

Help the Society in Selecting Their Best Life Insurance Cover (LIC) Using Data Mining Technique

Prathama Nemalekar ^[1], Prof. Amol Joglekar ^[2]

M.Sc Computer Science ^[1]

Department of Computer Science ^[2]

Mithibai College, Mumbai

Maharashtra - India

ABSTRACT

Investment is the key for making money. Investing small amount of money will give one large benefits and also help to improve an individual's lifestyle. Today in India, the life insurance sector is increasing so rapidly that most of the people are interested to invest some share of their income in the Life Insurance sector. Investing in LIC is not only for profit but it is also to compensate the loss. Company uses data mining techniques which help to give precise results and determine pricing, customer preferences and product positioning, impact on sales, customer satisfaction and corporate profits. This paper will help to predict which policy type is suited for an individual by considering various factors. These predictions will help to choose the best policy plan for an individual using association model.

Keywords:- Association Rules, Customer Information, Life Insurance Policies, LIC, Socio-Economic Factors

I. INTRODUCTION

Life Insurance Corporation of India (LIC) is the biggest life insurance player in Indian Insurance Industry. LIC offers a very vast and huge range of products catering to needs of various sections of people in India. It has introduced 22 crore policies and an Asset base of 8 lakh crore. LIC has the Trust that goes beyond generations. It is a government owned entity, so that our money resides on the safest hands [7].

Data mining can be used to search through large set of data to discover patterns and trends which cannot be analysed simply by data queries. Data mining uses mathematical algorithms that easily help to predict the output and evaluate probability of future events [5]. It will help to find different patterns and relationships among the data. Data mining can be accomplished by building different models based on algorithms which act on large dataset. These models will help to mine the data and generate new results. Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbour method etc., are used for knowledge discovery from databases [6]. Association rule mining is the most efficient data mining technique to search hidden or desired patterns which helps to get correlation relationships among various data attributes in large sets of database. A typical example of association rules mining is market basket analysis [11]. An association rule is implication relation in the form $X \rightarrow Y$ between two sets of items X and Y. Association rules aid to find relationships among different items according to

buying behaviour of a customer. These results are computed with the help of probability [12].

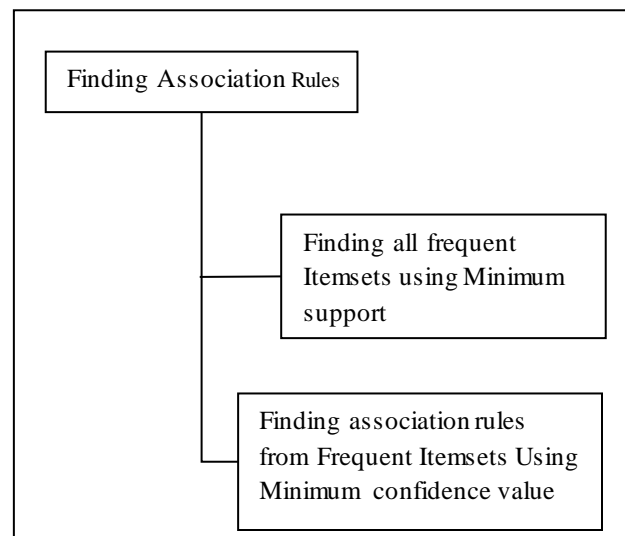


Fig 1: Generating Association rule

This paper mainly focuses to help the person to choose the best suitable life insurance plan according to his/her information and other factors. Association rules are useful for finding relationships among all attributes. In this paper, the different rules are generated based on individual's age, gender, monthly income, monthly expenses, etc. and their relationships with each other. All these factors can be considered for suggesting the right LIC plan for that person.

II. LITERATURE REVIEW

A.B. Devale and Dr. R. V. Kulkarni [1] investigated various data mining techniques such as clustering, association rules, classification and correlation for knowledge discovery in insurance business. By applying these data mining techniques, the companies can develop data about customer's buying pattern and behaviour also they can improve their business by reducing frauds and enhancing risk management.

V. SreeHari Rao, Murthy V. Jonnalagedda [2] experimented to extract the behavioural patterns for customer retention in health care insurance. The concept of Novel index tree (a variant of K-d tree) clubbed with K-Nearest Neighbour algorithm is proposed for efficient classification of data, as well as outliers and the concept of insurance dynamics is proposed for analyzing customer behavioural patterns. Among insurance policies that deal with automobiles, health, life, property, and so on, this paper limited the scope of this article to address the question of retention of customers in the health insurance sector. This paper dealt with the dynamical activity of the insurance industry, classifying the customers based on their characteristics.

Kuo Wang, and Jhieh-Yu Shyng [3], used definition of Customer Life Value (denoted as CLV), a suitable model were designed and customers' present values were estimated by given data from insurance company. Two data mining technologies (Rough Sets Theory and Decision Tree) were introduced and applied to find the rules and factors which might have influence on customers' values. The study is designed as follows. First, reviewing of all CLV models published previously and using them to calculate current values of customers. Then, we applied RST and DT to select useful variables and found out the rules affecting customers' value factors, based on the customers' information (such as age, gender, annual income and occupation) and internal information (such as category of insurance, amount of insured and gross premium) provided by an insurance company. The results and efficiencies were compared as well as the advantages of these two methods in this study.

Abdhesht Gupta, Anwiti Jain [4], introduced a web recommender system is proposed for life insurance sector based on web data mining using association rule which supports the insurance needy as well as life insurance representative to select best suitable life insurance plan for any particular person. They introduced Cold-start problem

which is discussing and proposing for new customers that how to acquire new customers information into system and get best recommendation for new customers. They also explored the data mining technique for recommendation system using association rule mining with some improvement in traditional recommendation system. In this work some high level association rule mining method is used to retain existing customer for new policy. Dual clustering method is used to overcome the limitation of single clustering method which gives more accurate and appropriate recommendation to solve cold-start problem.

III. RESEARCH METHODOLOGY

A. METHOD

Association rules are useful for analyzing and predicting customer behaviour. Association rules are used to identify strong rules according to measure of interestingness.

ITEMSETS

Let $I=\{i_1, i_2, \dots, i_d\}$ be the set of all items in market basket data and $T=\{t_1, t_2, \dots, t_N\}$ be the set of all transactions. Each transaction t_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 -itemset [8].

SUPPORT AND CONFIDENCE

An association rule is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint itemsets, i.e., $X \cap Y = \emptyset$.

The strength of the association rule can be measured in terms of its support and confidence.

Support determines how often a rule is applicable to a given dataset while confidence determines how frequently items in Y appear in transactions that contain X .

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}; \quad (1)$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}. \quad (2)$$

Support is an important measure because a rule that has very low support may occur simply by chance. For this, support is often used to eliminate uninteresting rules [8].

Confidence, on the other hand, measures the reliability of the inference made by a rule. For given $X \rightarrow Y$, the higher the confidence, the more likely it for Y to be present in transactions that contain X [8].

ASSOCIATION RULE

In general, association rule mining can be viewed as a two-step process:

1. Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, *min sup*.
2. Generate strong association rules from the frequent itemsets: By Definition, these rules must satisfy minimum support and minimum confidence [9].

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. 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 [10].

These are the steps of Apriori algorithm

STEP 1: Scan the Excel data to create first itemset with 0.1 support on items like {age, gender, monthly income, expenses and policy type}. Find support associated with each item to create an itemset. If support found of any one these items is large then it is a large itemset i.e. L1.

STEP 2: Itemsets created in iteration are joined together to make new large itemsets and use Apriori property to prune the unfrequented itemsets.

STEP 3: Find support of each itemset found in above iteration and compare the support with minimum support to find more itemsets.

STEP 4: The obtained set = Null
goto step 5 else goto step 2.

STEP 5: For each frequent itemset 1, generate all nonempty subsets of 1

STEP 6: Using nonempty itemsets best rules can be obtained using confidence of 0.9 which can be used to find insurance plan according to the given items.

C. DATA

Data mining is applied on the source of data using association rules. Fifty records were collected as a sample on which Apriori algorithm was applied. Socio-demographic 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.

Table 1: Customer’s Information

Sr No	Items	Options
1	Gender	M] Male F] Female
2	Age	A] Less than 18 years B] 19-30 C] 31-40 D] 41-50 E] 51-60
3	Marital status	Married Unmarried
4	Occupation	Government Private Student Housewife Medical Engineers
5	Monthly income	A] Less than 5000 Rs. B] 5001-15000 C] 15001-30000 D] 30001-45000 F] More than 45001
6	Monthly Expenses	A] Less than 5000 B] 5001-10000 C] 10001-20000 D] 20001-30000 E] More than 30001
7	Smoking and Alcohol	Yes No
8	Pre-existing Diseases	Yes No

Table 2: Insurance transaction data

Sr No	Item	Option
1	Policy type	T] Term plan P] Pension plan MB] Money back plan E] Endowment plan W] Withdrawn plan
2	Term	Less than 3 3-5 years 6-9 years 10-15 years 16-20 years More than 20 years
3	Periodicity	Monthly Quarterly Half yearly Yearly
4	Coverage	Below 100000 Rs. 100001-200000 200001-300000 300001-400000 More than 400000

IV. RESULTS AND DISCUSSIONS

The results for this paper are the rules generated after applying Apriori algorithm on samples. These rules 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.

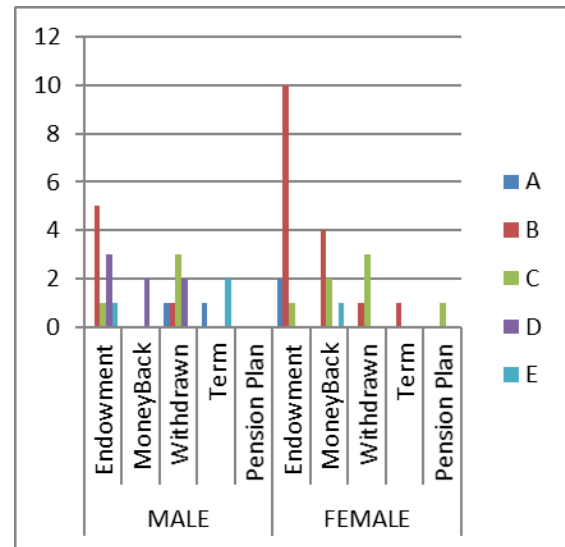


Fig 2: Distribution of policy types according to their Age

From Fig. 1, we can say that the age group of 19 to 30 and 31 to 40 are more interested in LIC for investment. Money Back and Endowment plans are more familiar among all age groups.

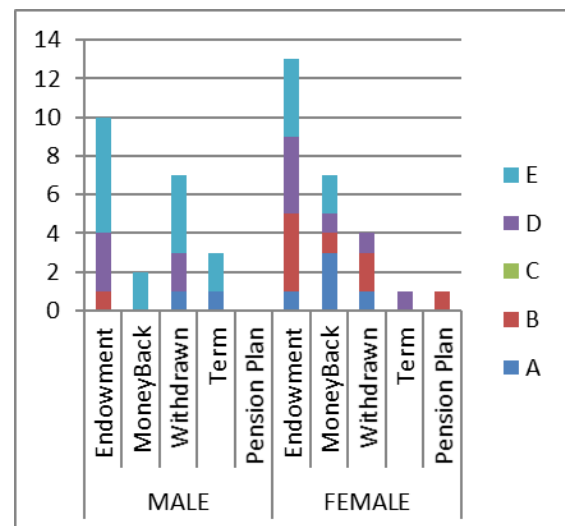


Fig 3: Distribution of Policies According To Their Income

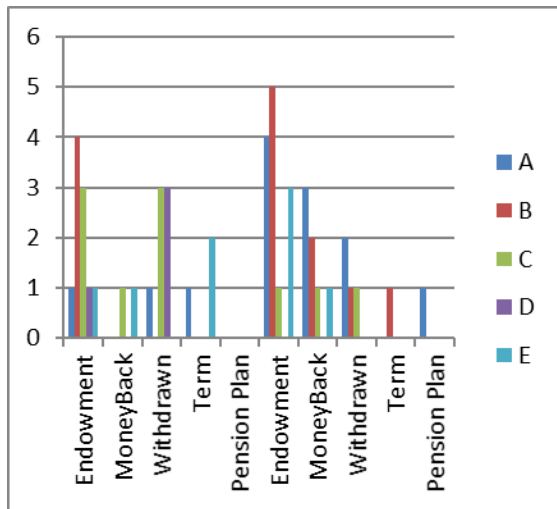


Fig 4: Distribution of Policies According To Their Expenses

From the research we can see that there are more female customers as compared to males which show that females are more awakened than males. Monthly income and expenses both can give the proper understanding to choose the correct policy as they both are independent of each other. This will give a good idea to choose the best suitable plan for investment by considering future safety.

V. CONCLUSION

This research 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.

REFERENCES

- [1] A. B. Devale and Dr. R. V. Kulkarni, "APPLICATIONS OF DATA MINING TECHNIQUES IN LIFE INSURANCE" International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.2, No.4, July 2012.
- [2] V. SreeHari Rao, Murthy V. Jonnalagedda , "Insurance Dynamics – A Data Mining Approach for Customer Retention in Health Care Insurance Industry" Cybernetics And Information Technologies • Volume 12, No 1 Sofia • 2012.
- [3] Kun-Shan Wu, Fang-Kuo Wang and Jhieh-Yu Shyng (2013), "Applied data mining techniques in insurance company: A comparative study of rough sets and decision tree" African Journal of Business Management.
- [4] Abdhesh Gupta, Anwiti Jain (2013), "Life Insurance Recommender System Based on Association Rule Mining and Dual Clustering Method for Solving Cold-Start Problem" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 10, October 2013.
- [5] <https://msdn.microsoft.com/enus/library/ms174949.aspx>
- [6] https://docs.oracle.com/cd/B28359_01/datamine.111/.../process.htm
- [7] http://www.licindia.in/individual_plans.htm
- [8] Introduction to Data mining pang Ningtan, Michigan state University, Michael Steinbach, University of Minnesota, Vipin Kumar, University of Minnesota.
- [9]]Dunham, Margaret H, Data Mining: Introductory and Advanced Topics, Prentice Hall.
- [10] Han and Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann, 2006.
- [11] Thabet Limani, Amor Lazzez, "Efficient Analysis of Pattern and Association Rule Mining Approaches".
- [12] Ms Shweta , Dr. KanwalGarg ,Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms , International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013 ISSN: 2277 128X .