Design and Implementation of a Marking Strategy to Increase The Contact Ability in The Call Centers Based on Machine Learning

Leonel Hernandez[#], Henner Guzman[#], Jose Ospino[#], Jean Freyle[#], Andri Pranolo[*]

[#] Engineering Faculty, ITSA Institución Universitaria, Carrera 45# 48-31, Barranquilla, Colombia
E-mail: lhernandezc@itsa.edu.co, hguzman@itsa.edu.co, jospino@itsa.edu.co, jfreyle@itsa.edu.co

[*] Informatics Department, Universitas Ahmad Dahlan, Jln. Kapas No 9 Semaki, Yogyakarta, Indonesia
E-mail: andri.pranolo@tif.uad.ac.id

Abstract—**Jamar is a company that belongs to the furniture sector, which manufactures and sells furniture and accessories for the home. Customer calls are one of the most trusted channels used in contact centers. Currently, the contactability indicator has a goal of 40% and is at 31%. The enemies of the efficiency of this channel are the terrible dimensioning, customers who evade answering these calls by identifying the numbers, non-market numbers in the databases, failures in the technological resources. Therefore, a proposal was made to design and implement a marking strategy in the call center, supported by a statistical model for dimensioning. Likewise, emerging technology such as Machine Learning is performed to help the marking strategy in outbound campaigns, reconfiguration of the dialplan to make it more efficient, and a redundant architecture design in the operators. Basic concepts of Teletraffic are explained, showing its primary functions, relevant for the management of the company's telephone system. In the same way, fundamentals of the Asterisk IP PBX are exposed, one of the most used in our environment due to its versatility and low implementation cost. The methodology of descriptive and applied research is used for the development of the project. The results and discussion show the dialing strategy and some call statistics from previous years, necessary to establish a correct dimensioning of the solution. The proposed solution allows having redundancy management for SIP and trunk operators, to have backup and reliability in case of failure.**

Keywords—**network; calls; dialplan; machine learning; asterisk.**

I. INTRODUCTION

Currently, in the Colombian market are involved many outsourcing companies of goods and services, such as call centers, which have provided and provide a customer service line where agents are available to help solving all problems. Agents can solve without having to go to the office to file a problem and also provide the line. Also, these companies provide the service to customers to resolve any concerns about the product portfolio, service sales, and collection management. In this way, the client finds all the support and leaves satisfied. For more than 15 years in Colombia, the necessary steps have been taken to outsource processes through the call center industry. This process of improvement has been dedicated only to voice, nowadays a more globalized concept focused on the BPO has been involved [1], [2].

Willing to change, there has been an improving trend in the Colombian call center and BPO industry in the last three years compared to 20 years ago. This change associated with the technological and market needs have to do with a vital interest in the growth and a proposal to become an emerging

sector of the Colombian economy and above all a World Class sector. In summary, the business of outsourcing in Colombia continues to progress by leaps and bounds, since the estimates of the Colombian Association of Call Center and BPO suggest that the sector's sales amounted to 1.7 billion pesos during 2012. This estimate means an increase of 11.21 percent compared to 2011 when its revenues reached 1.5 billion pesos [1]. The Colombian services sector shows multiple opportunities since it has sustained development and has become one of the most significant commercial bets in the country. Beyond a phone call to meet the requirements of the public or contact the customer, many software and hardware solutions are hidden. These software and hardware solutions support the customer service and collection scheme behind the telephone. These solutions have drawbacks such as errors in calls, such as clients that evade calls and repetitive failures at the software level. This enemy affects the efficiency of the call center, generating multiple reprocesses on its operations. Despite this, many companies neglect this aspect of vital importance, given that

they do not have the notion of the damage they generate to the process.

For all the above, this work aims to design and implement a marking strategy in Jamar, to sustain growth in the call center and make it more efficient, having statistical models to size and predict traffic. As well algorithms with emerging technologies such as Machine Learning to prevent the client from evading the call [3], database engines to have information updated and an information model to have the most efficient contact numbers.

As mentioned above, there are problems for the proper management of the call center, and to make it more efficient, in such a way that it is a factor that fosters or enhances the business of the company. Among these drawbacks are dialing as serial tone numbers, calls are` hung, calls with interference, ambulance tone, noise call, mute call. Additional some clients evade calls by identifying the numbers of the call center. Repeated failures at a system level affect the entire operation.

The big enemies of these software and hardware solutions are customer databases with wrong or unmarked numbers, such as shorter numbers, more extended numbers, alphanumeric data, no area code. Other possible problems are mobile call restriction, fixed call restriction, restriction to LDN call, restriction to call to LDI, empty field, non-contact numbers such as occupied, fax, answering machine, damaged, ringing and not responding, without sound, saturated. Other problems that are related are lousy sizing in technological resources such as SIP trunks, lack of monitoring of technical resources, lack of a strategy of marking customers. Finally, the most critical inconvenience has a marking system architecture without redundancy or contingency plans.

These enemies affect the correct operation of call centers, leading to rework in the business, inefficient call center agents, considerable loss of money due to continuous interruptions of the dialing system. These situations end up impacting the leading indicators of the call center, therefore, the monthly business goals. They impede the growth and development of this, stagnating the interests of the business in the short and long term.

There are challenges to have an efficient call center and meet business indicators such as contact ability. An adequate architecture is required to guarantee the continuity of the business, as a redundancy in the outbound operators, allowing overflowing calls in case of failures, redundancy in servers that support these technological solutions [4]. Likewise, a dimensioning must be achieved to handle all call traffic, get the technology to typify the results of the calls to feed the database, which seeks to reduce the percentage of non-dialing or non-contact numbers, design a customer-facing dialing strategy, to prevent the customer from identifying calls in outbound collections campaigns. Finally, monitoring of the states of each technological component to be more proactive before future failures [5].

II. Materials and Method

A. Materials

1) Machine Learning: Machine Learning is a scientific subject in the field of Artificial Intelligence that creates systems that learn automatically. Learning in this context means identifying complex patterns in millions of data. The machine that learns is an algorithm that reviews the data and can predict forthcoming behavior. Inevitably, also in this context, implies that these systems are improved autonomously over time, without human intervention [8] oriented to the Internet of things to allow the sending of alerts through an email. Another essential aspect to consider is the gateway, which allows the interconnection between the sensor network and a TCP / IP network. Finally, it is necessary to have the software required in which the application will be executed to obtain the information of the microcontrollers. This information is the one that will come from the node of the sensor. Machine Learning is a tool that allows finding patterns in the data, will allow us to study volumes of data that our brain is unable to analyze efficiently [9], [10]. This technology dates back to 1953, only now the focus of information systems has changed, and they are being created in a different way to how the discipline of machine learning has been carried out and popularized. The structure of the algorithm is shown in Figure1:
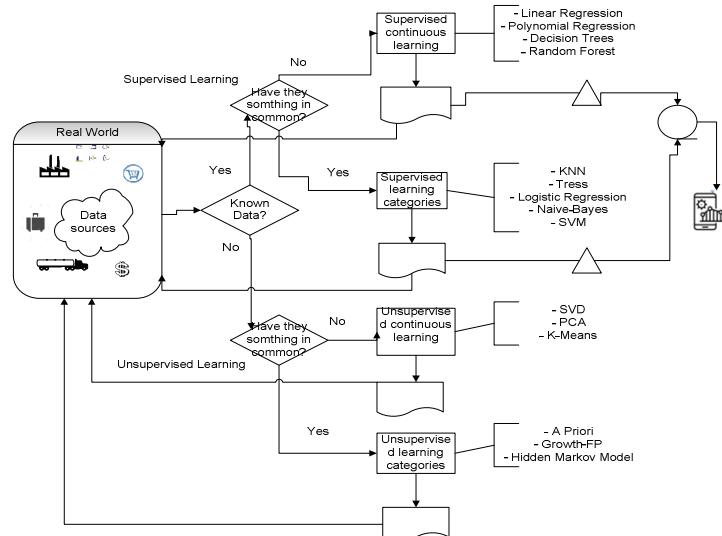

Fig. 1. Structure of the Machine Learning Algorithm

It is the systematic use of algorithms to process a vast amount of data (Big Data). It needs a series of algorithms that all together form the subfield of artificial intelligence called Machine Learning. To achieve the operation of the algorithm, the following aspects are required:

- Supervised Learning: for the known data.
- Unsupervised Learning: for unknown data.
- Continuous Learning: the data is made available in sequential order and for which some type of regression or grouping is needed, so the algorithm must adjust dynamically.
- Learning in Categories: some classification in the algorithm groups the data.
- Continuous Supervised Learning: based on regression which can be (Linear, Polynomial, tree or random forest).
- Continuous Non-Supervised Learning: based on Clustering and Dimensional Reduction.
- Unsupervised Learning in Categories: based on analysis of association either a priori or many patterns. Moreover, the Markov model in which hidden parameters are determined.

2) Teletraffic: It is the application of probability theories to the solution of planning problems, performance evaluation, operation and maintenance of telecommunications systems [11]. Primary functions of Teletraffic Engineering are shown in Fig. 2 [12]:
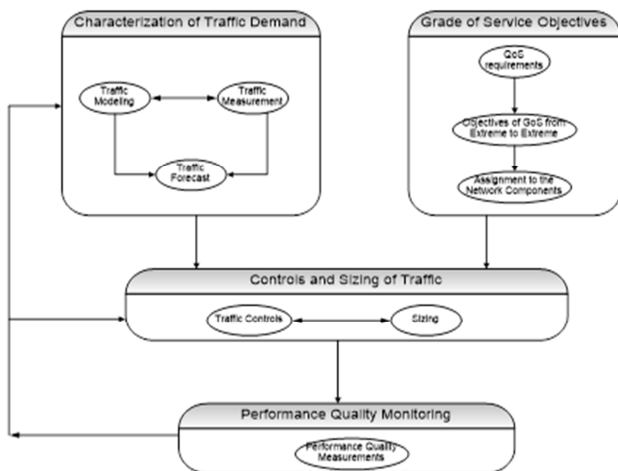

Fig. 2. Primary Functions of Teletraffic Engineering

It is considered a system with n identical resources, working in parallel.

- "Homogeneous group" (homogeneous group).
- A call is accepted in the system if there is at least one available resource, and each call occupies a single resource.
- "Complete accessibility" (full accessibility).

If all resources are busy, the system is "congested," and the call attempt is blocked.

- The call attempt in this case "disappears," there is no wait or retry.
- Loss Model (Lost Calls Cleared).

There is "congestion" when all the resources are busy, that is when the system is in the state "n":

- The probability that when arriving "arrives" finds the system in the state [n]" is equal to the stationary probability that the system is in the state [n].
- This event is known as the property "Poisson Arrivals See Time Average" or PASTA, demonstrated by Ronald W. Wolff in 1982.

There is congestion when all resources are busy, that is when the system is in the state [n]. One of the ways to find the probability of congestion is to use the formula of Erlang B. The traffic units that are analyzed in Teletraffic are the Erlang (One Erlang corresponds to an average traffic intensity of one hour per hour). Equivalent to a resource occupied in a permanent, and Hundreds of seconds per hour (HCS). An HCS corresponds to an average traffic intensity of 100 seconds per hour (can be associated with the duration of a typical call) 36 HCS = 1 E.

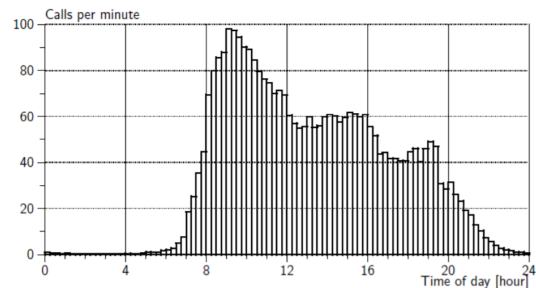The average number of calls per unit of time is shown in Fig. 3:

Fig. 3. Calls per Minute

The average duration of the occupation time, for example, the average length of the calls (mean holding time), is shown in Fig. 4:
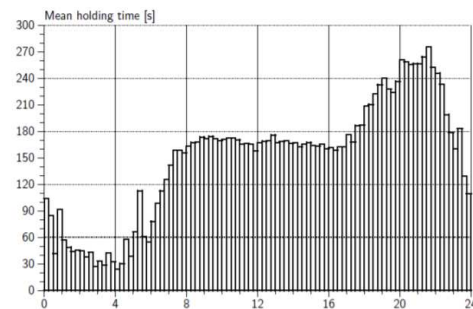

Fig. 4. Mean Holding Time

Figure 5 shows a traffic diagram based on call density and average occupancy:
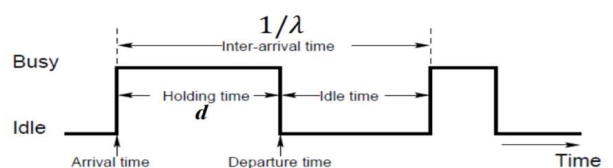

Fig. 5. Traffic Diagram

3

The different types of traffic are shown in Fig. 6. Some other important concepts are:

- Traffic Offered: It is the traffic that would be coursed if there was no rejection within the network. It could be done with "infinite" resources.
- Traffic Cursed: It is the traffic that went through the network. Typically "can be measured."
- Lost or Rejected Traffic: It is the traffic that could not be traced by the network.
- Overflow Traffic: It is the traffic that could not be traced by a network and is derived from another network.
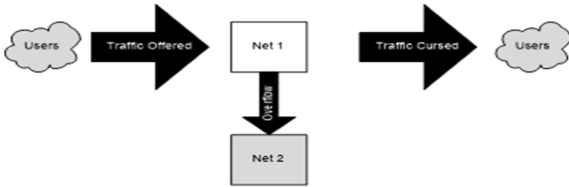

Fig. 6. Types of Traffic

3) Introduction to Asterisk: Asterisk is a free software program that offers functions of a telephone exchange (PBX). Any PBX can be connected to a specific telephone for telephones with access within the same organization and even access to communication outside of it through PSTN or connecting a VoIP provider or an ISDN as well as the primary [13]. Asterisk includes many features that were previously only available in expensive proprietary PBX systems, such as voicemail, conferencing, IVR, automatic call distribution, and many others. Users can create new features by writing a dialplan in the Asterisk script language or by adding modules written in C language or any other programming language supported on GNU / Linux [14].

Digium or other suppliers manufacture the FXO and FXS phone cards used to connect standard analog telephones. The first consists of any device that acts as a traditional telephone, connected to a conventional telephone line, that is, it can give a dial tone, ringing, picking up and hanging up. The second is an internal telephone line generator, provides power, and electrical signals (voltages) necessary to generate the ringtone and to establish audio communication.

Perhaps the most exciting thing about Asterisk is that it recognizes many VoIP protocols such as SIP, H.323, IAX and MGCP [15], [16]. Asterisk can interoperate with IP terminals acting as a registrar and as a gateway between them. One of the strengths of the Asterisk software is that it allows the unification of technologies: VoIP, GSM, and PSTN. Asterisk is beginning to be adopted in some corporate environments as a great low-cost solution together with SER (Sip Express Router). In the architecture of the project, Asterisk plays an important role, because all the VoIP traffic arrives on it and from there all services are consumed for the operation of this [17], [18].

B. Research Methodology

For the development of the project, it was necessary to apply two types of research methodologies, descriptive and applied [6], [7]. Descriptive research, due to it allows reviewing all the documentation related to the low level of

contact ability that currently affects the company Jamar in its portfolio department. Likewise, information related to machine learning and the implementation of Asterisk PBX must be reviewed. This information is critical to adequately develop the tagging strategy and implement the solution to the problem stated above. Applied research, because the theory analyzed in the descriptive study will be used to solve the problem practically. The design proposal is marked as a solution to the current need in the portfolio department in its call center area.

The study population is all companies for the BPO sector that want to improve the marking strategy for outbound campaigns. The sample on which the investigation will be carried out is the furniture retailer Jamar located in the downtown neighborhood of Barranquilla, specifically the portfolio area, with a staff of one hundred and eight (108) employees subject to information. Given the size of the sample, it is required to use a sampling procedure because it is not possible to analyze it thoroughly.

Meeting the objectives established in this study, some techniques and tools must be used to collect the necessary data to support the research, such as observation (it is intended to appreciate and perceive with attention the characteristics of the marking processes), interview (it is oriented to establish direct contact with people who are considered sources of information) and document review.

III. RESULTS AND DISCUSSION

The strategy has two specific objectives, select the most efficient number of the client and second, the best header or DID to make the call. To achieve these objectives, the strategy has a machine learning component where it becomes robust with each call made because each call is typed by the machine when it is hung. Also, the typing will be called rejected, call answered, busy number, mailbox voice, wrong call by the format of the destination number, number canceled, number out of coverage, call failed due to an internal error of the server, header or DID use for the call. These typifications will be the input of a model that will be responsible for selecting the efficient number of the client and better header or DID; this service will be consumed once the agent of the call center executes the call to the client. The technology to be used is Asterisk, in the database, it is in the study stage to support the predictive model of the marking strategy. Fig. 7 shows the marking approach:
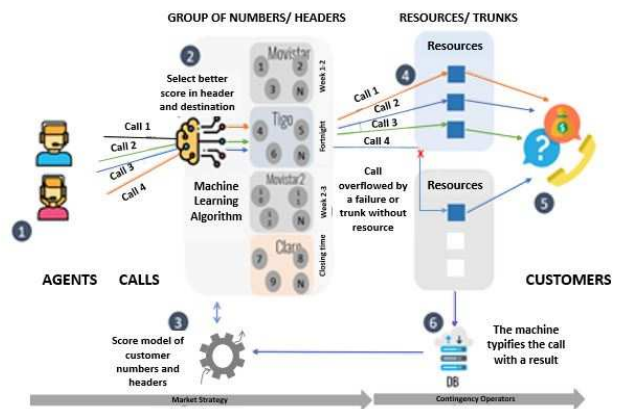

Fig. 7. Marking Strategy

In Figure 8 is shown a script with activities executed on the typing of calls:

Fig. 8 Dial Plan Implementation

Fig. 9 Traffic Outbound

TABLE I
TABLE OF TRAFFIC

Description	Value	Formula
The average duration in seconds	73	-
Calls/Time	3063	-
Time observation (sec)	3600	-
Traffic volume	0,850 833	Calls/Time Observation
Service rate	0.013 699	1/Average duration
Traffic Offered Erlang	62.11 083	Traffic volume/Service Rate
Congestion probability with 120 channels-redundant operation	0%	Erlang B
Congestion probability with 60 channels	12%	Erlang B

A. *Dimensioning with Statistical Model of Lost Technological Resources for the Traffic of Calls Outbound.*

Figure 9 shows an outbound traffic study was carried out for the years 2017 and 2018. The figure shows the number of monthly calls, the participation of calls by type of telephone (cellular or landline), and the participation of calls from the call center in the company and by last traffic per day of March 2018. It was necessary to carry out the study of traffic to perform a correct dimensioning:

To perform the outbound traffic sizing, it was necessary to validate the capacity of technological resources and intelligence of the traffic demand, being this input for a statistical model that would be used in sizing, the statistical model is supported in exponential and Poisson distribution with the model of losses or also called finite model resources [19] [20]. In Table I is shown the results and formula used:

Equation 1 shows the formula Erlang B:

$$p(n) = \frac{\frac{A^n}{n!}}{\sum_{j=0}^{n} \frac{A^j}{j!}} = E_B(n, A) \tag{1}$$

After applying the dimensioning, it is possible to reduce the probability of congestion to 0%, with a capacity to meet the estimated growth of 33 Erlang [21] [22]. A monitoring system allows verifying the equipment (hardware) and services (software) that are specified, warning when the

behavior of the same is not the desired one. It is a software that provides excellent versatility for consulting any parameter of interest of a system and generates alerts, which can be received by the responsible parties through (among other means) email and SMS messages when these parameters exceed the margins defined by the network administrator. Fig. 10 shows a monitoring system created in PHP that was implemented to perform a follow-up to the trunk lines:
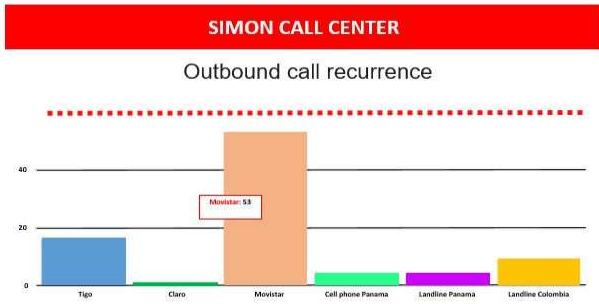

Fig. 10. Monitoring System

B. Design and Implementation of a Redundant Architecture of SIP Operators and Trunks

The redundancy refers to complete nodes that are replicated or components of these, as well as paths or other elements of the network that are repeated and one of its primary functions is to be used in case there is a system crash. In this case, we speak of SIP telephone trunks [23], [24]. The development of the algorithm for redundancy is performed in an Asterisk environment. The call is overflow to the second resource when the central resource is congested or when the central resource is out of service due to some failure, as shown in Fig. 11:
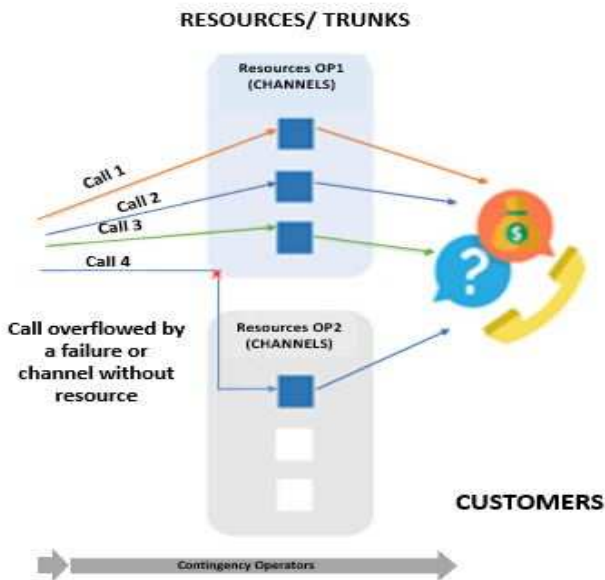

Fig. 11. Redundancy Diagram

In Figure 12 it can appreciate the detail of those above, an Asterisk dial plan script, where we execute the Dial application, which is responsible for making the call by the main trunk, additional assign the values to the variables sip

cause, hangup cause, and hangupcause_keys. These variables correspond to the response code sip, cause of hanging and stem sip respectively [25]:

Fig. 12. Asterisk DialPlan SIP Response

Table II shows a description of the SIP response codes:

TABLE II
SIP RESPONSE CODES

Sip Code	Hangup Cause	Description
480	20	Temporarily Unavailable
403	21	Forbidden
404	1, 2, 3, 26	Not Found
486	17	Busy Here
502	27	Bad Gateway
503	34, 38, 41, 42, 47	Service Unavailable

Finally, Fig. 13 shows how the call is taken through the contingent medium:

Fig. 13. Contingent Medium

IV. CONCLUSION

The project is underway and is being carried out by 40% in a company in the furniture sector that provides home furniture on credit. The call center has managed to improve contact ability with an increase of 5%, taking it from 31% to 36%, which has resulted in a substantial improvement in customer service, promoting business productivity in the same way. According to the contact ability indicator, the increase of five percent (5%) equals twenty-five thousand (25,000) additional active calls for the company, which match other vital indicators such as collection. In the schedule (and future work) come Machine Learning activities, which will provide an analysis of essential data, defining strategies that will facilitate corporate decision making. Additionally, the Asterisk PBX will be redefined, which will allow the business to be more competitive, obtaining higher profits and higher satisfaction of the system users and the company's customers.

Concerning the statistical component, several subsequent studies can be made, for example, multivariate analysis to find correlations between variables, and descriptive statistic to estimate the satisfaction level of call center users, employees, and clients, which will be exposed in future work.

REFERENCES

[1]   A. M. Gonzalez Vanegas, "Análisis de las strategies usadas por los servicios de tercerización de procesos de negocios (call center) en Colombia de frente a la fidelización de los clientes," Bogotá, 2014.

[2] M. C. Henao Robayo, M. J. Quiñonez Sánchez, and S. A. Cáceres Pinzón, "Estrategias De Tercerización En Colombia Como Centro De Operaciones Enfocado a Los Call Centers En Barranquilla," Universidad del Rosario, 2013.

[3] S. Ben-David and S. Shalev-Shwartz, Understanding Machine Learning: From Theory to Algorithms. 2014.

[4] I. Cheng, N. Boyette, and V. Krishna, "Towards a low-cost, high-quality service call architecture," in Proceedings - 2006 IEEE International Conference on Services Computing, SCC 2006, 2006.

[5] B. Legros, O. Jouini, and Y. Dallery, "A flexible architecture for call centers with skill-based routing," Int. J. Prod. Econ., 2015.

[6] R. Hernandez Sampieri, C. Fernandez Collado, and M. del P. Baptista Lucio, Metodología de la investigación. 2010.

[7] M. A. Saravia Gallardo, "Metodología de investigación científica," Conacyt, pp. 1–18, 2006.

[8] I. Goodfellow, Y. Bengio, and A. Courville, "Machine Learning Basics," Intell. Sens. Networks Integr. Sens. Networks, Signal Process. Mach. Learn., 2012.

[9] E. Alpaydın, Introduction to machine learning. 2014.

[10] I. H. Witten, E. Frank, and M. a Hall, Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). San Francisco: Morgan Kaufmann, 2011.

[11] V. Iversen, Teletraffic Engineering and Network Planning. 2005.

[12] I. J. Joskowicz, "Teoría e Ingeniería de Teletráfico Introducción Teoría e Ingeniería de Teletráfico." Montevideo, p. 131, 2017.

[13] B. Schwarz, "Asterisk open-source PBX system," 2004.

[14] S. Senthil Kumar, B. S. Dhivyalekshmi, S. Preethi, and R. Perumalraja, "PBX implementation in LAN using asterisk open source software," Int. J. Appl. Eng. Res., 2015.

[15] M. A. Khan and K. M. Shahriar, "ASTERISK Based Open Source IP-PBX System for Accountable Customer Support Service," in Proceedings - 2015 3rd International Symposium on Computational and Business Intelligence, ISCBI 2015, 2016.

[16] M. Maar, J. Sitarova, and M. Orgon, "Enterprise network with software Asterisk PBX based on the PLC technology," Int. J. Adv. Telecommun. Electrotech. Signals Syst., 2017.

[17] V. V Maksimov, N. S. Panasyuk, and S. A. Chmyhun, "Development of interactive voice response models based on IP-PBX Asterisk," in 2013 23rd International Crimean Conference "Microwave & Telecommunication Technology," 2013.

[18] J. B. Plaza, "Implantación de un sistema VoIP basado en Asterisk Director : José María San Agérico," Universidad Politécnica de Valencia, 2009.

[19] K. Walpole, Ronald; Myers, Raymond; Myers, Sharon; Ye, Probabilidad y Estadística para Ingenierías y Ciencias, Novena Edi. Mexico: Pearson, 2012.

[20] M. Mandjes and A. Ridder, "Large deviations approach to the transient of the Erlang loss model," Perform. Eval., 2001.

[21] M. D. Logothetis and I. D. Moscholios, "Teletraffic models beyond Erlang," in 10th International Conference, ELEKTRO 2014 - Proceedings, 2014.

[22] S. Zhang, D. Yin, Y. Zhang, and W. Zhou, "Computing on Base Station Behavior Using Erlang Measurement and Call Detail Record," IEEE Trans. Emerg. Top. Comput., 2015.

[23] T. Renier, H.-P. Schwefel, M. Bozinovski, K. Larsen, R. Prasad, and R. Seidl, "Distributed redundancy or cluster solution? An experimental evaluation of two approaches for dependable mobile internet services," in Lecture Notes in Computer Science, 2005.

[24] K. S. Chava and J. Ilow, "Integration of open source and enterprise IP PBXs," in Proceedings of the 3rd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, TridentCom 2007, 2007.

[25] "Asterisk config extensions.conf - VoIP-Info." [Online]. Available: https://www.voip-info.org/asterisk-config-extensionsconf/. [Accessed: 29-May-2018].