Applying Data Mining to Predict New Subscriber Activations in Telecom Companies

Kazi Shamsul Arefin, and Mohammad Majharul Islam

Abstract—As the great advance of new technologies, the competition in wireless telecommunication industry is getting severe. Therefore, Sales Management (SM) is one of the biggest concerns in a telecom sectors. Application of data mining can support the SM of telecom effectively, and a systematical architecture of data mining has been needed to support every aspects of SM roundly. Mobile companies wish to sale their packages and satisfy their needs. Hence, they need to predict the possible time to offer their packages and enhance revenue. To solve this problem, the systematical application structural design of data mining in telecom SM is established. In response to the difficulty of possible time for prediction of high sales, this study applies data mining techniques to build a model for prediction.

Index Terms—Data Mining, Sales Management (SM), New Subscriber Activations (NSA), Time Series (TS), Business Intelligence (BI), BIDS.

1 Introduction

TELECOMMUNICATIONS is one of the most data-intensive industries in the world. This is most likely because telecommunication companies routinely generate and store enormous amounts of high-quality data, have a very large customer base_and operate in a rapidly changing and highly competitive environment. Telecommunication companies maintain extensive customer information, such as billing information, as well as information obtained from outside parties, such as credit score information. This information can be quite useful and often is combined with telecommunication- specific data to improve the results of data mining. For example, while data of Call Detail Records (CDR) can be used to identify suspicious calling patterns, a customer's credit score is often incorporated into the analysis before determining the likelihood that fraud is actually taking place.

Data mining is used to analyze data within relational database. It is used to improve their marketing efforts, identify fraud, and better manage their telecommunication networks. However, these telecom companies face a number of data mining challenges due to the enormous size of their data sets, the sequential and temporal aspects of their data, and the need to predict very rare events—such as customer fraud and network failures—in real-time. Data mining can be used to extract knowledge directly from the data. It enables to view data from many different perspectives, categorize the data in new ways and summarize the result.

Some telecom company's managers have been reluctant to use data mining to their advantage. Arguments for this can run the gamut from too expensive to not enough time to lack of upper-management commitment. However, raising arguments like these will result, quite simply, in lost opportunities for many telecom companies. Conversely, if data mining is undertaken then many new opportunities will be created that will enable a company to become more profitable and competitive.

Data mining processes business activity in order to arrive at meaningful results, a carrier needs to align data mining with the goals and objectives of its business. Otherwise the results will be irrelevant and lacking context. Data mining also focuses on exploring different hypotheses, taking into account disparate data, where the end result of the data mining exercise could potentially drive building a better customer experience, creating new operational processes and improving the overall bottom line.

In this paper, we have introduced the TS data mining algorithm to accelerate the activation of new subscribers for mobile companies. We observed for 14 months (July 2008 to August 2009) the activation rate of mobile subscribers (Prepaid and Postpaid) and found some interesting results. Our vision is to provide to mobile companies a prediction of possible sales in upcoming days. Finally, an application example has been given to illuminate that telecom companies can make marketing strategies roundly and effectively with the support of data mining.

Kazi Shamsul Arefin is with the Department of Computer Science and Engineering, University of Asia Pacific (www.uap-bd.edu), Dhanmondi, Dhaka-1209, Bangladesh. E-mail: arefin@uap-bd.edu.

Mohammad Majharul Islam is with the Department of Computing and Information Technology, Islamic University of Technology (www.iut-dhaka.edu), Board Bazar, Gazipur -1704, Bangladesh. E-mail: majhar999@uap-bd.edu.

2 Background Information

2.1 NSA Status of Mobile Service Providers

In 1989, Bangladesh Telecom Limited (BTL) was awarded a license to operate cellular, paging, and other wireless communication networks [1]. Then in 1990 Hutchison

Bangladesh Telecom Limited (HBTL) was incorporated in Bangladesh as a joint venture between BTL and Hutchison Telecommunications (Bangladesh) Limited. HBTL began commercial operation in Dhaka using the AMPS mobile technology in 1993 and became the 1st cellular operator in South Asia [4]. Later that year Pacific Motors bought 50% of BTL. By 1996 HBTL was renamed as Pacific Bangladesh Telecom Limited (PBTL) and launched the brand name “Citycell Digital” to market its cellular products. Currently there are 6(six) mobile operators in Bangladesh. They are namely:

1. ***GrameenPhone –*** Grameenphone was the first Telenor venture in the Asian telecom market [3]. Today, Grameenphone is the largest mobile provider in Bangladesh, serving more than 21 million subscribers; an increase of 53 per cent in the subscriber base in 2007. Telenor and its partners have boosted network capacity and extended coverage to new and often remote areas, connecting millions of previously unconnected people. Telenor holds 62 per cent of Grameenphone. Grameenphone started trading its shares on the stock exchanges in Dhaka and Chittagong on November 16, 2009.
2. ***Banglalink –*** When banglalink entered the Bangladesh telecom industry in February 2005, the scenario changed overnight with mobile telephony becoming an extremely useful and affordable communication tool for people across all segments [6]. Within one year of operation, banglalink became the fastest growing mobile operator of the country with a growth rate of 257%. This milestone was achieved with innovative and attractive products and services targeting the different market segments; aggressive improvement of network quality and dedicated customer care; and effective communication that emotionally connected customers with banglalink.
3. ***Robi –***Axiata (Bangladesh) Limited is a dynamic and leading countrywide GSM communication solution provider [13]. It is a joint venture company between Axiata Group Berhad, Malaysia and NTT DOCOMO INC, Japan. Axiata (Bangladesh) Limited, formerly known as Telekom Malaysia International (Bangladesh), commenced its operation in 1997 under the brand name Aktel among the pioneer GSM mobile telecommunications service providers in Bangladesh. Later, on 28th March, 2010 the company started its new journey with the brand name Robi.
4. ***Warid –*** Warid Telecom International Limited, Bangladesh sold a 51 per cent stake in its two largest African networks to India’s Essar, and 70 per cent of its Bangladeshi network to Bharti Airtel [5]. In Bangladesh, Warid Telecom commenced its operations under a landmark MOU agreed upon by the Dhabi Group and the Government of Bangladesh worth USD 1 billion, out of which USD 750 million was exclusively committed for investment in the telecommunication sector of the country. Succeeding the MOU signing, the BTRC license for telecom service provision was issued to Warid Telecom, followed by the signing of interconnectivity agreement with all the existing telecom companies of Bangladesh. In May 10th, 2007, Warid Telecom launched its commercial operations in Bangladesh with a network encompassing 26 districts. By November 2007, the network had been expanded to cover 61 districts and being used by 2 million customers.
5. ***TeleTalk –*** Public limited company but 100% share have been owned by the government of Bangladesh [2].
6. ***Citycell –*** JointJoint venture with SingTel Asia pacific investment Pte limited [4]. According to BTRC (Bangladesh Telecommunication Regulatory Commission), the total number of Mobile Phone Active Subscribers has reached 52.43 million at the end of December 2009.

At the end of March 2010, the total number of Mobile Phone Subscribers reached 54.7 million [10]. The figures are shown below:

TABLE 1
STATUS OF SUBSCRIBERS OF MOBILE COMPANIES
(IN MARCH 2010)

Operators	*Active Subscribers (in millions)*
Grameen Phone Ltd. (GP)	23.9
Axiata (Bangladesh) Limited (Robi)	10.59
Orascom Telecom Bangladesh Limited (Banglalink)	14.22
PBTL (Citycell)	1.91
Teletalk Bangladesh Ltd. (Teletalk)	1.07
Warid Telecom International L.L.C (Warid)	3.01
Total	**54.7**

However, it was only **1028.19 thousand** PSTN Phone Subscribers [11]. The figures are shown below:

TABLE 2
STATUS OF SUBSCRIBERS OF PSTN COMPANIES
(IN MARCH 2010)

Operators	*Subscribers(in thousands)*
BTCL	872.41
Telebarta Ltd.	56.42
Jalalabad Telecom Ltd.	10.90
Onetel Communication Ltd.	39.57
Westec Ltd.	17.00
Sheba Phone Ltd. (ISL)	11.62
S. A. Telecom System Ltd.	18.03
Banglaphone Ltd.	2.24

Total	**1,028.19**

Undoubtedly, because of marketing policy and SM of mobile companies, this immense success was possible. SM of a telecom company attains the sales goals in an effective & efficient manner through planning, staffing, training, leading & controlling organizational resources. Revenue stimulates the telecom operators and the management of that process is the most important function.

2.2 Knowledge Management

The term of knowledge management is often problematic as there is little consensus regarding its definition [18], [19]. According to [23], a recent study outlined that there is not a single area of discourse within either the academic or popular management literature [21] that has been accepted. Many authors simply avoid the term completely, preferring to focus on specific aspects of the topic such as knowledge, innovation or learning [22]. Furthermore others argue that knowledge management is closely related to concepts such as organizational learning, organizational memory, information sharing, and collaborative work [20].

According to [23], there is no consensus on the definition of knowledge. Many authors avoided epistemological debate on the definition of knowledge by comparing knowledge with information and data [24]. A commonly held view is that data is raw numbers and facts, information is processed data and knowledge is authenticated information. Yet, as [24] highlighted, the presumption of hierarchy from data to information to knowledge with each varying along some dimension such as context, usefulness or interpretability is inaccurate.

[24] argued that the effective distinguishing feature between information and knowledge is not found in the content, structure, usefulness or interpretability, but rather "knowledge is information possessed in the minds of individuals: it is personalized information (which may or may not be new, unique, useful or accurate) related to facts, procedures, concepts, interpretations, ideas, observations, and judgments". Thus systems designed to support knowledge in organizations may not appear radically different from other forms of information systems, but will be geared toward enabling users to assign meaning to information and to capture some of their knowledge in information and / or data.

2.3 Business Intelligence Development Studio (BIDS)

According to [31], Business Intelligence Development Studio (BIDS) is one of the Reporting Services authoring environments that design reports and models. BIDS is the Visual Studio environment with enhancements that are specific to business intelligence solutions. BIDS provides solution files that enable to create and organize business intelligence project files.

BIDS is Microsoft Visual Studio 2008 with additional project types that are specific to SQL Server business intelligence [31]. BIDS is the primary environment that uses to develop business solutions that include Analysis Services, Integration Services, and Reporting Services projects. Each project type supplies templates for creating the objects required for business intelligence solutions, and provides a variety of designers, tools, and wizards to work with the objects.

3 DATA MINING

According to [17], data mining, is also called knowledge discovery in databases (KDD), is the process of extracting (unknown) patterns from data. In general, a data mining process includes several iterations of single data mining steps (algorithm executions). The goals of the data mining process are defined by the intended use of the system from the user perspective and can be classified into two types: verification, where the system is limited to verifying the user's hypothesis, and discovery, where the system autonomously finds new patterns. The discovery goal can be further subdivided into prediction, where the system finds patterns for predicting the future behavior of some entities, and description, where the system finds patterns for presentation to a user in a human-understandable form [7]. A variety of data-mining methods exist which help in achieving the goal of prediction and description.

3.1 Data Mining Methods and Algorithms

According to [8], data mining methods commonly involve the following classes of tasks: Inferring rudimentary rules, statistical modeling, constructing decision trees, constructing rules, mining association rules, linear models, instance-based learning, and clustering. For each of these methods a variety of data mining algorithms are available that are described in bellow:

3.2 The Data Mining Process

According to [17], the data mining process is an interactive and iterative process that involves numerous steps with many decisions made by the data miner. CRISP-DM [9] is a standard process model for data mining that depicts corresponding phases of a project, their respective tasks, and relationships between these tasks. According to CRISP-DM, the lifecycle of a data mining project consists of the following 6 (six) different phases:

1. ***Business Understanding*** - understanding the project objectives and requirements from a business perspective and converting this knowledge into a data mining problem definition;
2. ***Data Understanding*** - getting to know the data and to identify data quality problems;
3. ***Data Preparation*** – construct the final dataset from the initial raw data as input for the modeling;
4. ***Modeling*** - various modeling techniques are selected and applied, including the calibration of their specific settings;
5. ***Evaluation*** - assess how well the built model achieves the business objectives;
6. ***Deployment*** - the results of the data mining and the knowledge gained are delivered to the user, reaching from generating a simple report up to a

complex implementation of a repeatable data mining process.

The process in general is iterative, but also foresees stepping back between certain phases to adjust some of the decisions made. From the data mining perspective, the (business) user is mainly involved in the phases Business Understanding and Deployment, while the other phases are mostly performed only by the data miner. In terms of integration, it has to be distinguished between the (technical) deployment of the data mining solution as a whole, which might be done only once for a given business process, and the deployment of new data mining models, which might be done frequently.

3.3 Application of TS Data Mining

According to [15], [16], data mining is a technique of discovering useful pattern in data that are hidden and unknown in normal circumstances. Data mining consist of machine learning, statistics and database design [25]. It uses methods such as clustering, classification, association rule mining and probabilistic graphical dependency models to identify hidden and useful information from large databases [25], [26].

Data mining refers to extracting or "mining" knowledge from large amounts of data [27]. Keogh et al. [28] used piecewise linear approximation representation of TS for clustering, classification and association rule mining of TS data. They have examined the first extensive review and empirical comparison of TS segmentation algorithms from a data mining perspective. They have developed an efficient sequential pattern for identifying frequent temporal patterns. Faloutus et al. [29] have developed efficient TS similarity search methods namely subsequence matching. Povinelli et al. [30] have proposed a new signal analysis and classification method based on reconstructed phase space.

4 Microsoft TS Data Mining Algorithm

According to [14], the Microsoft TS algorithm provides regression algorithms that are optimized for the forecasting of continuous values, such as product sales, over time. Whereas other Microsoft algorithms, such as decision trees, require additional columns of new information as input to predict a trend, a TS model does not. A TS model can predict trends based only on the original dataset that is used to create the model. New data can be added to the model making a prediction and automatically incorporate the new data in the trend analysis.

The following diagram shows a typical model for forecasting sales of a product in four different sales regions over time. The model that is shown in the diagram shows sales for each region plotted as red, yellow, purple, and blue lines. The line for each region has two parts:

- Historical information appears to the left of the vertical line and represents the data that the algorithm uses to create the model.
- Predicted information appears to the right of the vertical line and represents the forecast that the model makes.

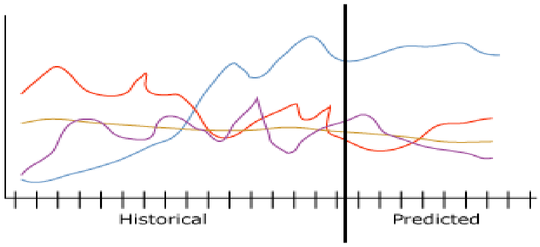


Fig. 1. Prediction from Historical Events.

The combination of the source data and the prediction data is called a series.

An important feature of the Microsoft TS algorithm is that it can perform cross prediction. If we train the algorithm with two separate, but related, series, we can use the resulting model to predict the outcome of one series based on the behavior of the other series. For example, the observed sales of one product can influence the forecasted sales of another product. Cross prediction is also useful for creating a general model that can be applied to multiple series. For example, the predictions for a particular region are unstable because the series lacks good quality data. We could train a general model on an average of all four regions, and then apply the model to the individual series to create more stable predictions for each region.

Example: The management team at Adventure Works Cycles wants to predict monthly bicycle sales for the coming year. The company is especially interested in whether the sale of one bike model can be used to predict the sale of another model. By using the Microsoft TS algorithm on historical data from the past three years, the company can produce a data mining model that forecasts future bike sales. Additionally, the company can perform cross predictions to see whether the sales trends of individual bike models are related.

Each quarter, the company plans to update the model with recent sales data and update their predictions to model recent trends. To correct for stores that do not accurately or consistently update sales data, they will create a general prediction model, and use that to create predictions for all regions.

4.1 How the Algorithm Works

In SQL Server 2005, the Microsoft TS algorithm used a single algorithm, ARTxp. The ARTXp algorithm was optimized for short-term predictions, and therefore, predicted the next likely value in a series. Beginning in SQL Server 2008, the Microsoft TS algorithm uses both the ARTxp algorithm and a second algorithm, ARIMA. The ARIMA algorithm is optimized for long-term prediction.

By default, the Microsoft TS algorithm uses a mix of the algorithms when it analyzes patterns and making predictions. The algorithm trains two separate models on the same data: one model uses the ARTxp algorithm and one model uses the ARIMA algorithm. The algorithm then blends the results of the two models to yield the best prediction over a variable number of time slices. Because ARTxp is best for short-term predictions, it is weighted more heavily at the beginning of a series of predictions. However, as the time slices that we are predicting move further into the future, ARIMA is weighted more heavily.

We can also control the mix of algorithms to favor either short- or long-term prediction in the times series. Beginning in SQL Server 2008 Standard, we can specify that the Microsoft TS algorithm use one of the following settings:

- Use only ARTxp for short-term prediction.
- Use only ARIMA for long-term prediction.
- Use the default blending of the two algorithms.

Beginning in SQL Server 2008 Enterprise, we can customize how the Microsoft TS algorithm blends the models for prediction. When we use a mixed model, the Microsoft TS algorithm blends the two algorithms in the following way:

- Only ARTxp is always used for making the first couple of predictions.
- After the first couple of predictions, a combination of ARIMA and ARTxp is used.
- As the number of prediction steps increases, predictions rely more heavily on ARIMA until ARTxp is no longer used.
- We control the mixing point, the rate at which the weight of ARTxp is decreased, and the weight of ARIMA is increased by setting the PREDICTION_SMOOTHING parameter.

Both algorithms can detect seasonality in data at multiple levels. For example, our data might contain monthly cycles nested within yearly cycles. To detect these seasonal cycles, we can either provide a periodicity hint or specify that the algorithm should automatically detect periodicity.

In addition to periodicity, there are several other parameters that control the behavior of the Microsoft TS algorithm when it detects periodicity, makes predictions, or analyzes cases. For information about how to set algorithm parameters, see Microsoft TS Algorithm Technical Reference.

4.2 Data Required for TS Models

When we prepare data for use in training any data mining model, we understand the requirements for the particular model and how the data is used.

Each forecasting model must contain a case series, which is the column that specifies the time slices or other series over which change occurs. For example, the data in the previous diagram shows the series for historical and forecasted bicycle sales over a period of several months. For this model, each region is a series, and the date column contains the TS, which is also the case series. In other models, the case series can be a text field or some identifier such as a customer ID or transaction ID. However, a TS model must always use a date, time, or some other unique numeric value for its case series.

The requirements for a TS model are as follows:

A single key time column: Each model must contain one numeric or date column that is used as the case series, which defines the time slices that the model will use. The data type for the key time column can be either a datetime data type or a numeric data type. However, the column must contain continuous values, and the values must be unique for each series. The case series for a TS model cannot be stored in two columns, such as a Year column and a Month column.

A predictable column: Each model must contain at least one predictable column around which the algorithm will build the TS model. The data type of the predictable column must have continuous values. For example, we can predict how numeric attributes, such as income, sales, or temperature, change over time. However, we cannot use a column that contains discrete values, such as purchasing status or level of education, as the predictable column.

An optional series key column: Each model can have an additional key column that contains unique values that identify a series. The optional series key column must contain unique values. For example, a single model can contain sales for many product models, as long as there is only one record for each product name for every time slice.

We can define input data for the Microsoft TS model in several different ways. However, because the format of the input cases affects the definition of the mining model, we need to consider our business needs and prepare our data accordingly. The following two examples illustrate how the input data affects the model. In below examples, the completed mining model contains patterns for four distinct series:

- Sales for Product A
- Sales for Product B
- Volume for Product A
- Volume for Product B

In the below example, we can predict new future sales and volume for each product. We cannot predict new values for product or for time.

TABLE 3

EXAMPLE FOR SALES PREDICTION

TimeID	Product	Sales	Volume
July-10	A	100	60
Aug-10	A	110	50
July-10	B	50	90
Aug-10	B	30	90

The TimeID column in the table contains a time identifier, and has two entries for each day. The TimeID column becomes the case series. Therefore, we would designate this column as the key time column for the TS model.

The Product column defines a product in the database. This column contains the product series. Therefore, we would designate this column as a second key for the TS model.

The Sales column describes the gross profits of the specified product for one day, and the Volume column describes the quantity of the specified product that remains in the warehouse. These two columns contain the data that is used to train the model.

4.3 Viewing a TS Model

After the model has been trained, the results are stored as a set of patterns, which can be explored or used to make predictions.

To explore the model, the TS Viewer can be used. The viewer includes a chart that displays future predictions, and a tree view of the periodic structures in the data.

The stored content for the model includes details such as the periodic structures detected by the ARIMA and ARTxp algorithms, the equation used to blend the algorithms, and other statistics.

4.4 Creating TS Predictions

We can create queries to return a variable number of predictions, and extra columns to the predictions to return descriptive statistics.

The Microsoft TS algorithm, in order to make predictions, needs to consider the following additional restrictions and requirements:

- Cross-prediction is only available when we use a mixed model, or a model based on the ARTxp algorithm. If we use a model based only on the ARIMA algorithm, cross-prediction is not possible.
- A TS model can make predictions that differ, sometimes significantly, depending on the 64-bit operating system that the server uses. These differences occur due to the way that an Itanium-based system represents and handles numbers for floating-point arithmetic, which differs from the way that an x64-based system does these calculations. Because prediction results can be specific to the operating system.

5 RESEARCH ANALYSIS

5.1 Experimental Evaluation

We used Microsoft TS algorithm of BIDS to predict from data. When we started the analysis, what we expected was turned out to be totally the opposite of reality. We thought, at the beginning of the month, the rate of NSA will be higher and it will be dropped gradually. However, in figure 2, we can see that the reality is completely different. NSA, in the figure, was almost constant from day 1-19th although some fluctuated and got the highest peak in 29-30th day. After then, there was a sudden drop of NSA. Here, we have considered August is the month of 30 days.

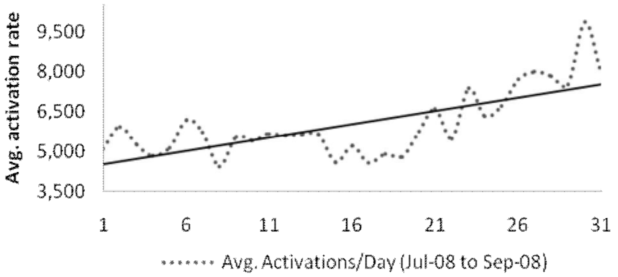
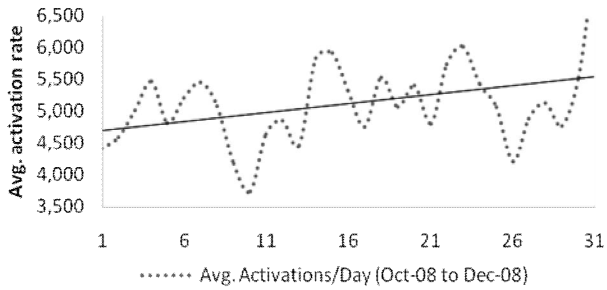


Fig. 2. Average NSA from July 2008 to September 2008.

In figure 3, the average rate of NSA fluctuated rapidly (October 2008 to December 2008) and the highest peak was at the end of month.

Fig. 3. Average NSA from October 2008 to December 2008.

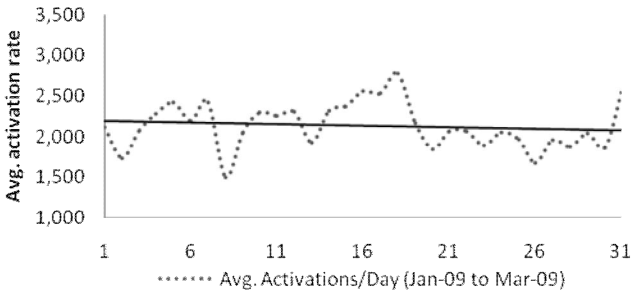
In figure 4, we got the lowest peak in the 8th day then it followed upward trend and claimed the maximum



point then declined again. However, there was a tendency to rise at the end of month.

Fig. 4. Average NSA from January 2009 to March 2009.

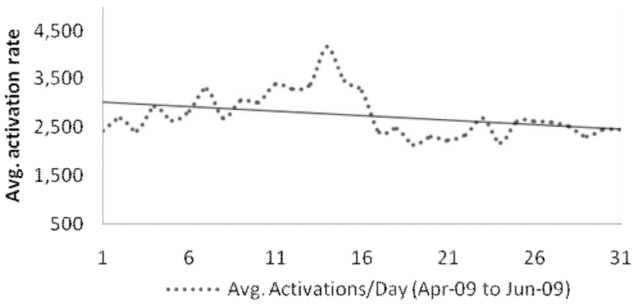
In figure 5, there was an increasing trend from day 1 to 14 and after then it dropped. This is because, in May 2009,



there was an offer to buy new packages which was finished in 14th May. Therefore we got the highest peak in the middle of month.

Fig. 5. Average NSA from April 2009 to June 2009.

In figure 6, there was an upward trend in NSA (July 2008 to June 2009) although some fluctuated and reached



to the peak at the end of month.

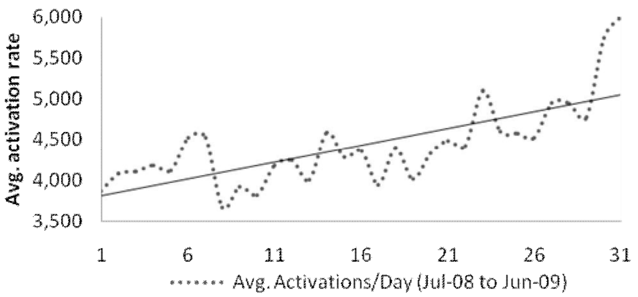


Fig. 6. Average New Activation from July 2008 to June 2009.

Now from above all observations, in July 2009, we were expecting that at the beginning of month will have the lowest peak and then it will follow an upward trend with some fluctuations and reach the maximum at the end of month then drop. In figure 7, we can see the prediction was moderately right.

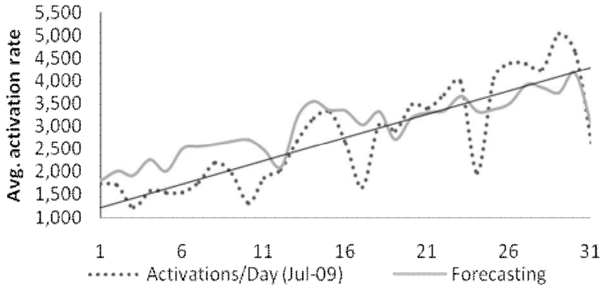


Fig. 7. NSA in July 2009.

In August 2009, prediction was almost same, though there was not any new package promotion offers. In figure 8, we again observed that the prediction through data mining was moderately right.

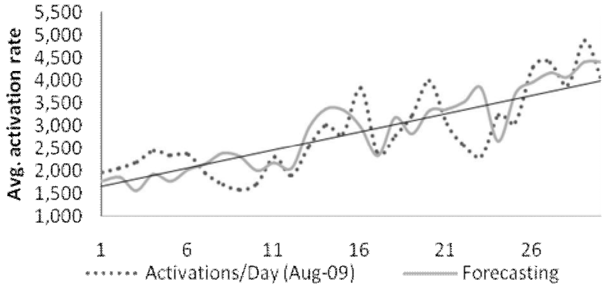


Fig. 8. NSA in August 2009.

In figure 9, at beginning of month the average rate of new subscriber actvations were arroud 2,900 and reached approximately 5,000 in 30th day in the month then it fell.

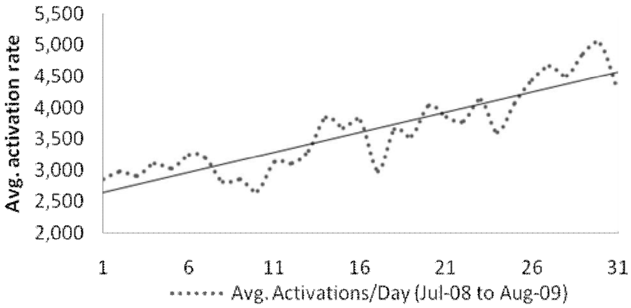


Fig. 9. Avg. NSA from July 2008 to August 2009.

5.2 System Configuration and Tools Used

System configuration was Core 2 Duo (2.13 MHz Processor), DDR3 4 GB RAM, and 4MB L2 Cache memory to retrive data.

Business Intelligence Development Studio (BIDS) of SQL Server 2008 was used to predict NSA from sample patterns.

6 Research Limitations

Data collection was extracted from data warehouse of a mobile service provider. It was the most challenging part of this research. Every telecom company tries to keep the data secret. Therefore, they only provided around 14 months of NSA data for the research. For this reason, some figures drastically fluctuated in the graphs. It may also be because of analyzing telecom market in Bangladesh.

7 Conclusion

The purpose of this research is to propose an estimation of possible sales to assist telecom Sales Management (SM). SM is always critical in broadly liberalized especially in telecom market. In order to keep the competitive advantage in telecom market, mobile service providers must be able to predict the rate of NSA. In this research paper, we utilized the Microsoft TS algorithm with SQL server BIDS to predict NSA and proved how the use of this model can predict the activation of new subscribers in upcoming days.

Acknowledgment

We would like to express our gratitude to Bangladesh Telecommunication Regulatory Commission (BTRC) because of their immense support.

References

[1] www.btrc.gov.bd (Last retrieval time: August 2010).
[2] www.teletalk.com.bd (Last retrieval time: August 2010).
[3] www.grameenphone.com (Last retrieval time: June 2010).
[4] www.citycell.com (Last retrieval time: May 2010).
[5] www.waridtel.com.bd (Last retrieval time: August 2010).

[6] www.banglalinkgsm.com (Last retrieval time: May 2010).

[7] Fayyad, U. et al., "From Data Mining to Knowledge Discovery in Databases," *AI Magazine 17*, (1996).

[8] Witten, I.H. et al., "Data Mining: Practical machine learning tools and techniques," *2nd Edition, Morgan Kaufmann*, San Francisco (2005).

[9] Hornick, M.F. et al., "Java Data Mining: Strategy, Standard, and Practice," *Morgan Kaufmann*, San Francisco (2006).

[10] www.btrc.gov.bd / newsandevents / mobile_phone_subscribers.php (Last retrieval time: August 2010).

[11] www.btrc.gov.bd / newsandevents / pstn_phone_subscribers.php (Last retrieval time: May 2010).

[12] www.telecompk.net / 2010 / 02 / 09 / telecom-bangladesh-common-pakistan (Last retrieval time: August 2010).

[13] www.robi.com.bd (Last retrieval time: August 2010).

[14] www.technet.microsoft.com / en-us / library / ms174923.aspx (Last retrieval time: August 2010).

[15] Lon-Mu Liu et al., "Data Mining on Time Series: an Illustration Using Fast-Food Resturant Franchise Data," *Computational Statistics & Data Analysis*, vol. 37, issue 04, pp. 455-476, (January 2001).

[16] I. Aydin et al., "The Prediction Algorithm Based on Fuzzy Logic Using Time Series Data Mining Method," *World Academy of Science, Engineering and Technology* 51 (2009).

[17] Dennis Wegener et al., "On Integrating Data Mining into Business Processes," *13th International Conference on Business Information Systems (BIS 2010)*, Berlin, Germany, (2010).

[18] D. Neef et al., "Making the case for knowledge management: the bigger picture," Management Decision 37(1):72-78 (1999).

[19] C. Bhatt et al., "Knowledge management in organisations: examining the Interaction between technologies, techniques, and people," *Journal of Knowledge Management* 5, (1): 68-75 (2001).

[20] G. Costello et al., "Knowledge Management in Strategic alliances: The Role of Information Technology," *Templeton College. Oxford*, University of Oxford (1996).

[21] S. Raub and C. C. Ruling: "The knowledge management tussle – speech communities and rhetorical strategies in the development of knowledge management," *Journal of Information Technology* 16(2): 113-130 (2001).

[22] T. Nguyen Manh et al., "Data warehouse design 2: sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application," *Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP, DOLAP'05, ACM Press,* New York (2005).

[23] Olusegun Folorunso et al., "Data Mining for Business Intelligence in Distribution Chain Analytics," *International Journal of the Computer, the Internet and Management*, vol. 18 no.1, pp 15-26 (January-April, 2010).

[24] M. Alavi et al., "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and ResearchIssues," *MIS Quarterly 25(1): 107-136* (2001).

[25] X. Feng and H. Huang, "A Fuzzy-Set-Based Reconstructed Phase Space Method for Identification of Temporal Patterns in Complex Time Series," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 5, pp. 601-613 (2005).

[26] R. J. Povinelli et al., "Data Mining of Multiple Nonstationary Time Series, in Proceedings of Artificial Neural Networks in Engineering," *St. Louis, Missouri*, pp. 511-516, (2005).

[27] J. Han et al., "Data Mining: Concepts and Techniques," *San Francisco: Academic Press*, 800 p. (2005).

[28] E. Keogh and P. Smyth, "A Probabilistic Approach to Fast Pattern Matching in Time series Databases," *in proc. Third International Conference on Knowledge Discovery and Data Mining* (1997).

[29] C. Faloutsos et al., "Fast Subsequence Matching in Time-Series Databases," *Proc. Sigmod Record (ACM Special Interest Group on Management of Data Conf.)*, pp. 419-429 (1994).

[30] R. J., Povinelli et al., "A New Temporal Pattern Identification Method for Characterization and Prediction of Complex Time Series Events," *IEEE Trans. On Knowledge and Data Engineering*, vol. 15, no. 2, pp. 339-352 (2003).

[31] www.msdn.microsoft.com / en-us / library / ms173767.aspx (Last retrieval time: August 2010).