Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products

S. Balaji
Research Scholoar,
Vels University
Chennai,India

S. K. Srivatsa
Phd, Senior Professor,
St.Joseph Eng.College
Chennai,India

ABSTRACT

Prediction analysis is a definite need of any business sector for retaining and attracting the most valuable customers .It is considered as a major challenge facing companies in this information age. Data mining enables companies, in the context of defined business objectives, discover new knowledge, to explore, visualise and understand their data, and to identify patterns, relationships and dependencies that impact on business outcomes.The main focus of this paper concerned with Naive Bayesian classification algorithm for customer classification and prediction on Life Insurance dataset.

Keywords:Data Mining,Naïve Bayesian,customer realationship Management.

1. INTRODUCTION

Data mining is the datadriven extraction of information from such large databases, a process of automated presentation of patterns, rules, and functions to a knowledgeable user for review and examination.(Marisa s.Viveros ,et.al,1996)Data Mining can help life insurance companies to make crucial business decisions and turn the new found knowledge into actionable results in business practices. Insurance firms can increase profitability by identifying the most lucrative customer segments and then prioritize marketing campaigns accordingly. Problems with profitability can occur if life insurance companies do not offer the right policy or the right rate to the right customer segment at the right time. The success of the life insurance profession depends, above all, upon the knowledge and integrity of the people who advise customers – and are their first, and most important point of contact.The term Insurance agents coined as Insurance advisor .An insurance advisor has the unique opportunity to earn the gratitude of people inaddition to highly rewarding one,both in terms of money and in terms of prestige and satisfaction. At the IRDA, the regulator's goal is to see that life insurers are increasingly able to attract, motivate and retain outstanding people, committed to adopting a 'needs-based' approach to financial advice.With DM operations insurance firms can now utilize all of their available information to better develop new products and marketing campaigns. Life insurance industry recorded a premium income of 2,91,605 crore during 2010-11 as against 2,65,447 crore in the previous financial year, registering a growth of 9.85 per cent.Data mining techniques can help insurance companies to guide the potential insurance advisors and customers and to map exact policy for proposal. Data mining enables companies, in the context of defined business objectives, discover new knowledge, to explore, visualise and understand their data, and to identify patterns, relationships and dependencies that impact on business outcomes.

2. LITERATURE SURVEY

Data Mining is a crucial step in the Knowledge Discovery in Database (KDD) process that consists of applying data analysis and knowledge discovery algorithms to produce useful patterns (or rules) over the datasets. Using data mining technology can filtrate and classify customer resources of insurance, divide credit customers into several grades, to predict the customer risk, thus investigating customer material of the low forecasted degrees of comparison can avoid deceiving policy effectively, and avoid service risk. Marisa .S.Viveros[1996]addresses the effectiveness of two data mining techniques in analyzing and retrieving unknown behavior patterns from gigabytes of data collected in the health insurance industry.Chien-Hsing wu,Shu-Vhen Kao and Yann-Yean Su(2005) presented that KDD/Data mining is utilized to explore decision rule to investigate the potential customers for an existing insurance product.Kanwal garg(2008) presented decision tree method for identifying customer behaviour of investment in life insurance sector. Patrick A Rivers(2010) examined some of the benefits and challenges of using data mining processes within the health-care arena. E.W.T.Ngai,L.Xiu and D.C.K.Chau(2009) analysis provides roadmap to guide future research and facilitate knowledge accumulation and creation concerning the application of data mining techniques in CRM.

Zhiyuan yao,Annika H.Holombom,Tomas Eklund and Barbro Back(2010) found that combined approach of SOM-Ward clustering and decision trees provide more detailed information about customer base for tailoring actionable marketing strategies.Abrahams et.al (2009) used decision trees to create a marketing strategy for a pet insurance company.Anna Jurek considered application of Naïve Bayes model for evaluation of risk connected with Life insurance of customers. M.Staudt,J(1998) reports on a project initiated at swiss Life for mining its data resources from the life insurance business.it lies on establishing comfortable data preprocessing support for normalised relational databases and on the management of meta data. Young Moon Cha,et al(2001) examined the characteristics of the knowledge discovery and data mining algorithms to demonstrate how they can be used to predict health outcomes and provide policy information for hypertension management using the Korea Medical Insurance Corporation database. Young Moon Cha,et al(2004) examined characteristics of the mining time dependent patterns to demonstrate how they can be used to predict hypertension outcomes and provide lifestyle information in order to prevent hypertension using data mining approaches.

Table 1.The advantages and limitations of data mining techniques for prediction analysis

Slno	Technique	Advantages	Limitations
1.	Regression	▪ Very rich literature on the use of model, ▪ Ease of application performing model(K.M.Osei-Bryson;2004)	▪ The difficulty of extracting classification rules ▪ the stability of their steady the optimal solution (John Hadden ; Ashutosh Tiwari ;Rajkumar Roy; Dymitr Ruta.2007)
2	Association Rule	Ability to discover hidden relationships among data behavioral the ability to sequence the events,phenomena customer behavior(Ding-An Chiang; Yi-Fan Wang; Shao-Lun Lee; Cheng-Jung Lin.,2003)	Total amount of items that do not frequent (Ding-An Chiang; Yi-Fan Wang; Shao-Lun Lee; Cheng-Jung Lin.2003)
3	Random Forest	▪ stable and steady ▪ The data subject has a good performance (K.W.De Bock and D.Van den Poel,2010)	The difficulty of performing construction(K.W.De Bock and D.Van den Poe,2010)
4	Decision Trees	very simple technique • provide reliable results • provide concrete rules (K.M.Osei-Bryson,2004)	the difficulty of extracting classification rules the stability of their steady the optimal solution(John Hadden ; Ashutosh Tiwari ;Rajkumar Roy; Dymitr Ruta.2007)
5.	Navie Bayes	The number of nominal variables than is the case New Bayes for better performance(Neslin, Gupta, Kamakura, Mason, C. 2006)	New Bayes method for the case of binary characters fare much less accuracy (Neslin, Gupta, Kamakura, Mason, C. 2006)
6.	Clustering	The most widely used method Initial assessment of customer data (Jonathan Burez; Dirk Van den Poel,2007)	Method's performance alone is not sufficient to predict customer behavior [3]

3. METHODOLOGY

Data mining methodology can often improve existing actuarial models by finding additional important variables, by identifying interactions, and by detecting nonlinear elationships.Insurance Market is purely based on customer penetration.Navie Bayes is the basis for many machine learning and data mining methods.In Bayesian classification is a classification method is applicable for huge dataset.Naive Bayesian classifier works with hypothesis H such as that the data tuple X belongs to a specified class C.The determination of P(H/X) that the hypothesis H holds given the evidence or observed data tuple X..P(H/X) is the posterior probability of H conditioned on X.Bayes' theorem is useful in that it provides a way of calculating the posterior probability ,P(H/X) ,from P(X/H) and P(X),

Bayes theorem is

P(H/X)=P(X/H)P(H)/P(X).

3.1 Naive Bayesian Classification Algorithm

The naive Bayesian classifier, or simple Bayesian classifier, works as follows:

1. Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, X=(x1, x2,…, xn), depicting n measurements made on the tuple from n attributes, respectively, A1, A2,.., An.

2. Suppose that there are m classes, C1, C2,…, Cm. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple x belongs to the class Ci if and only if

P (Ci|X)>P (Cj|X) for $1\leq j\leq m$, $j \neq i$

Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

3. As P(X) is constant for all classes, only P (X|Ci) P (Ci) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is, P(C1)=P(C2)

=…=P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class prior probabilities may be estimated by P(Ci)=|Ci,D|/|D|, where |Ci,D| is the number of training tuples of class Ci in D.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute P(X|Ci).

In order to reduce computation in evaluating P(X|Ci), the naïve assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes). Thus,

=P(x1|Ci)x P(x2|Ci)x… P(xm|Ci).

We can easily estimate the probabilities P(x1|Ci), P(x2|Ci),… ,P(xm|Ci) from the training tuples. Recall that here xk refers to the value of attribute Ak for tuple X. For each attribute, we look at whether the attribute is categorical or continuous-valued. For instance, to compute P(X|Ci), we consider the following:

(a) If Ak is categorical, then P(Xk|Ci) is the number of tuples of class Ci in D having the value xk for Ak, divided by |Ci,D|, the number of tuples of class Ci in D.

(b) If Ak is continuous valued, then we need to do a bit more work, but the calculation is pretty straightforward. A continuous-valued attribute is typically assumed to have a Gaussian distribution with a mean μ and standard deviation σ, defined by

$g(x,,\mu,\sigma)=1/\neg 2Д\ \sigma\ e^{-(x-\ \mu)2/2}\ \sigma^2$

So that

P(xk|Ci)=g(xk, μci, σci)

We need to compute μci and σci, which are the mean and standard deviation, of the values of attribute Ak for training tuples of class Ci. We then plug these two quantities into the above equation.

5. In order to predict the class label of X, P(X|Ci)P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of tuple X is the class Ci if and only if

P(X|Ci)P(Ci)>P(X|Cj)P(Cj) for 1 ≤ j ≤ m, j ≠ i

In other words, the predicted class label is the class Ci for which P(X|Ci)P(Ci) is the maximum

Product type A refers to Unit-Linked and B refers to Traditional one.

Policy type refers to A For Savings,B Class refers to Protection plans,C refers to Pension Plans and D for Child plans.

The maximum number of levels in the tree in the tree was limited to four and the minimum number of records in a node was set to 4000,inorder to prevent the Decision Tree from becoming very complex.

The Bayesian approach having the beauty that the probability of the dependent attribute can be estimated by computing estimates of the probabilities of the independent attributes.

There are 10000 (s=10000) samples and four classes.

The frequency of classes as A=3000,B=2000,C=3000,D=2000

The prior probabilities are obtained by dividing these frequencies by the total number in the training data.

P(A)=0.3,P(B)=0.2,P(C)=0.3,P(D)=0.2

Computation of posterior probabilities for the four classes,namely that the customer with attribute values X has Policy preference of Policy type A or type B or type C or type D.

The computation of P(X|Ci)P(Ci) for each of four classes are as

P(A)=0.3,P(B)=0.2,P(C)=0.3,P(D)=0.2 and these values are basis for comparing the four classes.

4. EXPERIMENTS AND ITS RESULTS

4.1.Data Set:-A IRDA Data set of Life Insurance Corporation Of India provided its transaction data for analyses. LIC contributed most of the business procured in this portfolio by garnering `123 crore of ndividual premium from 29.51 lakh policies and 138 crore of group premium

The entire data set covers the period from January 2011-December 2011.The dataset,containing 10,000 customer proposals.The dataset of customers contained customer product type and the plan preference under that category in addition to proposer's risk coverage details.classification of customers based on data for the study. It is known that too many attributes involved will very possibly result in discovered information that is difficult to interpret, or even meaningless. Therefore, by in-depth discussion with domain managers, we eliminated some of the attributes and finally came to a conclusion of 7 attributes, namely 1)Gender 2)age 3)Marital status 4)No of Kids- the customers of marital status married having possible values one or two or three5)Service category :-customer may be a minor or major .Non-minor customer may be in still service or retired 6)Product type – Unit Linked or Traditional product 7) Plan types –Savings plans,Protection plans,Pension plans and Child Plans. For the data set age values may be binned in to the following categories Unmarried in service,Unmarried not in service,Newly Married- in service-without kids,Married-inservice-withchildren,Married not in service,without children ,Married not in service with children. Their preference over the policy is transformed as A,B,C and D.In the above class Not in sevice refres to minor or jobless or retired from service

Predicting a class label using naïve Bayesian classification,we wish to predict the class label of a tupe using naïve Bayesian classification from the sample data .The data tuples are described by the attributes Married,service,age_sub_section ,kids and product type.

The policy type attribute has four distinct attributes namely {A,B,C,D}.

Let C1 coresspond to policy type preference A, C2 coresspond to policy type preference B, C3 coresspond to policy type preference C, C4 coresspond to policy type preference D.

I.The tupe to classify is

X={Married=yes.service=no,kids=YES,agesubsection(<40)=NO,producttype=A }

P(X/A)=1000/3000*3000/3000*3000/3000*2000/3000 *1000/3000
P(X/A)=1.2

P(X/B)=0*0*0*1000/2000*1000/2000=0

P(X/C)=3000/3000* 2000/3000*2000/3000*2000/3000*2000/3000=48/15=3.2

P(X/D)=2000/2000*0*2000/2000*1000/2000*1000/2000=0

P(X/Ci)P(Ci) is being computed to find the class that maximizes Ci,

P(X/Policy type =A)=0.3*1.2=0.36
P(X/Policy type=B)=0.2*0=0
P(X/Policy type=C)=0.3*3.2=0.96
P(X/Policy type=D)=0.2*0=0

Therefore the naïve Bayesian classifer predicts Policy type=C (Pension Plans) for tuple X.
Where as X is observed as Married=yes.service=no,kids=YES,agesubsection(<40)=NO,producttype=A }
Bayes theorem assumes that all attributes are independent and that the sample is good enough to estimate probabilities

II. The tupe to classify is

X={Married=yes.service=no,kids=YES,agesubsection(<40)= NO,producttype=B }

P(X/A)=1000/3000*3000/3000*3000/3000*2000/3000 *2000/3000=36/15=2.4

P(X/B)=0*0*0*1000/2000*1000/2000=0.0

P(X/C)=3000/3000* 2000/3000*2000/3000*2000/3000*1000/3000=24/15=1.6

P(X/D)=2000/2000*0*2000/2000*1000/2000*1000/2000=0

P(X/Ci)P(Ci) is being computed to find the class that maximizes Ci,

P(X/Policy type =A)=0.3*2.4=0.72
P(X/Policy type=B)=0.2*0=0
P(X/Policy type=C)=0.3*1.6=0.48
P(X/Policy type=D)=0.2*0=0

Therefore the naïve Bayesian classifer predicts Policy type=A(Savings Plans) for tuple X.
whichmaximizedCi.ThevalueX is observed as X={Married=yes.service=no,kids=YES,agesubsection(<40)= NO,producttype=B }.

III.The subsequent tupe to classify is

X={Married=NO.service=no,kids=YES,agesubsection(<40)= NO,producttype=A }

P(X/A)=2000/3000*3000/3000*3000/3000*2000/3000 *1000/3000=36/15
P(X/A)=2.4
P(X/B)=2000/3000*0*0*1000/2000*1000/2000=0
P(X/C)=0*2000/3000*2000/3000*2000/3000*2000/3000=48/15=0
P(X/D)=0*0*2000/2000*1000/2000*1000/2000=0

P(X/Ci)P(Ci) is being computed to find the class that maximizes Ci,

P(X/Policy type =A)=0.3*2.4=0.72
P(X/Policy type=B)=0.2*0=0
P(X/Policy type=C)=0.3*0=0
P(X/Policy type=D)=0.2*0=0

Therefore the naïve Bayesian classifer predicts Policy type=A(Savings Plans) for tuple X.
WhereasX is observed as Married=no.service=no,kids=YES,agesubsection(<40)=NO,producttype=A }

IV. The tupe to classify is

X={Married=NO.service=no,kids=YES,agesubsection(<40)= NO,producttype=B }

P(X/A)=2000/3000*3000/3000*3000/3000*2000/3000 *2000/3000=72/15=4.8

P(X/B)=2000/2000*0*0*1000/2000*1000/2000=0.0

P(X/C)=0*2000/3000*2000/3000*2000/3000*1000/3000=0.0

P(X/D)=0*0*2000/2000*1000/2000*1000/2000=0.0

P(X/Ci)P(Ci) is being computed to find the class that maximizes Ci,

P(X/Policy type =A)=0.3*4.8=14.4

P(X/Policy type=B)=0.2*0=0

P(X/Policy type=C)=0.3*0=0

P(X/Policy type=D)=0.2*0=0
Our Bayesian approach based model makes prediction analysis with maximum posteriori hypothesis.
Therefore the Naïve Bayesian classifer predicts Policy type=A (Savings Plans) for tuple X.Whichmaximized Ci.The valueX is observed as X={Married=no.service=no,kids=YES,agesubsection(<40)=NO,producttype=B }.

Bayesian classifier approach to insurance dataset observes customer preference towards the Savings Plans policy type based on attributes characterized mainly based on significant contribution of Marital status attribute of the customer .
The naïve bayesian classifier makes the assumption of class conditional independence,that is given the class label of a tuple ,the values of the attributes are assumed to be conditionally independent of one another.

5. CONCLUSION

In this paper application of classification technique for insurance product preferences towareds customers.Posteriori classification process is applied by looking at the data.Navie bayes classification method used to conduct policy preferences of life insurance customers.Navie bayes classifier is one of the effective classifier in comparison to decision tree and neural network classifier have found it to be comparable to all other classifiers. The result of the analysis demonstrate that Navie bayes can potentially be effective in conducting customer preference analysis over life insurance products. However, this paper observed the KDD/DM application in insurance domain, other issues may also be significant,such as considering other customer attributes (age group wise) policy preferences toward the insurance products.The analysis can be extended with the choice of the optimal combination of features for building classification models.The life style characteristics of the customer can also be considered for policy preference towards insurance products.

6. REFERENCES

[1] Marisa S. Viveros,BM Research Division T. J. Watson (1996) Applying Data Mining Techniques to a Health Insurance Information System, Proceedings of the 22nd VLDB Conference Mumbai (Bombay), India.

[2] Kamakura, Wagner A. & Michel W. (2000). Factor analysis and missing data, Journal of Marketing Research, Vol.37, pp. 490–498.

[3] Berry A.J.Michael ,&Linoff Gordon S.(2000) Mastering Data Mining: The Art and science of Customer Relationship Management ,Wiley,New Jersey

[4] Saundra Glover , Patrick A Rivers , Derek A Asoh , Crystal N Piper and Keva Murph ,(2010) Data mining for health executive decision support: an imperative with a daunting future., Health Services Management Research ,Vol. 23, Number 1 > pp. 42-46

[5] Jiawei Han and Micheline Kamber,Morgan,(2006),Data Mining concepts and Techniques , Kaufmann Publishers.

[6] Billungual report of Insurance Regulatory authority of India -2010-2011

[7] Zhiyuan Yao,Annika H.Holmbom,Tomas Eklund and Barbro Back,(2010),Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis,ICDM.LNAI 6171,pp.292-307.

[8] Khurana Sunayna. (2008)."Customer references in Life Insurance Industry in India", The ICFAIUniversity Journal of Services Vol. 6(3),pp. 61-68.

[9] Chien-Hsing Wu,, Shu-Chen Kao, Yann-Yean Su, Chuan-Chun Wu,(2005) Targeting customers via discovery knowledge for the insurance industry,Vol.29 .pp.291-299. Elsevier Ltd.

[10] Hokey Min,(2006), Developing the Profiles of Supermarket Customers through Data Mining[J],The Service Indusries Journal, Vol.26, No.7, pp.747–763

[11] E.W.T.Ngai,L.XiuandD.C.KChau(2009),Application of data mining techniques in customer relationship management,Expert systems with applicationsVol.36,pp.2592-2602.

[12] Abrahams,A,S.Becker,A,B.Sabido,D.D.Souza,R.Makriyiannis,G.Krasnodebski.,(2009),Inducing a Marketing strategy for a New Pet Insurance Company using Decision trees.Expert systems with applications,Vol.36,pp.1914-1921.

[13] .AnnaJurek,Danutazakrzewska,(2008),Improving Naïve Bayes models of Insurance risk by unsupervised classification,Proceedings of the International Multiconference on Computer Science and Information Technology pp.137-144.

[14] Kanwal garg,Dharminder kumar and m.c.garg,(2008),Data mining techniques for identifying the customer behavior of investment in life insurance sector in india,International journal information technology and knowledge management,Vol.1,pp.51-56.

[15] Yongqiang chen,Leifang Hu,(2005),Study on data mining in CRM system based on Insurance Trade,Proceedings of International conference on Emerging trends in Computers,.pp.839-842.

[16] M.Staudt,J-U.Kietz and U.Reimer,(1998),A data mining support environment and its application on Insurance data,American Journal for Aritifical Intelligence,pp.15-19.

[17] Young Moon Chae , Seung Hee Ho , Kyoung Won Cho , Dong Ha Lee and Sun Ha Ji,(2001),Data mining approach to policy analysis in a health insurance domain ,International Journal of Medical Informatics,vol. 62 ,pp.103–111.

[18] Young Moon Chaea, Dong Ha Leeb, Hwa Young Kima, Tae Soo Kima , Tae Min Songc(2004),Mining time dependency patterns in a Health Insurance Domain,International journal of Medical informatics,Vol.99.pp.1545-1555.

[19] Long WJ, Griffith JL, Selker HP, and D'Agostino RB.(1993), A comparison of logistic regression to decision-tree induction in a medical domain. Comp and Biomed Res Vol 26: 74-97.