ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 10, Issue, 10, pp.74117-74123, October, 2018

DOI: https://doi.org/10.24941/ijcr.32520.10.2018

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

ANALYSIS OF TELECOMMUNICATION COMPANIES' DATA MINING TECHNIQUES IN PHILIPPINES: A CASE STUDY OF GLOBE TELECOM

[1, *]Ashraf Mahfoudh Abdannabi Ali and [2]Abdualla Mohammed Alshibani

[1]Department of Computer Technology Department Tripoli Higher Institute of Science and Technology, Libya
[2]Faculty of Science, Baniwaleed University, Libya

ARTICLE INFO

Article History:
Received 20th July, 2018
Received in revised form
09th August, 2018
Accepted 24th September, 2018
Published online 30th October, 2018

Key Words:
Telecommunication, Data Mining.

ABSTRACT

This study was intended to analyze data mining techniques of telecommunication companies in the Philippines. This study was guided by the following objectives: to provide an overview on data mining; to examine the various data mining techniques of telecommunication companies in Philippines; to identify the challenges of data mining faced by telecommunication companies in Philippines. The study employed the descriptive and explanatory design; secondary means were applied in order to collect data. Primary and Secondary data sources were used and data was analyzed using the chi-square statistical tool at 5% level of significance which was presented in frequency tables and percentage. The study findings revealed that data mining significantly impacts on the performance of telecommunication industries.

Copyright © 2018, Ashraf Mahfoudh Abdannabi Ali and Abdualla Mohammed Alshibani. *This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Citation: *Ashraf Mahfoudh Abdannabi Ali and Abdualla Mohammed Alshibani*, 2018. "Analysis of telecommunication companies' data mining techniques in Philippines: A case study of globe telecom", *International Journal of Current Research*, 10, (10), 74117-74123.

INTRODUCTION

Telecommunications and information and communication technologies (ICTs) have been widely recognized as a driver of economic and social development, poverty reduction and wealth creation. Telecommunications/ICTs provide an opportunity for developing countries to facilitate trade and economic development in general, as well as business development and job creation, especially for poor and marginalized populations, including women, indigenous peoples and persons with disabilities. However; developing countries often face challenges of rapid pace of change of technologies and convergence, financial resources, lack of suitable technical experience in planning and deploying advanced technologies and networks. The telecommunications industry in the Philippines particularly the mobile service market is dominated by two (2) major players, a duopoly, namely SMART Communications and GLOBE Telecom.

SMART Communications is the leading wireless services providers in the Philippines having 54.2 million subscribers on its GSM network as of 2012. It is a wholly owned subsidiary of the Philippine Long Distance Telephone Co. (PLDT), the dominant telecommunications carrier. It operates a nationwide cellular network, a wireless broadband service, a satellite phone service, and mobile commerce services (SMART Communications Website, 2014).

Corresponding author: *Ashraf Mahfoudh Abdannabi Ali,
Department of Computer Technology Department Tripoli Higher Institute of Science and Technology, Libya.

Sun Cellular or Digitel Mobile Philippines, Inc. (DMPI), a wholly-owned subsidiary of Digitel used to be part of the telecommunications industry. In October 2011, PLDT acquired Digitel from JG Summit holdings (Sun Cellular Website, 2014). Globe Telecom on the other hand is a partnership of Ayala Corporation and Singapore Telecom (SingTel).

It is a major player in the telecommunications industry in the Philippines servicing at least 25 million subscribers as of 2012 (Rappler, 2014). These two big mobile service providers – SMART Communications and GLOBE Telecom - contribute a lot in the gross domestic product (GDP) of the Philippines. The telecommunication industry specifically the mobile service companies is a vital factor in the Philippine economy thus contributing a huge percentage to the Gross domestic product (GDP). In Evans' research (2014), it contributed 10% to the Philippines GDP. This leads to an important question of how these mobile service companies can provide service quality to their subscribers. Due to the enormous number of network messages generated, technicians cannot possibly handle every message. For this reason expert systems have been developed to automatically analyze these messages and take appropriate action, only involving a technician when a problem cannot be automatically resolved (Weiss, Ros and Singhal, 1998). This study is focused on Globe Telecom. Globe Telecom, commonly shortened as Globe or GT, is a major provider of telecommunications services in the Philippines. It operates one of the largest mobile, fixed-line, and broadband networks in the country. Globe in 2017 has about 65.8 million mobile subscribers, nearly 3.5 million broadband customers and 859 thousand landline subscribers. The company's principal shareholders are Ayala Corporation and Singapore Tele communications. It is listed on the Philippine Stock Exchange under the ticker symbol GLOBE and had a market capitalization of US$3.8 billion as of the end of June 2018. Globe's main competitors in the fixed-line telephone market are PLDT and Digitel. Bayan Tel used to be one of its competitors prior to its acquisition by Globe. In the mobile phone market, its main competitors are the Smart and Talk N Text brands of Smart Communications and Sun Cellular, a wholly owned subsidiary. In 2016, Globe introduced its Globe Lifestyle brand as a way to connect to its customers through fashion. Fraud is a serious problem for telecommunication companies, leading to billions of dollars in lost revenue each year. Fraud can be divided into two categories: subscription fraud and superimposition fraud. Subscription fraud occurs when a customer opens an account with the intention of never paying for the account charges. Superimposition fraud involves a legitimate account with some legitimate activity, but also includes some "superimposed" illegitimate activity by a person other than the account holder.

Superimposition fraud poses a bigger problem for the telecommunications industry and for this reason data mining technique is used for identifying this type of fraud. These applications should ideally operate in real-time using the call detail records and, once fraud is detected or suspected, should trigger some action. This action may be to immediately block the call and/or deactivate the account, or may involve opening an investigation, which will result in a call to the customer to verify the legitimacy of the account activity. However, this journal will examine various data mining techniques of telecommunication companies in Philippines, provide an overview of data mining, examine the various data mining techniques of telecommunication companies in Philippines and identify the challenges of data mining faced by telecommunication companies in Philippines. The objectives of this study is to provide an overview on data mining; to examine the various data mining techniques of telecommunication companies in Philippine; to identify the challenges of data mining faced by telecommunication companies in Philippine.

Literature Review

Philippines is located in Southeast Asia, a Philippines is a democratic nation of 75 million people. An archipelago of over 7,100 islands, the country is divided into 14 geographic and three administrative regions. Despite the dispersed nature of its topography, political and economic activity is centralized in the National Capital Region (NCR), where Manila, the capital is located. Though only 14% of Filipinos live in this region, it accounts for nearly a third of the Philippine Gross Domestic Product. For 20 years until 1986, the Philippines was under President Ferdinand Marcos, who centralized government under his authoritative rule. While several high-profile infrastructure projects were undertaken during his extended term, many other utilities were poorly developed under nationalized and/or monopolized agencies. It was only in 1987, upon the assumption of President Corazon Aquino, that the reform and the deregulation of these agencies and the industries to which they belong started. Telecommunications was among those that lagged behind, even if it were crucial in virtually connecting the country's many islands in the absence of a strong public works network.

In 1928, an exclusive franchise to develop the country's telecom backbone was granted to the Philippine Long Distance Telephone Company. Though this meant the installation of 30,000 new lines, it eventually translated to a decline in the long-term telephony in the country. Until 1987, teledensity in the country remained at 1:100. Fortunately, reforms in the industry have so far led to significant growth in the number of telephone lines in the country though actual subscription to these lines remain low because of personal cost considerations rations and the country's over-all weak economy following the Asian crisis. But the centerpiece of the country's telecommunications industry is the significant growth in the subscription to wireless services. Though introduced only in 1989, cellular mobile telephone services now have 11 million subscribers, or almost double the year 2000 figures. The growth is driven by the services only originally available in mobile phones (text messaging, pre-paid payment schemes) and the poor landline infrastructure in the country (compared to the presence of cell sites across the archipelago). That the communication sector is crucial in the country's development is reflected in the government's policy statements. Conversely, policies on the communication sector are important as they "inevitably influence the flow of information, and thus the trade of goods and services. More than ever before, telecom policy can affect the location of jobs and the competitive position of firms" (Fetekuty, 1992).

The 1987 Philippine Constitution states, "The State recognizes the vital role of communication and information in nation-building."The telecommunications industry generates and stores a tremendous amount of data (Han *et al.,* 2002). These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers (Roset *et al.,* 1999). The amount of data is so great that manual analysis of the data is difficult, if not impossible. The need to handle such large volumes of data led to the development of knowledge-based expert systems. These automated systems performed important functions such as identifying fraudulent phone calls and identifying network faults. The problem with this approach is that it is time consuming to obtain the knowledge from human experts (the "knowledge acquisition bottleneck") and, in many cases, the experts do not have the requisite knowledge. The advent of data mining technology promised solutions to these problems and for this reason the telecommunications industry was an early adopter of data mining technology (Roset *et al.,* 1999).

Telecommunication data pose several interesting issues for data mining. The first concerns scale, since telecommunication databases may contain billions of records and are amongst the largest in the world. A second issue is that the raw data is often not suitable for data mining. For example, both call detail and network data are time-series data that represent individual events. Before this data can be effectively mined, useful "summary" features must be identified and then the data must be summarized using these features. Because many data mining applications in the telecommunications industry involve predicting very rare events, such as the failure of a network element or an instance of telephone fraud, rarity is another issue that must be dealt with. The fourth and final data mining issue concerns real-time performance because many data mining applications, such as fraud detection, require that any learned model/rules be applied in real-time (Ezawa and Norton, 1995). Several techniques has also been applied is tackling all these issues in telecommunication companies. Telecommunication networks are extremely complex configurations of equipment, comprised of thousands of interconnected components. Each network element is capable of generating error and status messages, which leads to a tremendous amount of network data. This data must be stored and analyzed in order to support network management functions, such as fault isolation. This data will minimally include a timestamp, a string that uniquely identifies the hardware or software component generating the message and a code that explains why the message is being generated. For example, such a message might indicate that "controller 7 experienced a loss of power for 30 seconds starting at 10:03 pm on Monday, May 12." The term, Data mining is very generic and it refers to mining data to discover knowledge (information). In literature, it is defined as a process of extraction and analysis of patterns, relationships and useful information from massive databases. This mining process is also called as Knowledge Discovery in databases (KDD). In any data mining process, there are four subtasks involved. They are: classification, clustering, regression and association rule learning [Gary Cokins and Ken King].

Moreover, depending on the domain of application, data mining techniques are divided into two major categories: i). Verification oriented (the system verifies the user's hypothesis) and ii). Discovery oriented-the system finds new rules and patterns autonomously (Hangxia, Min and Jianxia, 2009). Verification Methods deal with evaluation of a hypothesis proposed by an external source. Statistical methods such as goodness-of-fit test, t-test of means and analysis of variance comes under this category. These methods are less associated with data mining techniques than their discovery oriented counter parts because most data mining problems are concerned with selecting a hypothesis (out of a set of hypotheses) rather than testing a known one (MO-Zan et al., 2007). But discovery methods are used to identify patterns in data automatically. Data mining techniques are applied in telecom data base for various purposes. Each uses different type of telecom data depending on the purpose. The first step in the data mining process is to understand the data. Without such an understanding, useful applications cannot be developed. In this section we describe the three main types of telecommunication data. The data generated by telecom industries are broadly grouped into 3 types:

Call Detail Data: Every time a call is placed on a telecommunications network, descriptive information about the call is saved as a call detail record. The number of call detail records that are generated and stored is huge. For example, AT and Tlong distance customers alone generate over 300 million call detail records per day (Cortes and Pregibon, 2001). Given that several months of call detail data is typically kept online, this means that tens of billions of call detail records will need to be stored at any time. Call detail records include sufficient information to describe the important characteristics of each call. At a minimum, each call detail record will include the originating and terminating phone numbers, the date and time of the call and the duration of the call. Call detail records are generated in real time and therefore will be available almost immediately for data mining. This can be contrasted with billing data, which is typically made available only once per month.

Call detail records are not used directly for data mining, since the goal of data mining applications is to extract knowledge at the customer level, not at the level of individual phone calls. Thus, the call detail records associated with a customer must be summarized into a single record that describes the customer's calling behavior. The choice of summary variables (i.e., features) is critical in order to obtain a useful description of the customer. Below is a list of features that one might use when generating a summary description of a customer based on the calls they originate and receive over some time period:

- average call duration
- % no-answer calls
- % calls to/from a different area code
- % of weekday calls (Monday – Friday)
- % of daytime calls (9am – 5pm)
- average # calls received per day
- average # calls originated per day
- # unique area codes called during P

These eight features can be used to build a customer profile. Such a profile has many potential applications. For example, it could be used to distinguish between business and residential customers based on the percentage of weekday and daytime calls. Most of the eight features listed above were generated in a straightforward manner from the underlying data, but some features, such as the eighth feature, required a little more thought and creativity. Because most people call only a few area codes over a reasonably short period of time (e.g., a month), this feature can help identify telemarketers, or telemarketing behavior, since telemarketers will call many different area codes.

The above example demonstrates that generating useful features, including summary features, is a critical step within the data mining process. Should poor features be generated, data mining will not be successful. Although the construction of these features may be guided by common sense and expert knowledge, it should include exploratory data analysis. For example, the use of the time period 9am-5pm in the fifth feature is based on the commonsense knowledge that the typical workday is 9 to 5 (and hence this feature may be useful in distinguishing between business and residential calling patterns).

Network Data: Telecommunication networks are extremely complex configurations of equipment, comprised of thousands of interconnected components. Each network element is capable of generating error and status messages, which leads to a tremendous amount of network data. This data must be stored and analyzed in order to support network management functions, such as fault isolation. This data will minimally include a timestamp, a string that uniquely identifies the hardware or software component generating the message and a code that explains why the message is being generated. For example, such a message might indicate that "controller 7 experienced a loss of power for 30 seconds starting at 10:03 pm on Monday, May 12."Due to the enormous number of network messages generated, technicians cannot possibly handle every message. For this reason expert systems have been developed to automatically analyze these messages and take appropriate action, only involving a technician when a problem cannot be automatically resolved (Weiss, Ros and Singhal, 1998).

As was the case with the call detail data, network data is also generated in real-time as a data stream and must often be summarized in order to be useful for data mining. This is sometimes accomplished by applying a time window to the data. For example, such a summary might indicate that a hardware component experienced twelve instances of a power fluctuation in a 10-minute period.

Customer Data

Telecommunication companies, like other large businesses, may have millions of customers. By necessity this means maintaining a database of information on these customers. This information will include name and address information and may include other information such as service plan and contract information, credit score, family income and payment history. This information may be supplemented with data from external sources, such as from credit reporting agencies. The customer data maintained by telecommunication companies does not substantially differ from that maintained in most other industries. However, customer data is often used in conjunction with other data in order to improve results. For example, customer data is typically used to supplement call detail data when trying to identify phone fraud.

Data Mining Applications

The two main factors on which Data Mining relay on include the availability of the problem that has to be approached and solved by the Data Mining and the availability of Data for implementing the technologies. The telecommunications industry was an early adopter of data mining technology and therefore many data mining applications exist. The main reason behind the significance of Data Mining in the Telecommunications industries is the availability of tremendously large volume of data. Several typical applications are described below. The primary application areas include Fraud detection, marketing and Customer Relationship Management and Network Management.

Fraud Detection: Fraud is very serious issue that the telecommunication industry faces since it leads to the loss of revenue by billions of dollars. As provided by Gosset and Hyland (1999) the telecommunication fraud can be defined as any activity by which telecommunication service is obtained without intention of paying. Fraud in telecommunication companies can lead to billions of dollars in lost revenue each year. Telecommunication Fraud can be divided into two categories namely: subscription fraud and superimposition fraud.

Subscription fraud: It occurs when a customer opens an account with the intention of never paying for the account charges.

Superimposition fraud: It involves a legitimate account with some legitimate activity, but also includes some "superimposed" illegitimate activity by a person other than the account holder. Telecommunication companies consider Superimposition frauds are the most significant problems which occurs when a perpetrator gains illegal access to the account of a legitimate customer. Both subscriptions fraud and Superimposition fraud should be detected immediately and customer account should be deactivated. Superimposition fraud poses a bigger problem for the telecommunications industry and for this reason we focus on applications for identifying this type of fraud. These applications should ideally operate in real-time using the call detail records and, once fraud is detected or suspected, should trigger some action. This action may be to immediately block the call and/or deactivate the account, or may involve opening an investigation, which will result in a call to the customer to verify the legitimacy of the account activity. The most common method for identifying fraud is to build a profile of customer's calling behavior and compare recent activity against this behavior. Thus, this data mining application relies on deviation detection. The calling behavior is captured by summarizing the call detail records for a customer, as described earlier in this chapter. If the call detail summaries are updated in real-time, fraud can be identified soon after it occurs. Because new behavior does not necessarily imply fraud, one fraud-detection system augments this basic approach by comparing the new calling behavior to profiles of generic fraud and only signals fraud if the behavior matches one of these profiles (Cortes and Pregibon, 2001).

Customer level data can also aid in identifying fraud. For example, one sample rule that combines call detail and customer level data for detecting cellular fraud is: "People who have a price plan that makes international calls expensive and who display a sharp rise in international calls are likely the victim of cloning fraud." This same basic approach has been used to identify cellular cloning fraud, which occurs when the identification information associated with one cell phone is monitored and then programmed into a second phone (cloning fraud was a very serious problem in the 1990's, until authentication methods were developed to eliminate this type of fraud). This data mining application analyzed large amounts of cellular call data in order to identify patterns of fraud (Fawcett and Provost, 1997). These patterns were then used to generate monitors, each of which watches a customer's behavior with respect to one pattern of fraud. These monitors were then fed into a neural network, which determined when there is sufficiently evidence of fraud to raise an alert. Data mining can also help detect fraud by identifying and storing those phone numbers called when a phone is known to be used fraudulently. If many calls originate from another phone to numbers on this list of "suspect" phone numbers, one may infer that the account is being use fraudulently (Cortes and Pregibon, 2001).

Cellular cloning was a very serious issue in 1990's. This was eliminated with the Authentication methods. Deviation detection and Anomaly detection are the most common techniques used for detecting superimposed fraud. Combined use of customer signatures dynamic clustering and pattern recognition are some other methods which are recently applied in this area. Absolute analysis and differential analysis are considered as the two main sub categories of approaches for fraud detection. According to , the most often used techniques for fraud detection in telecommunication include statistical modeling, Bayesian rules, visualization methods, clustering, rule discovery, neural network, Markov models as well as combinations of more than one method. Fraud applications have some characteristics that require modifications to standard data mining techniques. For example, the performance of a fraud detection system should be computed at the customer level, not at the individual call level. So, if a customer account generates 20 fraud alerts, this should count, when computing the accuracy of this system, as only one alert; otherwise the system may appear to perform better than it actually does (Rosset, Murad, Neumann, Idan and Pinkas, 1999). More sophisticated cost-based metrics can also be used to evaluate the system. This is important because misclassification costs for fraud are generally unequal and often highly skewed (Ezawa and Norton, 1995). For this reason, when building a classifier to identify fraud, one should ideally know the relative cost of letting a fraudulent call go through versus the cost of blocking a call from a legitimate customer. Another issue is that since fraud is relatively rare and the number of verified fraudulent calls is relatively low, the fraud application involves predicting a relatively rare event where the underlying class distribution is highly skewed. Data mining algorithms often have great difficulty dealing with highly skewed class distributions and predicting rare events. For example, if fraud makes up only .2% of all calls, many data mining systems will not generate any rules for finding fraud, since a default rule, which never predicts fraud, would be 98.8% accurate. To deal with this issue, the training data is often selected to increase the proportion of fraudulent cases. For example, Ezawa and Norton (1995) increase the percentage of fraudulent calls from 1-2% to 9-12%. However, the use of a non-representative training set can be problematic because it does not provide the data mining method with accurate information about the true class distribution (Weiss and Provost, 2003).

Marketing/Customer Profiling: Telecommunication companies maintain a huge volume of data about their customers, in addition to the general customer data most businesses collect, telecommunication companies also store call detail records, which precisely describe the calling behavior of each customer. This information can be used to profile the customers and these profiles can be used for marketing and forecasting purposes. The emphasis of marketing application in telecommunication industry has moved from identifying new customers to measuring customer value and then taking steps to return the profitable customers. This shift has happened because it is expensive to acquire new customers than retaining the existing ones. A numerous Data Mining methods can be used to generate the customer life time value (the total net income a company can expect from a customer over time) for telecommunication customers. Different Data Mining techniques are used to model customer life time value for telecommunication customers.

The key element of modeling the life time value for a telecommunication customer is to estimate how long he or she will remain with their current network. One of the serious issues that telecommunications companies face is customer churn. It is a process by which a customer leaves one telecommunication company for another. Customer churn is a significant problem because of the associated loss of revenue and the high cost of attracting new customers. Some of the worst cases of customer churn occurred several years ago when competing long distance companies offered special incentives, typically $50 or $100, for signing up with their company-a practice which led to customers repeatedly switching carriers in order to earn the incentives. Data mining techniques now permit companies the ability to mine historical data in order to predict when a customer is likely to leave. These techniques typically utilize billing data, call detail data, subscription information (calling plan, features, contract expiration data) and customer information (e.g., age). Based on the induced model, the company can then take action, if desired.

For example, a wireless company might offer a customer a free phone for extending their contract. One such effort utilized a neural network to estimate the probability h(t) of cancellation at a given time t in the future (Mani, Drew, Betz and Datta, 1999). In the telecommunications industry, it is often useful to profile customers based on their patterns of phone usage, which can be extracted from the call detail data. These customer profiles can then be used for marketing purposes, or to better understand the customer, which in turn may lead to better forecasting models. In order to effectively mine the call detail data, it must be summarized to the customer level as described earlier in this chapter. Then, a classifier induction program can be applied to a set of labeled training examples in order to build a classifier. This approach has been used to identify fax lines (Kaplan, Strauss and Szegedy, 1999) and to classify a phone line as belonging to a business or residence (Cortes and Pregibon, 1998). Other applications have used this approach to identify phone lines belonging to telemarketers and to classify a phone line as being used for voice, data, or fax.

Network Fault Isolation: Telecommunication networks are comprised of highly complex configurations of hardware and software. Since the companies requires optimum network efficiency and reliability, most of the network elements have the capability of self-diagnosis and generating status and alarm messages. Expert systems were developed to handle alarms. In order to effectively manage the network, alarms must be analyzed automatically in order to identify network faults in a timely manner or before they occur and degrade network performance. A proactive response is essential to maintaining the reliability of the network. Because of the volume of the data, and because a single fault may cause many different, seemingly unrelated, alarms to be generated, the task of network fault isolation is quite difficult. Data mining has a role to play in generating rules for identifying faults. The Telecommunication Alarm Sequence Analyzer (TASA) is one tool that helps with the knowledge acquisition task for alarm correlation (Klemettinen, Mannila and Toivonen, 1999). This tool automatically discovers recurrent patterns of alarms within the network data along with their statistical properties, using a specialized data mining algorithm. Network specialists then use this information to construct a rule-based alarm correlation system, which can then be used in real-time to identify faults.

TASA is capable of finding episodic rules that depend on temporal relationships between the alarms. For example, it may discover the following rule: “if alarms of type link alarm and link failure occur within 5 seconds, then an alarm of type high fault rate occurs within 60 seconds with probability 0.7.” This information can be used to generate a rule based alarm correlation system, which can be used for identifying faults in real time. Before standard classification tasks can be applied to the problem of network fault isolation, the underlying time-series data must be represented as a set of classified examples. This summarization, or aggregation, process typically involves using a fixed time window and characterizing the behavior over this window. For example, if n unique alarms are possible, one could describe the behavior of a device over this time window using a scalar of length n. In this case each field in the scalar would contain a count of the number of times a specific alarm occurs. One may then label the constructed example based on whether a fault occurs within some other time frame, for example, within the following 5 minutes. Thus, two time windows are required.

Once this encoding is complete, standard classification tools can be used to generate "rules" to predict future failures. Such an encoding scheme was used to identify chronic circuit problems (Sasisekharan, Seshadri and Weiss, 1996). The problem of reformulating time-series network events so that conventional classification based data mining tools can be used to identify network faults has been studied. Weiss and Hirsh (1998) view this task as an event prediction problem while Fawcett and Provost (1999) view it as an activity monitoring problem. Transforming the time-series data so that standard classification tools can be used has several drawbacks. Time weaver (Weiss and Hirsh, 1998) is a genetic-algorithm based data mining system that is capable of operating directly on the raw network-level time series data (as well as other time-series data), thereby making it unnecessary to re-represent the network level data. Given a sequence of time stamped events and a target event T, Time weaver will identify patterns that successfully predict T. Time weaver essentially searches through the space of possible patterns, which includes sequence and temporal relationships, to find predictive patterns. The system is especially designed to perform well when the target event is rare, which is critical since most network failures are rare. In the case studied, the target event is the failure of components in the 4ESS switching system.

METHODOLOGY

The research design used for this study was the descriptive research design. Since data characteristics were described using frequencies and percentages, and no manipulations of data or variables were necessary, the researcher chose this research design. The researcher discarded other alternatives such as the causal and explanatory research designs, because accurate findings and data analysis may not be achieved.

DISCUSSION OF RESULTS AND CONCLUSION

This paper explained how data mining is used in the telecommunications company. Data Mining plays a significant role in the telecommunication industry due to the availability of large volume of data and the rigorous competition in the sector. Three main sources of telecommunication data (call detail, network and customer data) were described, as were common data mining applications (fraud, marketing and network fault isolation). The paper also highlighted some key issues affecting the ability to mine data, and commented on how they impact the data mining process. One central issue is that telecommunication data is often not in a form or at a level suitable for data mining. Other data mining issues that were discussed include the large scale of telecommunication data sets, the need to identify very rare events (e.g., fraud and equipment failures) and the need to operate in real-time (e.g., fraud detection). Data mining applications must always consider privacy issues. This is true in the telecommunications industry, since telecommunication companies maintain highly private information, such as whom each customer calls. Most telecommunication companies utilize this information conscientiously and consequently privacy concerns have thus far been minimized. A more significant issue in the telecommunications industry relates to specific legal restrictions on how data may be used. In the United States, the information that a telecommunications company acquires about their subscribers is referred to as Customer Proprietary Network Information (CPNI) and there are specific restrictions on how this data may be used. The Telecommunications Act of 1996, along with more recent clarifications from the Federal Communications Commission, generally prohibits the use of that information without customer permission, even for the purpose of marketing the customers other services. In the case of customers who switch to other service providers, the original service provider is prohibited from using the information to try to get the customer back (e.g., by only targeting profitable customers). Furthermore, companies are prohibited from using data from one type of service (e.g., wireless) in order to sell another service (e.g.. landline services). Thus, the use of data mining is restricted in that there are many instances in which useful knowledge extracted by the data mining process cannot be legally exploited. Much of the rationale for these prohibitions relates to competition. For example, if a large company can leverage the data associated with one service to sell another service, then companies that provide fewer services would be at a competitive disadvantage. The telecommunications industry has been one of the earliest adopters of data mining technology, largely because of the amount and quality of the data that it collects. This has resulted in many successful data mining applications. Given the fierce competition in the telecommunications industry, one can only expect the use of data mining to accelerate, as companies strive to operate more efficiently and gain a competitive advantage. The recent developments in the Data Mining and the implementation and enhancement of existing techniques and methods ensure the continuous growth and compatibility of telecommunication companies that make use of them.

REFERENCES

Aggarwal, C. 2007. Data Streams: Models and Algorithms New York: Springer.

Cortes, C., Pregibon, D. 1998. Giga-mining: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining; 174-178, 1998 August 27-31; New York, NY: AAAI Press.

Cortes, C., Pregibon, D. 2001. Signature-based methods for data streams. Data Mining and Knowledge Discovery; 5(3):167-182.

Ezawa, K., Norton, S. 1995. Knowledge discovery in telecommunication services data using Bayesian network models. Proceedings of the First International Conference on Knowledge Discovery and Data Mining; Montreal Canada. AAAI Press: Menlo Park, CA.

Fawcett, T, Provost, F. 1999. Activity monitoring: Noticing interesting changes in behavior. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 53-62. San Diego. ACM Press: New York, NY.

Fawcett, T., Provost, F. 1997. Adaptive fraud detection. Data Mining and Knowledge Discovery; 1(3):291-316.

Fetekuty, G. 1992. "Negotiating the World Information Economy." In Sapolsky, H. (Ed). The Telecom Revolution: Past, Present and Future

Gary, C. and Ken. K. 2011. Managing Customer Profitability and Economic Value in the Telecommunication Industry: SAS Institute White paper.

Han, J., Altman, R. B., Kumar, V., Manila, H., Pregibon, D. 2002. Emerging scientific applications in data mining. Communications of the ACM; 45(8): 54-58.

Hangxia Ma, Min Qin, Jianxia Wang, 2009. "Analysis of the Business Customer Churn Based on Decision Tree

Method", The Ninth International Conference on Control and Automation, Guangzhou, China.

Kaplan, H., Strauss, M., Szegedy, M. 1999. Just the fax differentiating voice and fax phone lines using call billing data. Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms. 935-936. Baltimore, Maryland. Society for Industrial and Applied Mathematics: Philadelphia, PA.

Klemettinen, M., Manila, H., Toivonen, H. 1999. Rule discovery in telecommunication alarm data. *Journal of Network and Systems Management,* 7(4):395-423.

Liebowitz, J. 1988. Expert System Applications to Telecommunications. New York, NY: John Wiley.

Berry M. and G. Linoff, 2000. Mastering Data Mining. John Wiley and Sons, New York, USA.

Mani, D. R., Drew, J., Betz, A., Datta, P. 1999. Statistics and data mining techniques for lifetime value modeling. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 94-103. San Diego. ACM Press: New York, NY.

MO Zan, ZHOA Shan, LI Li, LIU Ai-Jun, 2007. "A predictive Model of Churn in Telecommunications Base on Data Mining". IEEE International Conference on Control and Automation", Guangzhou, China.

Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., and Kaushansky, H. 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunication industry

PAKDD (2006). Data Mining Competition. http://www3. ntu. edu. sg/SCE/pakdd2006/competition/overview. htm

Pareek, D. 1999. Business Intelligence for Tele communications. Auerbach Publications, Taylor and Francis Group LLC.

Roset, S., Murad, U., Neumann, E., Idan, Y., Pinkas, G. 1999. Discovery of fraud rules for telecommunications challenges and solutions. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 409-413, San Diego CA. New York: ACM Press.

Sasisekharan, R., Seshadri, V., Weiss, S. 1996. Data mining and forecasting in large-scale telecommunication networks. IEEE Expert. 11(1):37-43.

Weiss, G. 2004. Mining with rarity: A unifying framework SIGKDD Explorations.

Weiss, G. M., Hirsh, H. 1998. Learning to predict rare events in event sequences. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. 359-363. AAAI Press.

Weiss, G. M., Provost, F. 2003. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research,* 19:315- 354.

Weiss, G. M., Ros, J, Singhal, A. 1998. Network monitoring using object-oriented rule. Proceedings of the Tenth Conference on Innovative Applications of Artificial Intelligence; 1087-1093. AAAI Press, Menlo Park, CA.

Yossi Ritcher, Elad Yom-Tov, Noam Slonim, 1999. Predicting Customer Churn in Mobile Networks through Analysis of Social Groups. SIAM.

****