Analysis of Best Life Insurance Cover Selection using Data Mining Technique

Mr. T Udaya Kumar[1], MuneeshaVunnam[2], KeerthiSonga[3], MounikaVattigunta[4], Venkata Ravi TejaVemulakonda[5]
[1]Asst. Professor, [2,3,4,5]B.Tech Students
Dept. of CSE, LBRCE, Mylavaram.

Abstract - There are millions of Indians who are continuously putting the efforts to invest regularly in an insurance plan. Investing money will not only give large benefits but also help the individual to improve his lifecycle. Life Insurance sector offers a huge range of products fulfilling the needs of people like financial protection, tax benefits, health expenses cover etc., This paper will help the individual to predict which policy is best suitable for him by considering various factors. Finding frequent patterns are useful for analysing and predicting customer behaviour. This Frequent itemset mining can help the individual to choose the perfect life insurance plan.

Keywords - Life insurance sector, frequent patterns, Frequent itemset mining

I. INTRODUCTION

Data mining is the process of sorting through large datasets to identify patterns and establish relationships to solve problems through data analysis. In data mining, association rules are created by analyzing data for frequent patterns, then using support and confidence criteria to locate the most important relationships with the data.

Data mining can be accomplished by building different models based on algorithms which act on large dataset which will help to mine the data and generate new results. In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact business. Frequent pattern mining is an important data mining task and a focussed theme in data mining. The patterns that appear frequently in a dataset are called Frequent patterns. This can be explained easily by Market-basket analysis. It gives the relationships among the attributes.

This paper mainly focuses to help the person to choose the best suitable life insurance plan according to customer information and other factors. In this paper, frequent patterns are generated based on individual's age, gender, monthly income, monthly expenses, etc. and their relationships with each other. We will generate the frequent item sets and its analysis of best plan suitable to the person.

II. RESEARCH METHODOLOGY

A. Method - Finding frequent item sets to determine the measure of interestingness.

i). Itemsets - Let $V=\{v_1,v_2,......,v_n\}$ be the set of items in market basket data and $P=\{p_1,p_2,.......,p_n\}$ be the set of transactions. Each transaction v_i contains a subset of items chosen from i. In association analysis, a collection of zero or more items is termed as itemset. If an itemset contains k items, it is called a k-itemsets.

ii). Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min sup.

B. Algorithm - The Apriori algorithm is the most well known association rule algorithm and is used in most commercial products. It uses the following property, which we call the itemset property: *Any subset of a large subset must be large.*

The basic idea of the Apriori algorithm is to generate itemsets of a particular size and scan the database to count these to see if they are large. During each scan of the items, counts are counted using minimum support and confidence. Only those itemsets that have support more than minimum support can be considered as large itemsets and those can be used to generate contenders for the next pass. Li that is levels are used to generate next counts. An itemset is considered as a rule only if all its subsets also are large. These itemsets are then joined to make of large itemsets found in previous pass.

The following are the steps of Apriori algorithm:

1. Scan the data in excel sheet which contains the items such as {age, gender, monthly income,..etc}.
2. Calculate support associated with each and every item to create an itemset.
3. If support of any of these items is large it is considered as a large itemset i.e. L1.
4. The itemsets generated in the iteration are joined to make large itemsets which is known as candidate-itemset generation and uses Apriori property to prune un-frequent itemsets.
5. Compare the support of each itemset with the minimum support count.
6. If the obtained set is Null Goto step 4.
7. Generate nonempty subsets for each frequent itemset.
8. The frequently generated item itemsets is used to find insurance plan according to given items.

C. Data - Data mining is applied on the source of data using association rules. Some records were collected as a sample on which Apriori algorithm was applied. The variables, such as demographic profile, lifestyle, and asset allocation would affect the purchase of insurance products.

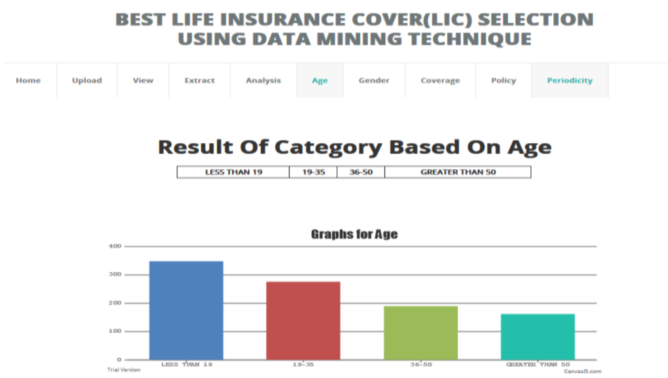
So for this paper, the basic customer's information (age, sex, monthly income, marital status, monthly expenses, etc) and insurance transaction data (policy type, policy plans, mode of payment, term, etc) is used.

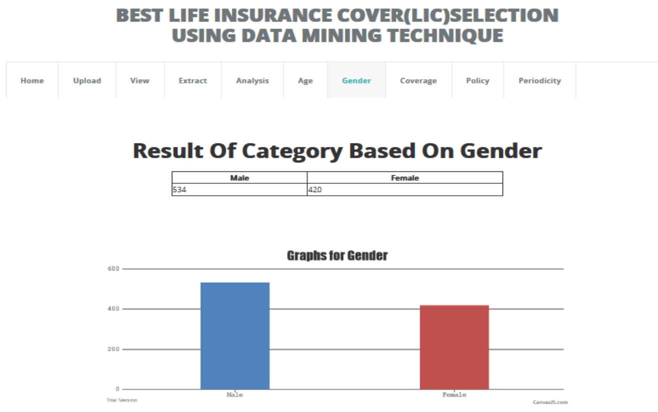
III. RESULTS AND DISCUSSIONS

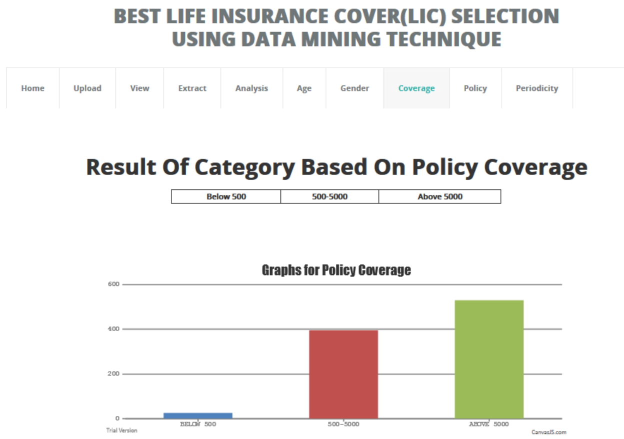
The results for this paper are the frequent patterns generated after applying Apriori algorithm on samples. These patterns will be used in the making of an application which will give the best insurance plan to the customer according to their personal and financial information. This will guide the customer to make an informed choice from the plans available for investment. This will give idea to the customer that where to invest. Itemsets can be created based on age, gender, monthly income and expenses which will conclude relationships with policy types.

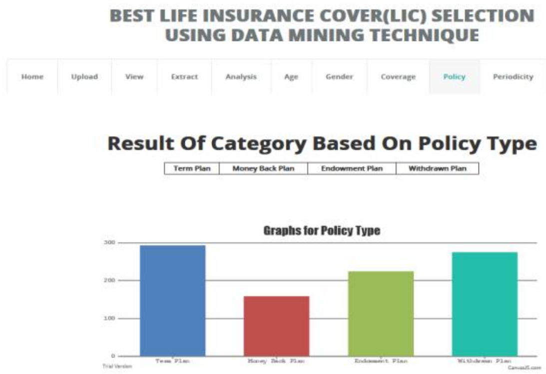
These are the results after applying the apriori algorithm on the sample data :

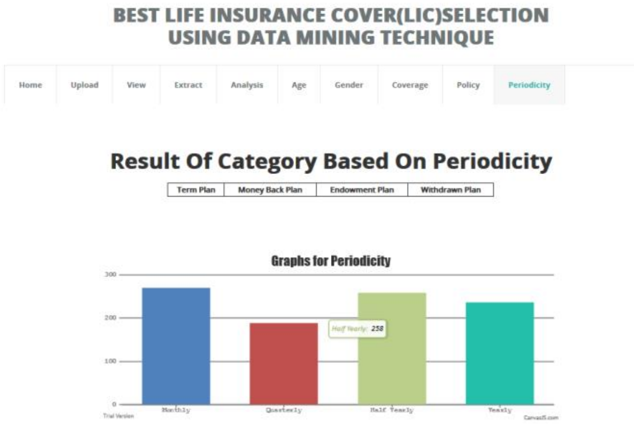
Withdrawn Plan : 75
Money Back Plan : 46
Term Plan : 81
Endowment Plan : 65
Money Back Plan, Endowment Plan : 14
Withdrawn Plan, Term Plan : 34
Endowment Plan, Withdrawn Plan : 25
Money Back Plan, Term Plan : 19
Endowment Plan, Term Plan : 23
Money Back Plan, Withdrawn Plan : 14
Money Back Plan, Endowment Plan, Withdrawn Plan : 2
Money Back Plan, Withdrawn Plan, Term Plan : 3
Endowment Plan, Withdrawn Plan, Term Plan : 2











IV. CONCLUSION

This paper proposed an association model using Apriori algorithm based on some market basket analysis. These rules can guide the client to choose the perfect life insurance plan. This paper only focuses on life insurance policies offered by LIC of India and many private companies can be added to enhance the feature of proposed system and for which one might need to change the algorithm. Also more user friendliness can be provided with a good accuracy so that it can be implemented in real life.

V. REFERENCES

[1]. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining, And OLAP",MC Graow–Hill, 1997.
[2]. Bigus and Joseph P, "Data Mining With Neural Networks", MC Graw–Hill, New York 1996.
[3]. Christopher J. Matheus, Gregory Piatetshy–Shapiro and Dwight Mcneill", Selecting and Reporting what is Interesting The Kefir Application to Health Care Data", Advances in Knowledge Discovery and Data Mining, AAA1 Press/The MIT Press, 1996
[4]. David Cheung, Vincent T., Ada W. Fu and Yongjian Fv, "Efficient Mining of Association Rules in Distributed Databases", IEEE, 1996.
[5]. Hongjun LU, Ling Feng and Jiawei Han, "Beyond Intratransaction Association Analysis: Mining Multidimensional Intertransaction Association Rules", ACM Transactions on Information Systems, Vol. 18, October 2000.
[6]. Huan Liu, Farhad Hussain, Chew Lim Tan and Manoranjan Dash, "Discretization: An Enablin Technique", Data Mining and Knowledge Discovery", vol. 6 No. 4, October 2002.
[7]. Raymond T. Ng, Laks V. S. Lakshmanan, Jiawei Hon and Alex Pany, "Exploratory Mining and Pruning Optimizations of Constrained Associations Rules", ACM 1998 page 13.
[8]. Ramakrishnan Srikant and Rakesh A. Grawal, "Mining Generalized Association Rules", Proceedings of The '21st VLDB Conference", Zurich, Switzerland, 1995.
[9]. Ramakrishnan Srikant and Rakesh A. Grawal, "Mining Quantitative Association Rules in Large Relational Tables", Proc Sigmod '96, 6/96 Montreal Canada, 1996 ACM.
[10]. Natalya Friedman Noy and Carole D. Hafner, "The State of The Art in Ontology Design", AI Magazine Vol. 18, No. 3, Fall 1997.
[11]. Mr. A. B. Devale and Dr. R. V. Kulkarni "A REVIEW OF DATA MINING TECHNIQUES IN INSURANCE SECTOR" Golden Research Thoughts Vol - I , ISSUE - VII [January 2012]