# CHAPTER 1 INTRODUCTION

1.1 Background of Problem

Telecom service provider industry is an industry that is always evolving with the development of telecommunications technology. Technological development is encouraging business development of broadband telecommunication services in order to offer a variety of new products and services. This leds to intense competition among service providers in maintaining and keeping customers for their continuation of their business. Consequently, many telecom service provider companies suffer losses from the loss of valuable customers who move to a competitor or any other telecommunications service providers; This is known as customer churn.

PT Telkom Indonesia is challenged to retain the existing customers and attract the customer to remain loyal to use of Telkom's products; for this reason, it always improves its services. From 2010 to 2011 the number of customer churn, especially in the field of broadband internet can be shown in the figure below:

Proportion of fixed broadband users who switched provider in the last twelve months in selected countries in 2010 and 2011
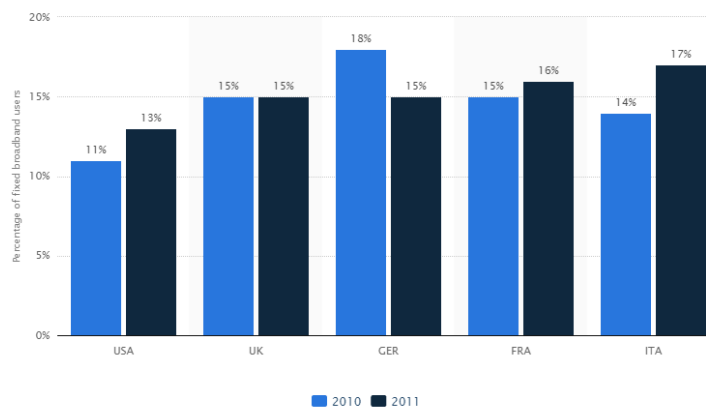
Additional Information:
Worldwide; Ofcom; October 2011; 5,090 Respondents; 18-64 years; Internet users

© Statista 2015
Source: Ofcom

Figure 1-1 Churned Broadband User in 2010 and 2011

However, the technical data mining to build churn prediction still has some limitation due to the spesific properties of data churn. There are two major spesific properties of data churn that usually become the problems in making a churn prediction model:

- The data usually has imbalanced class, the number of churn customer is usually very small or minority of the data;
- The data usually contains of annoying noise.

A dataset is called imbalanced if there are one or more classes that are dominating the whole dataset as majority classes, and other classes becoming minority. In addition, the data are usually contain annoying noise, which reflects inaccuracies in observations and the nature of the underlying process that misleads the representations and in turn it misleads the interpretation of the data collected [1]. The unwanted presence of outlier data (annoying noise) in the training data leads to a biased model.

Furthermore the Imbalanced data and the annoying noise in data will cause difficulty in making a churn prediction model so that the results of customer churn prediction are inaccurate and often reduces the accuracy of a model.

1.2   Problem Statement

Cases of churn are serious issue that often happens in the telecommunications company because it would affect the revenue earned by the company. Consequently, in order to overcome this issue the churners must be recognized before they churn.

Customer churn prediction model needs comprehensive schema due to some problems especially of data churn explained in Chapter 1.1. Consequently, it is necessary to develop an accurate churn prediction model so that it can minimize the loss of revenue from the loss valuable existing customers.

In addition, Broadband internet services are potentially one of the greatest sources of revenue for telecom service providers and consequently attractive features and services that are very influential in their marketing campaigns. Until now either very little churn prediction has been implemented on the broadband Internet services over fixed line networks, or the literature of churn prediction in telecommunication does not provide the details of methodologies for churn prediction using broadband information [2] [3]. Therefore, it is necessary to investigate the churn prediction in Fixed Broadband Internet service field.

1.3   Objective

Many data mining techniques have been used to create the prediction models, including the classification, clustering, association or a combination of both [2] [1] [3] [4]. In addition, the previous method using combined data mining techniques by combining two or more techniques has proven to provide better performances accuracy than many single techniques. However, the proposed method is expected to be able to handle the specific properties of data churn to produce an accurate churn prediction model.

For this reason, in this research is to buid a new method for churn prediction by developing the previous work combining with an other data mining technique. The research analyzes a data mining

method approach to identify churn on Broadband Internet Customer, especially in PT Telkom Indonesia.

The main purpose of this research was to minimize the loss revenue because of customer churn with the improvement of the accuracy and recall value of churn prediction. The method focuses on the used of Artificial Neural Network as a classifier. There are two methods that are going to be built, Combined Artificial Neural Network (ANN+ANN) and Combined Artificial Neural Network and SMOTEBoost Algorithm (ANN+SMOTEBoost+ANN). Later analyze the improvement the accuracy between those methods. Based on several findings [1] [5] [6] [7], there are some challanging points in prediction of churn in telecommunication sector are:

- How to develop a churn model which is able to predict accurately so that it can improve the performance of churn prediction.(Accuracy, Recall & F1 Measure)
- How to handle the imbalance and outlier data which usually occur in the data customer churn. (Issue of Data)
- How to implement churn prediction that can have an impact on churn rate reduction and revenue protection. (Revenue Protection)
- How to deal with the real data on specific Fixed broadband internet customer. (system implementation).

The proposed method is expected to be able to overcome the challenges above and work well as a churn predict on broadband internet customer in Telkom also in another broadband internet service provider.

1.4   Scope and Delimitation

To get the results of the research to achieve the expected goals, limitations problems are defined as follows:

1. Focus on Broadband Internet Customer of PT. Telkom Indonesia especially dataset of Pre Non Productive Customer (Pre-NPC).
2. The Time framed used is between December 2014 until December 2015.
3. The dataset used is Jakarta Region with the consideration that the region Jakarta has the greatest amount of data Pra NPC and has a variety of high profile people and competition product of internet broadband, so expected churn models might be able to be used in other regions.
4. Focusing on analyzing the used of Artifical Neural Network as a classifier to solve the churn problem on Fixed Broadband Internet.
5. The study is conducted in the prototype stage.

1.5   Contibution and Significant

Churn Prediction is a phenomenon which is used to identify the possible churners in advance before they leave the network. Thereby, the potential loss of the company could be avoided. This research utilizes data mining techniques to identify the churners.

Previous works using combined techniques data mining have been performed to improve the performance accuracy compared to single technique (conventional technique). This study proposes new schemas of customer churn prediction using combined technique in which consider not olny can used to overcome the problem of noise, but also can be used to overcome the problem of imbalanced class data.

This method using combined feed neural network and SMOTEBoost Algorithm . While the feed forward neural network is used to overcome the problem of noise, SMOTEBoost used to overcome the problem of imbalanced class. SMOTEBoost is combined the standard Boosting procedure (AdaBoost) with an oversampling technique called SMOTE(Synthetic Minority Oversampling Technique) which creates new minority class examples by extrapolating between existing examples. The proposed schemas will be usefull for :

1. Produce a powerful model to predict the subscribers who are likely to churn using dataset contains of annoying noise using and imbalanced data . While the feed forward neural network is used to overcome the noise of data, SMOTEBoost used to overcome the imbalanced data.
2. Sampling techniques approach for handling imbalanced data using SMOTE with oversampling technique. This technique create additional data from minority class (does not reduce the potensial data), in contrast using undersampling technique which can reduce the data record that causes big loss of potential data from the majority class in imbalanced data, so that it reduces prediction performance.
3. This proposed method can be applied to real data on specific Fixed broadband internet customer with high accuracy .

Finally, this proposed method will be able to minimize the losses suffered by Telecom's operator caused by churn especially in PT.Telkom Indonesia. By doing this research, this helps the company to prevent subscribers who are likely to churn in future by taking the required retention policies to attract the likely churners and to retain them.

1.6   Thesis Organization

This thesis consists of some chapters. Each chapter is related to another chapter.
a) Chapters 1 is Introduction. It tells about overview of churn, the problem and the objective.
b) Chapter 2 is Literatature Review. It tells about existing works related to churn prediction.

c) Chapter 3 is Methodology and System Design. It explains the methodology used in this study and the system design of the proposed schemas.

d) Chapter 4 is Implementation and Analysis. It shows the experimental result and give the analysis of the result.

e) Chapter 5 is Conclusion and Recommendation. It tells the conclusion from this study and future work for next development related to this study.