ISBN:978 0 9943656 37
www.apiar.org.au

BUSINESS INTELLIGENCE FOR PROFILING OF TELECOMMUNICATION CUSTOMERS

Rokhmatul Insani [a], Hira Laksmiwati Soemitro [b]
[ab] Institute of Technology Bandung, Indonesia
Corresponding email: 23513097@std.stei.itb.ac.id

Abstracts

Business Intelligence is a methodology that exploits the data to produce information and knowledge systematically.Business intelligence can support the decision making process. Some methods in business intelligence are data warehouse and data mining. Data warehouse can store historical data from transactional data. For data modelling in data warehouse, we apply dimensional modelling by Kimball. While data mining is used to extract patterns from the data and get insight from the data, Data mining has manytechniques, one of them is segmentation. For profiling of telecommunication customers, we use customer segmentation according to customer's usage of services, customer invoice and customer payment. Customers can be grouped according to their characteristics and can be identified as the profitable customers. We apply K-Means Clustering Algorithm for segmentation. The input variable for that algorithm we use RFM(Recency, Frequency and Monitary) model. For all the processes in data mining, we use tools IBM SPSS modeller.

Keywords : Business Intelligence, Customer Segmentation,Data Warehouse, Data Mining

1.Introduction

Nowdays, the business management concept is greatly envolved with the customer-oriented idea. Managing customers has become the most critical factor of an enterprise's success, including in the telecommunications industries. Profiling of customers can capture the business and its customers. Profiling customers can help enterprises to understand the life cycle of customer and provide efficient and personalized services for their customer. We can use Business Intelligence to profiling of customer. Business Intelligence is methodology, architecture or tools that can process the data to the usefull information and knowledge for decision maker. The ilustration of knowledge discovery by Han, Kamber and Pei (2012):

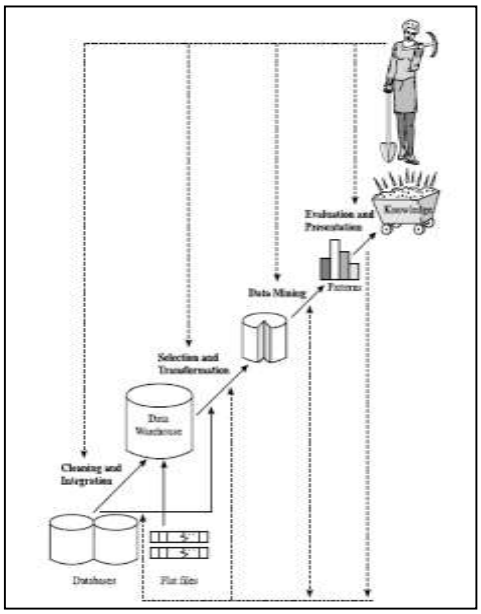


Figure 1Knowledge discovery by Jiawei Han

ISBN:978 0 9943656 37
www.apiar.org.au

The First step is to create data warehouse then analyze the data using data mining. Data warehouse is used to store historical data from transactional data. Data mining is the process to extract useful information and knowledge. Data mining helps users to identify valuable pattern contained in data to help the major decision. The data qualities is the most critical factor that affect to the output of data mining.

There are many techniques in data mining. One of them is segmentation. Customer segmentation is to clasify the customer according to the customer's attributes, behaviour, values and others. Customer segmentation can provide appropriate product, services and marketing models to the customer. Segmentation enable the enterprise to understand which customer have higher values customer.

2.Related Work

Some research about business intelligence for customers has been developed. Customer segmentation of bank based on data warehouse and data mining, proposed by Shuxia Ren, Qianing Sun and Yuguang Shi, used SOM and K-Means Algorithm to segment the customers (2010). Customer segmentation using decision tree to identify VIP customer in mobile communication industry by Yihua (2010). Xiong Weiwei et al. (n.d.) discussed the customer segmentation using RFM value and grey relation for the logistic market. This is where they used AHP for identify indicator weight. Calculating customer value using RFM model in airline industries proposed by Liu Jiale and Du Huiying (n.d.), they use AHP (Analytical Hierarchy Proses) for calculating potential value. Based on their research, there is no research about the whole process about business intelligence to profiling of telecommunication customer which is to create data warehouse for store the historical data then analyze the customer segmentation using data mining.

Data Architecture of Business Intelligence

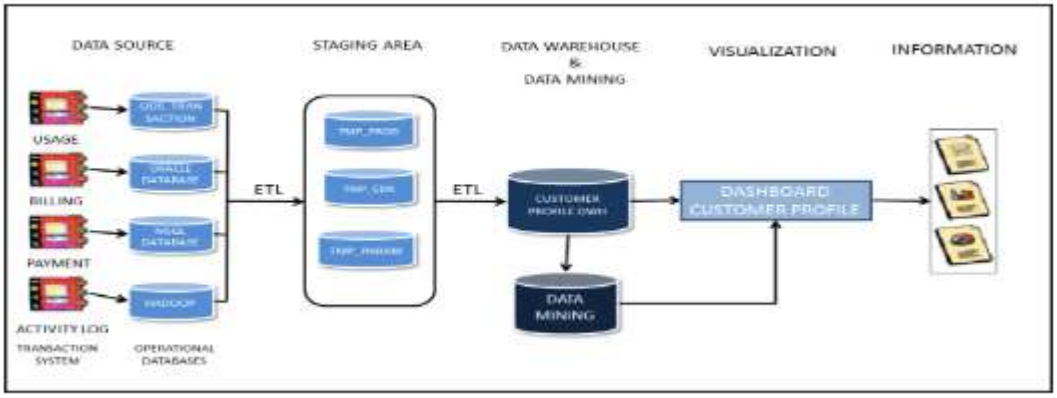


Figure 2 Data Architecture for Customer Profile

Data source from transactional system that has different operational databases and can be of any format. Staging area is a bridge of transactional data to data warehouse. ETL (Extract-Transform-Load) is a process to identify, integrating and load the data from operational database to data warehouse. Detail data warehouse created to store historical transaction based on business process in this company. Data in data warehouse is used to analyze the pattern of data using data mining. After that, the data is visualized in a dashboard so the information can be understood by users. In this research, we only propose about data warehouse modelling and data mining.

3. Data Warehouse for Telecommunication Customer

The definition of data warehouse according to Bill Inmon: A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process. Subject-Oriented: A data warehouse can be used to analyze a particular subject area. Integrated: A data warehouse integrates data from multiple data sources. Time-Variant: Historical data is kept in a data warehouse. Non-Volatile: Once

ISBN:978 0 9943656 37
www.apiar.org.au

data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

Thera are two approaches for modelling data warehouse introduced by Bill Inmon and Ralph Kimball. Bill Inmon's approach (the top-down design): A normalized data model is designed first. Then, the dimensional data marts, which contain data required for specific business processes or specific departments are created from the data warehouse. Ralph Kimball's approach (the bottom-up design): The data marts facilitating reports and analysis are created first; these are then combined together to create a broad data warehouse. For building data warehouse, Kimball's approach takes lesser time than Inmon's approach and the data modelling design of Kimball's approach is simpler then Inmon's approach, so we choose Kimball's approach for modelling data warehouse.

4. Dimensional Modelling

Kimball's data warehousing architecture is also known as Data Warehouse Bus. The database-independent bus architecture allows for incremental data warehouse and business intelligence (DW/BI) development. It decomposes the DW/BI planning process into manageable pieces by focusing on the organization's core business processes, along with the associated conformed dimensions. Conformed dimensions are master dimensions that are managed once in and reused by multiple fact tables. They support the integrate data from multiple business processes.

With Modelling in data warehouse, there are fact tables and dimensional tables. Fact table consist of measurement or fact of a business process. Dimensional table consist of reference information about a measurable event. First step for modelling data warehouse is to define bus matrix. Bus matrix discribe the business process in a company. According to Kimball and Margy Ross, the enterprise data warehouse built on the bus architecture "identifies and enforces the relationship between business process metrics (facts) and descriptive attributes (dimensions)". Before, we define bus matrix, we must describe the company's bussiness process. Figure 3 below, shows the description about business process in telecommunication company.

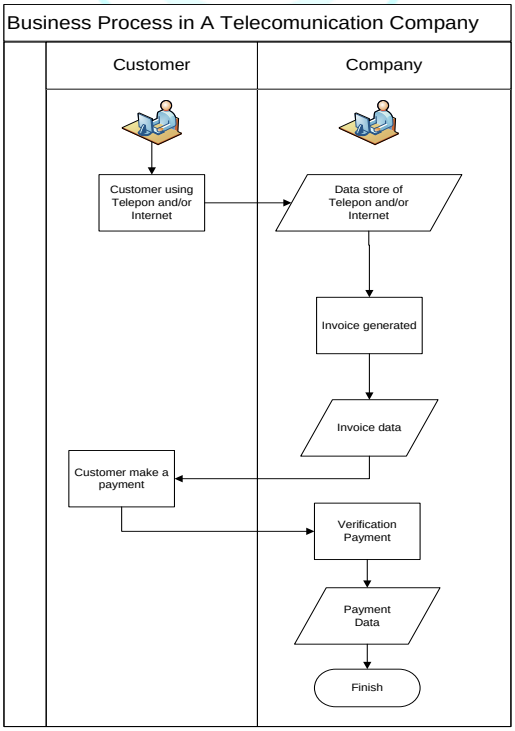


Figure 2 Flowchart of Business Process in Telecommunication Company

ISBN:978 0 9943656 37
www.apiar.org.au

In a telecommunication company which has many branches in Indonesia, there are two services that are telephone services and internet services. The customer can use more than one service. For example, a customer have 2 line of telephone services and 2 line of internet services. Usage of services will be store in operational database. Every month, the company will generate invoice to its customer. Then, the customer will make payment. Every business process store in transactional database. After we know about the business process in this company, we describe that business process into bus matrix. Figure 4 below, show the bus matrix about business process in this telecommunication company.

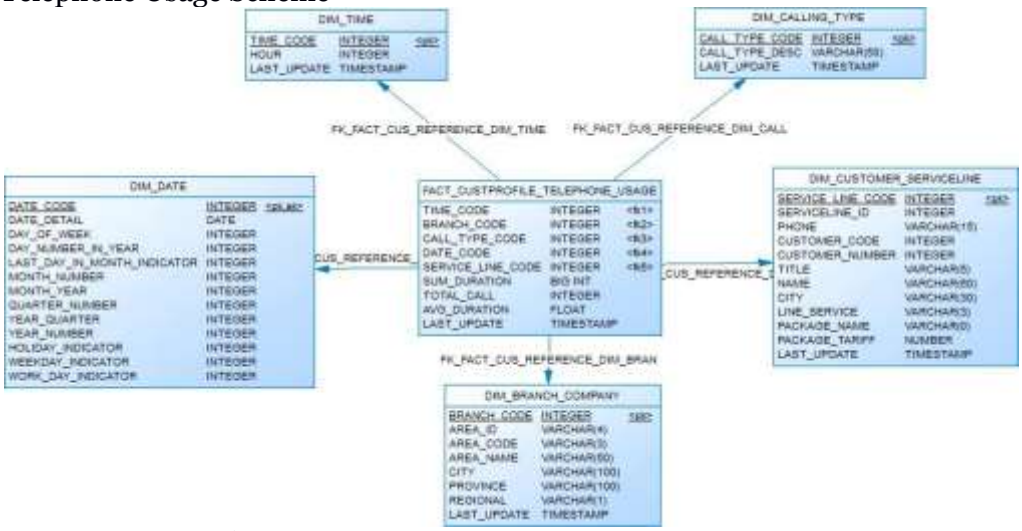
	date	time	branch company	customer - serviceline	product	channel list
telephone usage	X	X	X	X		
internet usage	X	X	X	X		
invoice	X		X	X	X	
payment	X		X	X	X	

Figure 3 Bus Matrix in a Telecommunication Company

Bus matrix above (figure 4) illustrates the mapping of fact tables candidate (left side) and the dimension tables candidate (top side). And mark X means that the fact table associated with the dimension table. The bus matrix is blue print fordimensional modelling of data warehouse.

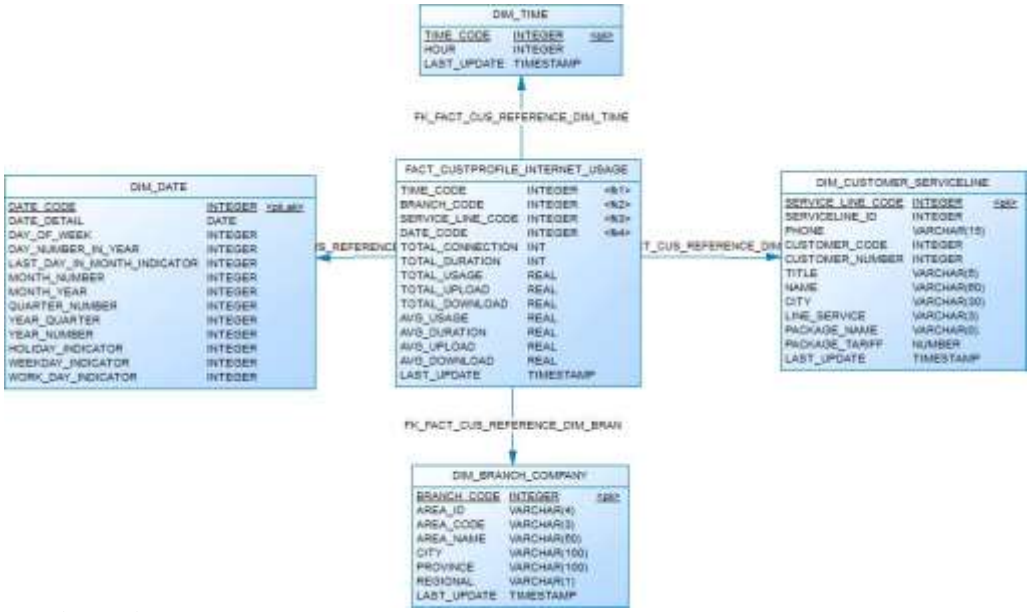
There are two types of dimensional modelling that are star schema and snowflake scheme. A star schema consists of one fact table and a number of dimension tables which associated with fact table. The Snowflake scheme has more complex structure than star scheme. In the snowflake schema, to be normalized to eliminate redundancy, the dimension table is grouped into several tables. The dimensional modeling is not only based on business processes, but also based on existing transactional data. Based on bus matrix, we have four schema that use star scheme. The four scheme of dimensional modelling are:

1. Telephone Usage Scheme

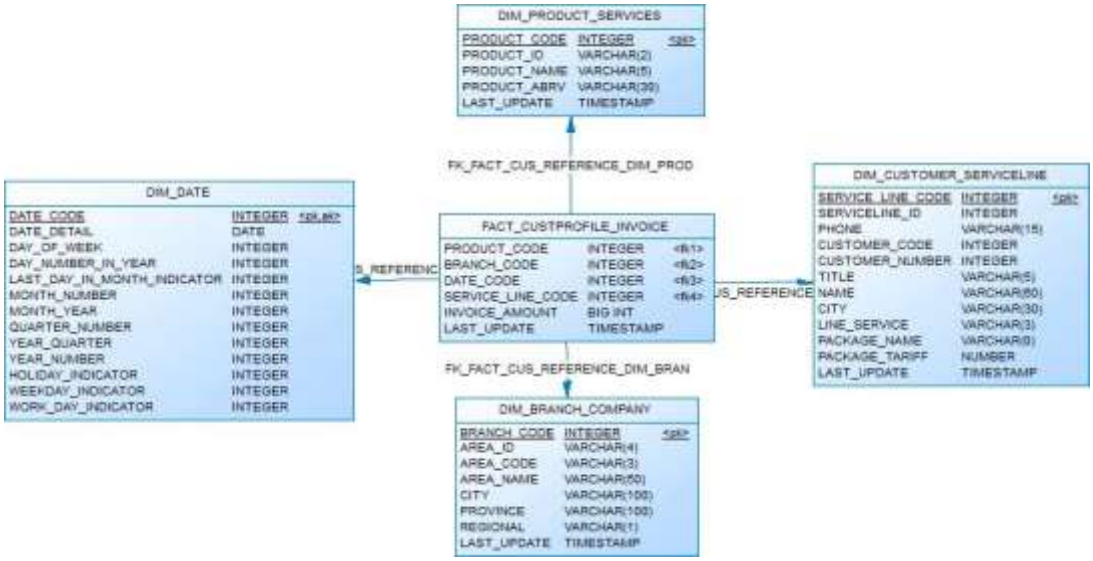


2. Internet Usage Scheme

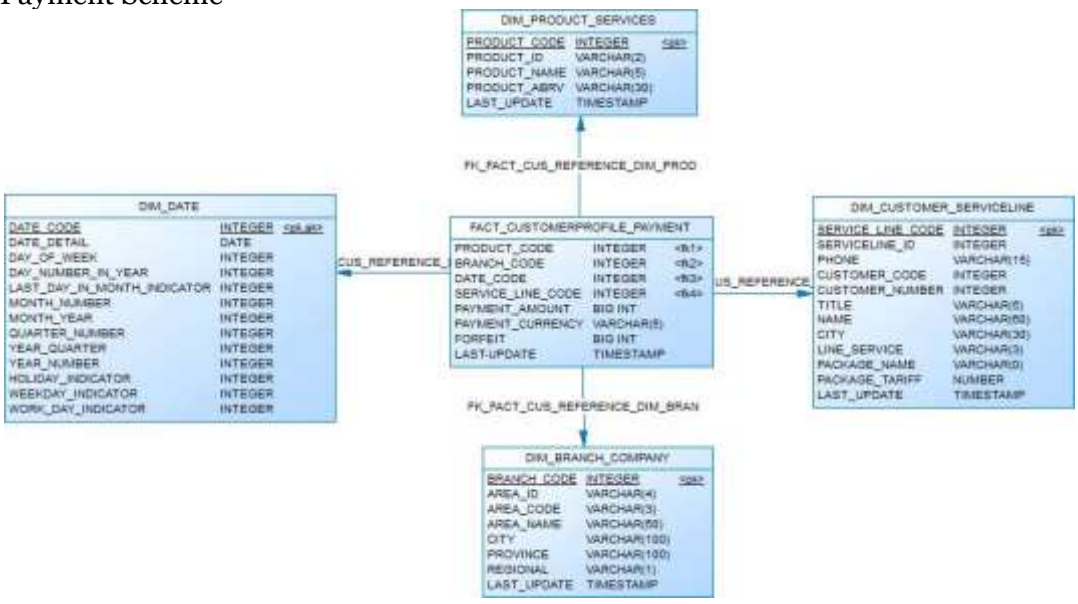
ISBN:978 0 9943656 37
www.apiar.org.au



3. Invoice Scheme



4. Payment Scheme



ISBN:978 0 9943656 37
www.apiar.org.au

5. Data Mining for Telecommunication Customer

According to Jiawei Han, many people treat data mining as a synonym for another popularly used term,knowledge discovery from data(KDD), while others view data mining as merely anessential step in the process of knowledge discovery. The method of data mining include description, association rule, classifiation, clustering, outlier analysis, etc.

Customer Segmentation

Customer segmentation is grouping customers based on specific factors, such as age, interest, behaviour, etc. Companies also hope to gain a deeper understanding of their customers. Customer segmentation can be used to identify profitable customer.

In data mining, we can use clustering method for segmenting. One of the clustering algorithm is K-Means Clustering. K-Means is the simplest clustering algorithm. K-means groupedthe objects into K clusters. K is the number of clusters that will be generated, defined by the user.
Step in K-Means Clustering Algorithm is :

1. Decide the number of clusters k
2. Initialize the center of the clusters
3. Attribute the closest cluster to each data point
4. Set the position of each cluster to the mean of all data points belonging to that cluster
5. Repeat steps 3-4 until convergence

In the input for K-Means Alorithm, we propose using RFM model. The RFM model is proposed by Hughes in 1994. The model is used to present customer behaviour characteristics. The model is distinguishing customer by three variable i.e. interval of customer consumption, number of customer usage and money amount. The variabel model is : Recency (R) is customer's last purchase, Frequency (F) is the total number of purchase during a spesific period and Monitary (M) is monitary is the amount of money used to purchases in during a spesific period.

Based on the data warehouse scheme above, Recency can initialized as maximal time for customer usage either telephone and internet services. The smaller the value will be better, it means that the customer just make a transactions. Frequency can be initialized as the total call of customers and the total internet connection of customers in a specific period. The larger value will be better, it means customer transactions often. Monitary can be initialized as the total payment of customer in a specific period. The larger value will be better, it means customer is providing high income to company.

For the data mining process, we use IBM SPSS Modeller. IBM SPSS Modeler is a data mining and text analytics software application built by IBM. It has a visual interface which allows users to leverage statistical and data mining algorithms without programming.

To demostrate of our proposed methode, we use telecommunication customer data. Data period is from January 1st 2015 until 7 September 2015.The First step is data preparation, Integration and transform the data, sumarizing data in a particular month. The sample data for customer segmentation shown in table below:

Customer_id	last_usage	total_usage	product	payment_amount
4734988	20151116	50	Phone	1546603
22121437	20151020	30	Phone	185502
7253024	20151121	45	Phone	434500
2298452	20151029	100	Internet	451584

ISBN:978 0 9943656 37
www.apiar.org.au

4734988	20121114	112	Internet	695000

Table 1: Sample data for customer segmentation

The process of customer segmentation using SPSS modeller shown in Figure 5.

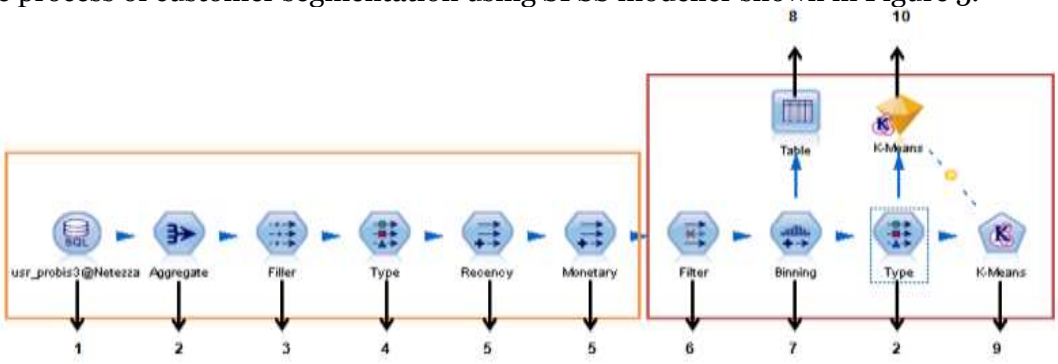


Figure 4: Customer Segmentation using SPSS Modeller

Description:

1. Data Source node : Retrieve the data from database, in this research we use DBMS netezza.
2. Aggregate node : To agregatedata based on customer_id
3. Filler node : To transform the value in a columnand change the same column according to the new value
4. Type node : Determining a data source property
5. Derive node : To transform the value and add new column for that value.
6. Filter node : Choose or rename the variable name that will be used
7. Table node : Looking at the data
8. Binning node : Simplifying the value, for example change the data range from 1 until 100.
9. K-means node : Create a cluster model based on K value, we use K=4 to grouping customer in 4 segment.
10. Generate node : Results of the k-means clustering node

And the results are shown below:

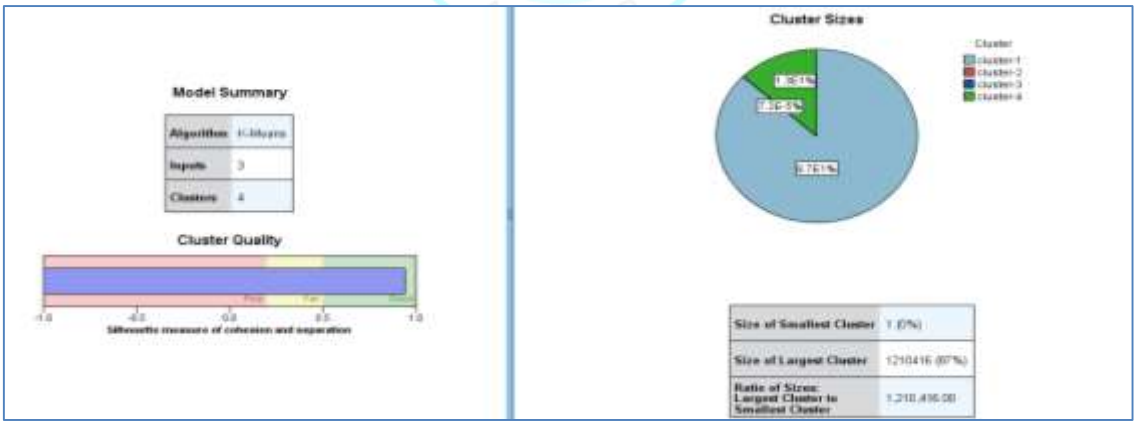


Figure 5 Result of customer segmentation

ISBN:978 0 9943656 37
www.apiar.org.au

Figure 6 Clustering of Customer

From the figure 7 shown that:

Cluster 1: high recency, low frequency, low monitary; it's means the customer rarely used the services, indicates that the service has not been used, this customers have high possibility for customer churn.
Cluster 2: low recency, high frequency, low monitary; it's means the customer in this group customers often use the service, it's indicated loyal customer.
Cluster 3: moderate recency, low frequency, moderate monitary; compared with the other cluster, the customer spend a lot of money to the company. It's indicated profitable customer.
Cluster 4: low recency, low frequency, low monitary ; this cluster have the same characteristics with cluster 1. The difference is the latest usage of customer. The last usage by customer relatively recent usage, so the churn rate is lower than cluster 1.

Conclusion

This research attempts to try business intelligence applying in the field of customer of telecommunication based on data warehouse and data mining. Data warehouse is used to centralizing the transactional data and store the historical data. Data warehouse store data summary in certain granularity so that the stored data is less than transactional data, its impact in increasing query performance for data retrieval.
Data mining for customer profile is used for customer segmentation. To structure the customer segmentation model, this research take RFM model and K-Means Clustering. The result shows that the methode is feasible to be built. The results of this research, the customer can be identified loyal customers, profitable customers and customers who will be churn. After customers are grouped, the company can provide treatment to each customer group.

Acknowledgement

The research is suported by Indonesian Endowment Fund for Education (LPDP Scholarship Program).

ISBN:978 0 9943656 37
www.apiar.org.au

References

i. Abbott, D., 2014. *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. USA: Willey.

ii. Chen, Y., Zhang, G., Hu, D. & Wang, S., 2006. *Customer Segmentation In Customer Relationship Management Based On Data Mining*. IEEE

iii. Danhua, J., Xiaoegeng, Z. & Runrun, W., 2012. *Research on the Amount of Customers in Telecom Package Preview Based on Data Mining*. IEEE

iv. Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concept and Techniques*, 3rd edn, USA: Elsivier.

v. Inmon, W. H., 2005. *Building the Data Warehouse*, 4th edn, John Wiley and Sons, Inc.

vi. Kimball,R., Ross, M., Thornthwaite, W., Mundy, J. & Becker, B., 2007. *The Data Warehouse Lifecycle Toolkit*, 2nd edn, USA: Willey.

vii. Ranjan, J., 2005-2009. Business Intelligence: Conceps, Components, Techniques and Benefits. *Journal of Theorical And Applied Informatiom Technology*.

viii. Ren, S., Sun, Q. & Shi, Y., 2010. *Customer Segmentation of Bank Based on Data Warehouse and Data Mining*. IEEE

ix. Yihua, Z., 2010. *Vip Customer Segmentation Based on Data Mining in Mobile-communication Industry*. IEEE