of records. These data originate from various sources such as social media, sensors, digital transactions and many other resources (Russom, 2011).
- Variety: Data can be in multidimensional fields, different formats and generated by various sources. Big data can be structured (numeric data in traditional databases) or unstructured (text documents, audio, email, stock ticker data, video, and financial transactions) formats (Russom, 2011).
- Velocity: The frequency of data transfer or/and generation. Today the speed of data generating is increasing rapidly (Russom, 2011).
- Veracity: Need an analysis of big data to gain predictions in unpredictability (Beulke, 2011).
- Value: Which set of big data gives maximum benefit or /and insight on economically the best insights adding higher value (Demchenko, Grosso, De Laat, & Membrey, 2013; White, 2012).

Big Data Analytics

In generally, 'Big Data Analytics' (BDA) is defined as a set of techniques and technologies which used in order to investigate 'Big Data' where it from usual ones to more complicated, and of a sizeable massive scale and disclose large hidden values. It is a process of organising, collecting, as well as analysing extensive data sets with the use of different patterns and other beneficial information (Armour, Espinosa, Kaisler & Money, 2012; Begoli & Horey, 2012; Chen & Zhang, 2014). The BDA can directly apply in different application, including real-time analysis such as social networks, biomedicine, astronomy, navigation as well as business. When it comes to business context, 'Big Data Analytics' can be identified as a process of examining 'Big Data' and discovering unknown correlations, hidden patterns, customer preferences, market trends and various other information which relate with businesses (Bifet & Fan, 2014; Kumar, Wang & Yadav, 2013).

Primarily, BDA focuses on solving new problems/challenges or old problems/challenges in better and effective ways.

2.2. Machine Learning Algorithms

Machine learning algorithms can recognise as mathematical and logical programs that can self-adjust in order to perform better and expose to more data. After the live ware (human) specify the suitable machine learning algorithm while performing the learning cycles, the machine learning algorithm can adjust its parameters, with respective to feedback on its previous performance making predictions with a given dataset (Nicholson, 2019). Those algorithms can be simple to very complicated.

Artificial neural networks (ANN) are a higher demanding algorithm in use today. Rojas (1996) defines ANNs as distributed, adaptive, generally non-linear learning machines built from many different processing elements (PEs). The PEs sum all these contributions and produce an output that is a non-linear (static) function of the sum (Rojas, 1996). Deep learning also a type of neural network. Unlike ANNs, deep

learning has more hidden layers to obtain composing simple and non-linear modules which can transform the one hidden layer (starting with the raw input) into a representation at a higher, slightly more abstract level. Its algorithm has very complex functions (Bengio, Hinton & LeCun, 2015). Therefore, deep learning neural networks construct so many other algorithms such as Convolutional neural networks (CNN), Recurrent neural networks (RNN), and Natural Language Processing (NLP). Natural Language Processing is the analysis of linguistic data, such as data in the form of textual (publications, documents, and other text formats) using computational methods. The main goal of NLP is to generate a digital representation of the text in structure to the unstructured natural languages by gathering insights from linguistics (Cohen & Verspoor, 2013). So, NLP algorithms can be used on text type input and get a pattern or a classification as a result.

2.3. Business Intelligence (BI)

Kimble & Milolidakis (2015) state that business intelligence is the ability of a company to make meaningful insights by using available data. Further, they explain that Business intelligence contains a range of areas such as customer intelligence, competitor intelligence, product intelligence, market intelligence, technological intelligence, strategic intelligence and the business counterintelligence (Kimble & Milolidakis, 2015). In 2004, Thomson conducted a survey and suggested significant benefits of BI are improving business decision making (78%), more accurate reporting and generating faster (81%), improving customer service (56%) as well as increasing company revenue (49%) (Gong & Xia, 2014). BDA can go hand in hand with BI in various applications such as managing supply chain risk, providing real-time business insights in order to optimise pricing and help make essential decisions, generating competitive intelligence, as well as improving customer satisfaction (Davenport, 2014; Alexander & Wang, 2015).

In the business world, many tools are using for BI in various stages of management and those tools can be customised into different industries. At the operational level,

BI act through enterprise performance management (EPM) and generates preformatted reports; at the middle management level, BI creates 'On-line Analytical Processing' (OLAP) tools that provide a high-level aggregated view of data; for the top-management level, BI interact through the BI dashboards (Hurbean, 2006). Today in the business world, there are many modern, and sophisticated tools and software such as Microsoft Power BI, SAP Business Intelligence, MicroStrategy, SAS Business Intelligence, and Tableau (Haije, 2019) are using in order to make smooth and advance their business.

Big Data Analytics in the Telecommunication sector

The telecommunications sector is a highly expanding sector in the era. As mentioned earlier, the number of users, the complexity and the data exchanging are increasing day by day within the sector. In 2017, Wang, Wei, Zhan and Sun stated that the mainstream telecom operators in multiple countries already had applied the big data analytics for their telco businesses. Those applications cover the growth of both internal as well as the external applications. Therefore, when examines empirical evidence, it can find many applications of big data analytics which in use today. A major application of BDA in the telco is to handling network traffic. Since there is a massive number of requests are coming to the networks, today BDA is in use for network traffic handing. An extensive scale (commercial cellular network with 4.2 Tera bytes input volume every day) network traffic analysis and monitoring proposed by Liu, Ansari & Liu (2014). This system is based on a Hadoop platform, and it has proved that the system has the capability of processing big network-generated data and revealing the traffic also to monitor the users' behaviour. Another popular BDA application in telecommunication sector is to real-time identifying network problems. This system helps to identify network failures and alert them to resolve soon. It helps to improve the quality of the networks and ultimately, improvement of customer satisfaction (He, Qiu, Robert, Yao, Yin, Yu, Zhao & Nan, 2016). Moreover, after analysing the customer base and user behaviour, churn management or the churn prediction also done by the BDA using modern algorithms (Teradata, 2014). These

insights regarding the existing customer base are highly significant to the company in order to provide better customer service.

2.4. Role of Big Data Analytics in Decision marketing

Most of traditional decision support systems provide insights for the business by processing transaction systems such as ERPs. Therefore, these decisions are mostly bound to the internal business decisions process (Davenport & Dyché, 2013). Instead of analysing only the transaction data, if it considers structured and unstructured big data which is related to the business, it will provide more insights to make more creative decisions for the company. Therefore, most of the leading companies use BDA in their decision-making (Jeble, Kumari & Patil, 2018). BDA techniques use by Wal-Mart to an understanding of patterns and redesign their arrangements and placement of products together so they can improve sales (Shaw, Subramaniam, Tan, & Welge, 2001). So, it is evidence that BDA has the capacity to predict the demand for the products or products bought together.

Moreover, from a survey done by Jeble, Kumari and Patil (2018), provide numerous empirical evidences that BDA has a critical role in the decision-making process. Such as Predicting the spread of flu by geography by regions (Mayer-Schönberger & Cukier, 2013), Store layouts redesign to place such products together (Fawcett & Waller, 2013), Redesign Routes leading to the saving of millions of gallons of fuel (Davenport & Dyché, 2013), and Plan product strategy (Jeble, et al., 2018). Wells Fargo, Bank of America and Discover use BDA to understand their customer relationship through big data gathered from various data resources which were a mix of structured, semi-structured and unstructured data such as website clicks, ATM transactions, call centre logs, transaction records, and more. With using BDA, they built customer profiles, and this process helped to understand the reasons for customer attrition, customer opportunities and problems as well as to make decisions to attract customers (Davenport & Dyché, 2013).

2.5. Psychographic Segmentation

Kotler and Keller (2005) explain market segmentation as a process of partition a market into distinct subsets of customers, and distinct marketing mix, those customer subsets can be selected as a target market. Based on demographic, geographic, psychological and behavioural primary factors, market segmentation performs. By considering these factors, it gives a better knowledge of the relevant target audience, and it helps to prepare more effective market segments (Furnham & Gunter, 1992).

The psychological segment classifies customers concerning their values, attitudes, lifestyles, opinions and interests (Broderick & Pickton, 2005). Therefore, psychographic segmentation identifies two primary types of customers of lifestyle profiles and personality profiles (Furnham & Gunter, 1992). Further, Armstrong and Kotler (2005) state that some markets use personality variables in order to segment the market. Mintz (2017) points it out that the psychological segment is highly crucial to marketing since it has a higher impact when people purchase non-essential goods. Also, psychological factors caused to make a person's lifestyle (Mintz, 2017). However, Sandeep Vij (2012), states that many factors could affect to shape psychographic segmentation (Vij, 2012). Due to this complex situation, the Big Data analytics can be used to identify the psychographic segment behaviour of users with Big data (Huang, Liu, Liu, Wang, and Wang, 2019).

2.6. Clustering Analysis of Decision-Making with BDA

Clustering analysis has recognised as a method of solution for the multiple criteria decision-making problems (MCDM) (Ortt & Rezaei, 2013). Clustering performers, according to an unsupervised method, in order to divide data into several clusters or groups. In other words, clustering is the grouping data which has similar characteristics into a similar cluster, less similar characteristics into a different cluster. It uses in some different applications such as pattern classification, pattern analysis,

decision-making (Flynn, Jain & Murty, 1999). Faber, Hochberg, Kelly, Thomas and White (1994) have discussed different types of clustering techniques.

There is practical and empirical evidence that clustering analysis widely uses in MCDM (Alfian, Kwon & Sya, 2017). Clustering analysis has used to identify the optimal maintenance, rehabilitation and replacement alternatives. It also has applied by a broiler company to an analysis of supplier segmentation with regards to their willingness and capabilities (Ortt & Rezaei, 2013). Moreover, Keilera Röthlisberger and Zischg (2017) proposed clustering analysis to handle risk reduction problems; Wu and Xu (2018) have used it to break down the complex problem into smaller pieces; Delaney, de Bruin, Parker, and Strough (2015) have identified clustering analysis influences on decision-making processes-based differences on gender and age.

2.7. Customer Segmentation

Customer segmentation refers to the practice of grouping customer base into separate segments by considering similar interests, habits, lifestyle and other characteristics (Rouse, 2019). After a survey done by Fiedler and Rossberger (2014), they found that customer segmentation is a useful analytic marketing tool which can use in businesses. Further, they state that it delivers reliable insights for strategic marketing decisions. Judy Bayer (2010) has identified four customer segments which are critical to the telecommunication sector, namely, customer value segmentation, customer behaviour segmentation, customer life cycle segmentation, and customer migration segmentation (Fiedler & Rossberger, 2014).

Customer Segmentation in Telecommunication

Data which use for segmentation could contain a massive amount of data, with a range of customers demographics, educational achievements, profiles, incomes, and especially a wide array of evolving consumer demands. The real challenge is to analyse all these big data and transform it into an efficient and effective marketing

mix which contains price, product, place, promotion in order to satisfy the people, and processes (Dibb & Simkin, 2001). To achieve this complicated task, it BDA for customer segmentation with cluster analysis techniques. Especially with regarding the telecommunication sector, the amount of big data and number variables are very high.

3. METHODOLOGY

This research has followed **Positivism** philosophy with **Deductive** approach because it does not examine new knowledge but existing inductive theories. Since it has the objective of finding an application of BDA for strategic marketing decisions, **Action research** will take place as the research strategy. Moreover, **Quantitative mono method** would be followed as the research method since it investigates on URL data with a **Cross-sectional** time horizon for data collection method due to this study examines the customer base in the present time (has a specific moment).

3.1. Sample Method for Data Collection

The data collection for this study has been done with considering the objective of using Big Data Analytics to use for effective strategic marketing decisions. After making a series of discussions with the Chief Operating Officer (COO), Chief Information Officer (CIO) and top level of the marketing department of the telecommunication company, the researcher identified their requirement of customer segmentation in order to make effective marketing decisions. To market new products and services to the target customer, the telco company has a separate cloud-based system. However, the company does not have a proper method to cluster customers based on their Psychographic segment.

Therefore, this action research finds how BDA could be used for customer segmentation. The customer base will segment into six segments of entertainment, NEWS, sports, business, health care, and education according to their interests by analysing URL (Uniform Resource Locators) data.

Population

The population of this study or the customer base of the telco company is 2 million+ and among those customers, 966,925 broadband users. Therefore, the target population is all those internet users of 966,925 (Telco company Annual Report, 2019).

Sample Size

After analysing the broadband customer base, two samples have been identified—the first sample based on the highest usage of the internet for the past thirty days (Sample A). The expectation of selecting the highest usage of the internet was to collect a higher number of most popular URLs. For this sample, the top 100 customers were considered. For the second sample, another 150 customers were randomly chosen who has been active users during the past thirty days (Sample B). The expectation of this random sample was to select unbiased, genuine sample of the population. Therefore, a total of 250 customers were selected in two samples.

In order to analysis, the Psychographic segment of the broadband customer base, the researcher finds that the user accessed Uniform Resource Locators (URLs) to be analysed. Therefore, the user ID, devinfo (device information) and the previously accessed URLs in the past 30 days of 250 samples were extracted from the Hadoop ecosystem at the telco company. The data collection is about 98 gigabits of data in CSV format.

4. DATA ANALYSIS

Today in the world of business, data analysis plays a significant role in making decisions more scientific and supporting businesses operate more effectively (Gong & Xia, 2015).

Sample A and Sample B data collection which has been dumped from the telco company includes data of Customer ID (IDs), Device information (devinfo), and Accessed URLs (url). Customer ID has been hashed (encrypted) in order to anonymous the actual users. A sample data set is shown in *Figure 2*. Also, some of Data tuples has missing values in Nan format. Those Nan data tuples have removed in the first place.

Figure 1: Sample data set

IDs	devinfo	url
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	Unknown	exmaple.com
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	Unknown	exmaple.com
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	Unknown	exmaple.com
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	Unknown	exmaple.com
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	Unknown	exmaple.com

Step – 1 URL Sorting

URL data contains many servers, hosts details and IP addresses as well. Therefore, the same URL domain may appear in different formats. As an example, the domain of 'google' appears in 'clients3.google.com', 'www.google.com' and 'google.lk' URL formats. To avoid the error can occur in this situation, it has to split URLs and extract only the domains. Therefore, by running a simple algorithm with 'split' and 'join' methods, it can extract only the domains from the URLs and replace existing URL data with extracted domains. A sample data set after obtaining domains is shown in (*Figure 3*).

Figure 2: Data set after obtaining domains

904ad7c6ccba5d053eae58d78d37116d548c690baba6dc423186671243f7b592	ceipmsn
0414eaff716cb8dd043f2368da0049514bd0936dc2f53b74c24e70a29c2c4974	charityengine
0414eaff716cb8dd043f2368da0049514bd0936dc2f53b74c24e70a29c2c4974	neotracki
0414eaff716cb8dd043f2368da0049514bd0936dc2f53b74c24e70a29c2c4974	teamviewer
9a9937192fc7e3e75eded50adb6f01c270b061ea7f89bf52cb30bc3c4c8ced08	exmaple

The researcher has identified four types of unidentified domain and URL types. Those are IP address (such as 178.128.222.128), search engines (such as Google), dummy domains (such as exmaple.com), and computer requests (such as i5.c.eset.com, adobe). In order to remove all unnecessary domains, the researcher has used the method of maintaining a blacklist. All unnecessary domains will be added to the blacklist and remove data rows which are listed under the blacklist. A sample data set of the blacklist is shown in (*Figure 4*).

Figure 3: Blacklist Domains

blacklist
exmaple
google
ceipmsn
microsoft
teamviewer
firefox

In order to identify the top 250 domains, another algorithm was introduced to count the number of unique domains and list according to the higher to lower frequency. A sample data set of the top 250 domains are shown in (*Figure 5*).

Figure 4: Top 250 domains

domain
facebook
youtube
alibaba
office365
etisalat
lankadeepa
bbc
godaddy

After identifying the top 250 domains, it has to visit each domain manually and should label into the relevant segment. A sample labelled data set of the top 250 domains is shown in (*Figure 6*).

Figure 5: Labelled data set of the top 250 domains

domain	segment
facebook	entertainment
youtube	entertainment
alibaba	business
office365	business
etisalat	business
lankadeepa	news
bbc	news
godaddy	business

At the end of Step 1, the researcher has identified the most user accessing 250 domains and further, it has categorised into six segments which the company expects to segment the customer base.

Step – 2 Labelling Customers

Sample B data set will be used to identify and label the customers appropriately. For supervised learning, pre-processing data with labelling is essential (Bie, Braga & Maia, 2009).

After cleaning the URLs, the domains have to obtain with the suitable Customer ID. Then each domain has to compare with listed top 250 domains and filter only the domains similar to the top 250 list. A sample data set with Customer ID and domains is shown in (*Figure 7*).

Figure 6: Data set with Customer ID and domains

c4d779a3fb0355188c49dffbd9c5ee65eae4d1be86152cee923640940b633865	adaderana
fe890eae32c36ac613dbfb563787905f51bce56dd5dcf1a07949a8b1f5341743	youtube
9da715fef51a727e5380e3522dfc17ce1db5711d564db64b8d2fd5ebce385d20	xiaomi
cc5ff3ed17de40ffe2620138b21ac7276385e0f8cd0f7cd3d5344268363d6815	bbc

Then after, each domain has to remapped with the labels of the top 250 shown in *Figure 6*, in order to identify the segment of each data row concerning the given segments. A sample data set with the Customer ID, domains and the mapping are shown in (*Figure 8*).

Figure 7: Data set with the Customer ID, domains and the mapping

c4d779a3fb0355188c49dffbd9c5ee65eae4d1be86152cee923640940b633865	adaderana	news
fe890eae32c36ac613dbfb563787905f51bce56dd5dcf1a07949a8b1f5341743	youtube	entertainment
9da715fef51a727e5380e3522dfc17ce1db5711d564db64b8d2fd5ebce385d20	xiaomi	business
cc5ff3ed17de40ffe2620138b21ac7276385e0f8cd0f7cd3d5344268363d6815	bbc	news

In order to label customers to suitable segment, The next task is to label each customer into the appropriate segment that they belong. The researcher observes that most of the customers have a height number of 'entertainment' category domain accessed, and the lowest number of 'health' category domain accessed. Therefore, it has to filter the

customers who have more interest in a specific category more than others. In order to perform this task, the researcher finds that creating a new ratio, 'segment ratio' introduced. This ratio will construct by counting the number of domains with relevant to each segment, and it divides from the count of all the domains.

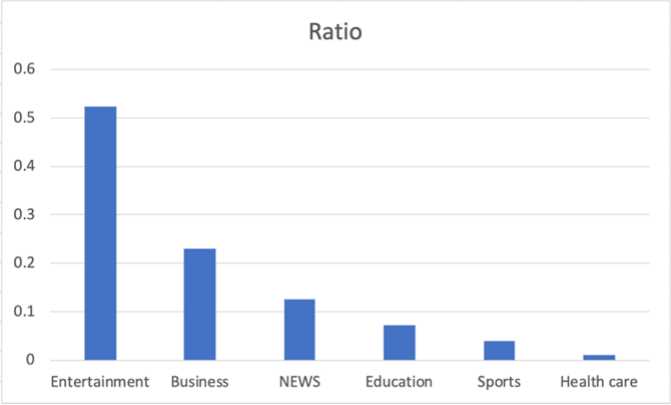
seg_rat = (number of domains in a segment) / (total number of domains)

Figure 9 shows the six ratios of 'segment ratio', and Figure 10 shows the 'segment ratio' in a bar chart.

Figure 8: Segment ratio

Segemnt	Ratio
Entertainment	0.523852049
Business	0.229386736
NEWS	0.125987369
Education	0.072284741
Sports	0.038930051
Health care	0.009559054

Figure 9: Segment ratio chart



The customer classification algorithm in Step 2 will be done in two sub-steps. First, calculating the 'customer segment ratio' (cus_seg_rat) for all six segments by counting the number of domains with relevant to each segment for each customer, and it divides from the count of all the domains for each customer.

cus_seg_rat = (number of domains in a segment of a customer) / (total number of domains of a customer)

Secondly, the 'customer segment ratio' will reduce by the 'segment ratio', and the answer will compare it with each segment answers. Then the segment which holds the largest value will be the category of the customer.

category = segment_name{max(cus_seg_rat - seg_rat)}

By using this algorithm, every customer in Sample B has been labelled. The intention of Step 2 is to create labels to each customer and create targets for the supervised learning process for the machine learning algorithm in Step 3. A sample data set with the Customer ID and the segment label is shown in (Figure 11).

Figure 10: Data set with the Customer ID and the segment label

customer_id	label
fe890eae32c36ac613dbfb563787905f51bce56dd5dcf1a07949a8b1f5341743	business
cc5ff3ed17de40ffe2620138b21ac7276385e0f8cd0f7cd3d5344268363d6815	entertainment
6105cfd8d440a77e7b79d4d2a447863c22aeead35aa7cd6b4dd3663ca6353db9	entertainment
2e456bde22addb2e9d734e08f51f314d1b916b0aacd60e4b21c394fbe762bfdd	news
8c85304efdfb53a98d7f6d1058cec8ab0e523d53dd7dfcbac9a01c37dc90e67c	entertainment

Step 3 - Training NLP Algorithm

The objective of Step 3 is to apply Big Data Analytics techniques on this Action research by constructing an Artificial Neural Network (ANN) in order to cluster customers (Brinkkemper, Fotaki, Meijer & Spruit, 2014).

The importance of creating ANN to cluster the customer base into a relevant segment is, it can perform the task much faster and considering other facts that were missing in Step 1 and Step 2. In an ANN, it is possible to input data without pre-processing (without extracting Domains) since the machine learning algorithm processes the input as it is. This specific task has involved with String data (non-numeric data). Therefore, the ANN constructs with NLP (Natural Language Processing) techniques associate with a supervised clustering algorithm. The hybrid algorithm is consisting of Natural Language Toolkit (NLTK) (Cohen & Verspoor, 2013) and k-nearest neighbours (KNN). So, the training of ANN, the input data format is a 3-dimension (3D) array of customer Id, URLs, and label. This method of data analysing using Big Data has been practised in previous studies (Dave, 2018; Evgeniou, 2018).

(customer_id, (URL_1, URL_2, ..., URL_n), lable_id)

The training split has done in 3:1 ratio of input into training: testing on a random selection basis. After performing five epochs, the best result came by 63.56% of accuracy. However, the accuracy level would make high by adjusting the optimisers; the accuracy level increased up to 68.222%. After training the ANN, the machine saved, and it is ready to deploy to the telco servers.

Step 4 - Implementing the Algorithm in Telco Servers

The ANN machine can be directly transferred into the telco servers. However, there is a significant move that can be improved. The severs at the telco company is running in a Hadoop environment. Therefore, the performing level of this ANN algorithm can be significantly improved by rewriting the script in the PySpark language.

4.1. Overall of the Application

After implementing the algorithm and performing customer segmentation in broadband user customer base in the telco company (approximately 1 million users), the customer base will be categorised by Psychographic segment of interests. When the marketing division wanted to make effective strategic marketing decisions, the managers can identify customers by the customer ID and make marketing decisions according to each customer Psychographic segment. Especially, these insights can be used in promotion and advertising platform to host target market campaigns.

5. CONCLUSION, DISCUSSION AND RECOMMENDATION

The research is intended to investigate how Big Data Analytics can be applied to make effective strategic marketing decisions. In order to study the context, the researcher examined the empirical evidence regarding both BDA and strategic marketing decisions.

5.1. Conclusion

The telecommunication sector companies have to involve a massive number of data every second. Also, these data contain various essential insights of the customers in various formats. However, extracting insights from these data with traditional statistical methods have various limitations, and it is highly challenging. Therefore, to analysis such Big data, it has to use the modern analytical method of Big Data Analytics. This research examined how BDA can be used to extract valuable information in the business and marketing point of view.

The study investigated the process and the importance of strategic marketing, including segmentation, targeting and positioning (S-T-P). Further, it drilled down market segmentation process and reflected the value of segmenting customers based on Psychographic segment. The research clustered the broadband customer base of the Telco company based on its Psychographic segment with using BDA. The Psychographic segment contains the interests of people and their lifestyles (Broderick & Pickton, 2005). Therefore, these individual clusters can be used in target marketing (Camilleri, 2018), and also those distinct features of customers in each segment can be used to make effective strategic marketing decisions. This analysis was based on the URLs data of broadband customers. After the pre-processing the data, customers ware manually labelled into six segments using logical algorithms for supervised learning purpose. Then after using NLTK and KNN hybrid machine learning ANN algorithm, the clustering has been done. This application is a clear indication of the possibility of using BDA for modern data analytics for businesses. Further, after the optimisation, this model performed with an accuracy level of 68%, which is highly challenging to achieve by using traditional analytical methods.

5.2. Discussion

The research findings of this research are highly valuable for any large - scale companies. After performing the customer segmentation for the entire broadband customer base, then the marketing division can take strategic decisions based on the characteristics of an individual segment. This telco company gets to market new products which can categorise under six segments (entertainment, NEWS, sports, business, health care, and education) mainly. Today the promotions have been done by random customer selection basis. After segmenting customer base using BDA techniques, the company will be able to perform target marketing campaigns for potential customers with a higher level of productivity. Furthermore, this BDA application reflects that BDA is an efficient method of analysing big data to extract essential and critical insights for making effective strategic marketing decisions.

Therefore, switching from traditional analyses methods, modern big data algorithms can improve the decision-making process fast and more accurate in any industry.

5.3. Recommendations

As it pointed out, selecting random customers for promotions and marketing can have a low level of productivity. Therefore, it can recommend implementing BDA based customer segmentation and based on the results of the clustering, planning the marketing decisions.

To improve the machine learning algorithm, it can recommend several modifications to the process. For this research, the sample has only considered one month of URL data. If a certain month had a significant event or particular incident, then those data can contain errors. Therefore, the researcher recommends considering more than two months of data as the time frame. Further, Sample A has been considered of the top 100 users who had most internet usage in a month. However, after the analysis of these users, it was visible that the majority of those users were categorised under the entertainment category. Therefore, the researcher recommends selecting a large number of random users for Sample A as well. Then the top 250 domains will consist of balanced domains for each category. To improve the training to the machine, it can recommend increasing the number of training data set (Sample B). This research had the ability to identify only the router-based customer (ADS). It can recommend to cross analysis the data with 'device information' to recognise individual customer level.

The telco company operates in a Hadoop ecosystem. By using multiple query engines, it is possible to introduce the Apache Spark environment to the Hadoop ecosystem. Therefore, the researcher recommends implementing parallel Spark ecosystem to process BDA. The Spark ecosystem has a higher level of support towards machine learning algorithms than the Hadoop ecosystem (White, 2012). This implementation will express the speed of processing any BDA method, including customer segments, to improve providing real-time insights. Imran and Yayah (2017) suggest that BDA

can use in telecommunication companies for Network Infrastructure Service, Churn Prevention, Customer Relationship and Experience, Product Development and Fraud Detection. Offering new products to the customers is an essential requirement to be retained as the market leader in the telecommunication sector (Phavy, 2013). However, recognising new customer needs is a challenging objective with a lack of information. Therefore, telecommunication companies have established new projects to collect data on user behaviour in tablets and mobile phones in order to create buying patterns using BDA. These projects have been significantly improved in the revenue of the companies (Imran, et al., 2017). Hence, it can recommend to the Telco company to implement BDA in other applications such as network monitoring, forecasting and identifying consumer patterns relating to telecommunication.

5.4. REFERENCES

Agrawal, S., Pate, B., Patel, A. & Verma, J. P., 2016. Big Data Analytics: Challenges And Applications For Text, Audio, Video, And Social Media Data. *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI),* 5(1), pp. 41-51.

Akter, S.; Chopin, G.; Edwards, A.; Fosso Wamba, S.; Gnanzou, D., 2015. "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study,". *International Journal of Production Economics..*

Alexander, C. A. & Wang, L., 2015. Big Data Driven Supply Chain Management and Business Administration.. *American Journal of Economics and Business Administration,* Volume 7, p. 60–67.

Ali, H., Sultan, K. & Zhang, Z., 2018. Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks. *Institute of Electrical and Electronics Engineers Inc.,* Volume 6.

Armstrong, G. & Kotler, P., 2005. *Marketing: An Introduction.* 7th ed. Upper Saddle River: N.J. Prentice Hall.

Annual Deport Dialog (PLC), 2018. *Annual Deport* , s.l.: Annual Deport Dialog (PLC).

Annual Report Hutch (PVT), 2018. *Annual Report Hutch (PVT),* s.l.: Hutch (PVT).

Annual Report SLT (PLC), 2019. *Annual Report SLT (PLC),* s.l.: SLT (PLC).

Armour, F., Espinosa, Kaisler, . S., J. A. & Money, W., 2013. *Big Data: Issues and Challenges Moving Forward.* s.l., 2013 46th Hawaii International Conference on System Scien 1530-1605/12, 2012 IEEE.

Begoli, E. & Horey, J., 2012. ―Design Principles for Effective Knowledge Discovery from Big Data. *2012 Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture.*

Bengio, Y. Hinton, G., 7 LeCun, Y., 2015. Deep Learning.. *Nature*, Volume 521, pp. 436-44.

Beyer, M. A. & Laney, D., 2012. *The Importance of "Big Data": A Definition..* Stamford, CT, s.n.

Bie, T. D., Maia, T. T. & Braga, A. P., 2009. Machine Learning with Labeled and Unlabeled Data. *European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning.* , Volume 9, pp. 22-24.

Bifet, & A. Fan, W., 2014. Mining Big Data, Current Status and Forecast to the Future. *SIGKDD, Explorations*, 14(2), pp. 1-5.

Broderick, A. & Pickton, D., 2005. *Chapter 17: Identifying target audiences and profiling target markets.* 2nd Ed ed. In Pickton, David & Broderick: Amanda: Integrated marketing communications.

Camilleri, M., 2018. Market Segmentation, Targeting and Positioning. *Travel Marketing, Tourism Economics and the Airline Product* , Volume 4, pp. 69-83.

CBSL, 2019. *Economic and Social Statistics of Sri Lanka – 2019*, Colombo: Central bank of Sri Lanka .

Chen, C. & Zhang, C.-Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on big data.. *Information Sciences - Journal*, Volume 275, p. 314–347.

Chen, M., Mao, S. & Liu, Y., 2014. *Big Data: A Survey.* New York: Springer Science+Business Media.

Cohen, K. B. & Verspoor, K., 2013. *Natural Language Processing.* s.l.:s.n.

Davenport, T. H. & Dyché, J., 2013. *Big data in big companies..* s.l.:International Institute for Analytics.

Davenport, T. H., 2014. How strategists use 'big data' to support internal business decisions, discovery and production.. *Strategy & Leadership.*, 42(4), p. 45 – 50..

Demchenko, Y., Grosso, P., De Laat, C. & Membrey, P., 2013. Addressing big data issues in Scientific Data Infrastructure. *IEEE International Conference on Collaboration Technologies and Systems*, p. 48–55.

Dave, R., 2018. *towardsdatascience.com.* [Online] Available at: https://towardsdatascience.com/industrial-classification-of-websites-by-machine-learning-with-hands-on-python-3761b1b530f1 [Accessed 28 4 2020].

Dibb & Simkin, 2001. Market Segmentation: Diagnosing and Treating the Barriers.. *Industrial Marketing Management*, 30(8), p. 609–625.

Dutta, P., 2019. *ubuntupit.com.* [Online] Available at: https://www.ubuntupit.com/best-big-data-applications-in-todays-world/ [Accessed 28 4 2020].

Evgeniou, T., 2018. *inseaddataanalytics.* [Online] Available at: https://inseaddataanalytics.github.io/INSEADAnalytics/CourseSessions/Sessions45/ClusterAnalysisReading.html [Accessed 28 4 2020].

Fawcett, S. E. & Waller, M. A., 2013. Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain.. *Journal of Business Logistics*, 34(4), pp. 249-252.

Fiedler, M. & Rossberger, T., 2014. Customer Segmentation Analysis in Major Sporting Goods Companies and its Influence on Strategic Marketing Decisions. *Review of European Studies,* 6(4), pp. 268-276.

Furnham, A. & Gunter, B., 1992. *Consumer Profiles: An Introduction to Psychographics (Consumer Research and Policy).* London: Routledge.

Gandomi, A. & Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* , 35(2).

Gong, . P. & Xia, B. S., 2014. Review of business intelligence through data analysis.. *Benchmarking: An International Journal.,* Volume 21, p. 300– 311.

Gong, P. & Xia, B. S., 2015. Review of business intelligence through data analysis. *Benchmarking,* 21(2), pp. 300-311.

Gong, . P. & Xia, B. S., 2014. Review of business intelligence through data analysis..

Aheleroff, S., 2011. Customer Segmentation For a Mobile Telecommunications Company Based on Service Usage Behavior.

Haije, E. G., 2019. *Top 15 Business Intelligence Tools: An Overview.* [Online] Available at: https://mopinion.com/business-intelligence-bi-tools-overview/ [Accessed 15 3 2020].

Hurbean, L., 2006. Business intelligence: applications, trends, and strategies. *Analele Stiintifice ale Universitatii "Alexandru Ioan Cuza" din Iasi - Stiinte Economice,* Volume 52, pp. 307-312.

Imran, G. K., Yayah, F. C. & Ting, C.-Y., 2017. Adopting Big Data Analytics Strategy in Telecommunication Industry. *Journal of Computer Science & Computational Mathematics,,* 7(3), pp. 57-67.

Jain, A., Flynn, P., & Murty, M.1999. Data Clustering: A Review.. *ACM Computing Surveys,* Volume 31, p. 264–323.

Jeble, S., Kumari, S. & Patil, Y., 2018. Role of Big Data in Decision Making. *OPERATIONS AND SUPPLY CHAIN MANAGEMENT,* 11(1), pp. 36 - 44.

Kalton, G. & Moser, C. A., 1971. *"Survey methods in social investigation." Survey methods in social investigation.* 2nd Ed ed. s.l.:s.n.

Kimble, C. & Milolidakis, G., 2015. Big Data and Business Intelligence: Debunking the Myths.. *Global Business and Organizational Excellence.,* Volume 35, p. 23 – 34.

Kotler, P. & Keller, K., 2009. *Marketing management.* 13th Ed ed. Prentice Hall: Pearson.

Kumar, M., Wang, S. & Yadav, C., 2013. Algorithm and Approaches to handle large Data- A Survey. *IJCSN*, 2(3).

Mayer-Schönberger, V. & Cukier, K., 2013. *Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt..* [Online] Available at: www.amazon.in/Big-Data-Revolution-Transform-Think/dp/0544227751 [Accessed 12 3 2020].

Mauro, A., Greco, M. & Grimaldi, M., 2016. A Formal Definition of Big Data Based on its Essential Features. *Library Review,* 65(3), p. 122 – 135.

McAfee, A. et al., 2012. Big data. *The management revolution. Harvard Business Review,* 90(10), pp. 61-67.

Mintz, J., 2017. Psychographic Segmentation. *CEPSM.*

Nicholson, C., 2019. *pathmind.* [Online] Available at: https://pathmind.com/wiki/machine-learning-algorithms [Accessed 10 3 2020].

Ortt, R. & Rezaei, J., 2013. Multi-criteria supplier segmentation using a fuzzy preference relations based AHP. *European Journal of Operational Research,* Volume 225, p. 75–84.

Osuagwu, L., 2016. A Model of Strategic Marketing Decision Premises. *International Journal of Marketing Studies,* 8(3), pp. 145-153.

Petrov, C., 2019 . *techjury.net.* [Online] Available at: https://techjury.net/stats-about/big-data-statistics/ [Accessed 20 7 2019].

Phavy, O., 2013. *New Service Development Process in Telecom Industry: The Telefonica Case Study.* 8th Ed ed. Stockholm: s.n.

Riahi, Y. & Riahi, S., 2018. Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering,* 5(9), pp. 524-528.

Rojas, R., 1996. Neural Networks: A Systematic Introduction. *Springer-Verlag Berlin Heidelberg.*

Rouse, M., 2019. *searchcustomerexperience.* [Online] Available at: https://searchcustomerexperience.techtarget.com/definition/customer-segmentation [Accessed 5 4 2020].

Russom, P., 2011. The Three Vs of Big Data Analytics. *TDWI.*

Sya, M. A., Alfian, G. & Kwon, Y. S., 2017. Customer Purchasing Behavior Analysis as Alternatives for Supporting In-Store Green Marketing Decision-Making. *Sustainability*, Volume 9.

Shaw, M. J., Subramaniam, C., Tan, G. W. & Welge, M. E., 2001. Knowledge management and data mining for marketing.. *Decision support systems,* 31(1), pp. 127-137.

Schroeck, M. et al., 2012. *Analytics: The realworld use of big data.* New York, NY: IBM Institute for Business Value , Said Business School.

Sun, L. & Zhu, Q., 2020. Big Data Driven Anomaly Detection for Cellular Networks. *nstitute of Electrical and Electronics Engineers Inc,* Volume 8, pp. 31398-31408.

Sun, Y., Wang, Z., Wei, G., & Zhan, Y., 2017. Big data in telecommunication operators: data, platform and practices. *Journal of Communications and Information Networks,* 2(3), pp. 78-91.

Teradata, 2014. Big Data Framework for Telecommunications. *Teradata.*

Davenport, T. H. & Dyché, J., 2013. Big data in big companies.. *International Institute for Analytics.*.

Vij, S., 2012. Psychographic Segmentation Based on Belief Factors Underlying Attitude toward Advertising in General. *International Journal for Management Research* , 1(4), pp. 17-28.

White, M., 2012. Digital workplaces: Vision and reality.. *Business Information Review*.

White, T., 2012. *Hadoop: The definitive guide.* s.l.:O'Reilly Media, Inc..

Wu, Z. & Xu, J., 2018. A consensus model for large-scale group decision making with hesitant fuzzy information and changeable clusters.. *Search Results Web results Information Fusion - Journal,* Volume 41, p. 217–231.