

# Prediction of Online Vehicle Insurance System using Bayes Classifier – A Proposed Approach

S. S. Thakur

Department of Computer Science & Engineering  
MCKV Institute of Engineering, Liluah  
Howrah – 711204

J. K. Sing

Department of Computer Science & Engineering  
Jadavpur University  
Kolkata – 700032

## ABSTRACT

A classification technique (or classifier) is a systematic approach used in building classification models from an input data set. Some examples include decision tree classifier, rule based classifiers, neural networks, support vector machines and naïve Bayes classifiers. Each technique employs a learning algorithm to identify a model that best fits the relationships between the attribute set and the class label of the input data. The model generated by the learning algorithm should both fit the input data well and correctly predict the class labels of records it has never seen before. Therefore, a key objective of the learning algorithm is to build models that accurately predict the class labels of previously unknown records. In this work prediction has been done for the entire test set data, and accuracy is found to be good, compared to the relative performance of different classifiers namely decision tree classifier.

## General Terms

Data Mining

## Keywords

Bayes' classifier, Web services, Conditional Probability, Posterior Probability, Online Insurance etc.

## 1. INTRODUCTION

Classification systems were developed earlier to organize a large collection of objects. Automated classification has been a subject of intensive research for many years. Many pattern recognition problems also require the discrimination of objects from different classes. Examples include speech recognition, handwritten character identification, and image classification. The subject of classification is also a major research topic in the field of neural networks, statistical learning, and machine learning. An in-depth treatment of various classification techniques is given in the book by Cherkassky and Mulier [1].

In many applications the relationship between the attribute set and the class variable is non-deterministic. In other words, the class label of a test record cannot be predicted with certainty even though its attribute set is identical to some of the training examples. The situation may arise due to noisy data or the presence of certain confounding factors that affect classification but are not included in the analysis. For example, consider the task of predicting whether a person is at risk for heart disease based on person's diet, and workout frequency. Although most people who eat healthily and exercise regularly have less chances of developing heart diseases, they may still do so because of other factors such as heredity, excessive smoking, and alcohol abuse. Determining whether a person's diet is healthy or the workout frequency is

sufficient is also subject to interpretation, which in turn may introduce uncertainties into the learning problem.

This paper outlines the implementation of Bayes classifier using Bayes theorem for prediction of online vehicle Insurance system.

## 2. BAYES THEOREM PRELIMINARIES

In probability theory Bayes' theorem (often called Bayes' Law) relates the conditional and marginal probabilities of two random events [2] [3]. It is often used to compute posterior probabilities given observations. For example, a patient may be observed to have certain symptoms. Bayes' theorem can be used to compute the probability that a proposed diagnosis is correct, given that observation. As a formal theorem, Bayes' theorem is valid in all interpretations of probability. However, it plays a central role in the debate around the foundations of statistics [4] [5] frequentist and Bayesian interpretations disagree about the ways in which probabilities should be assigned in applications. Frequentists assign probabilities to random events according to their frequencies of occurrence or to subsets of populations as proportions of the whole, while Bayesian describe probabilities in terms of beliefs and degrees of uncertainty.

Let  $X$  and  $Y$  be a pair of random variables. Their joint probability,  $P(X = x, Y = y)$ , refers to the probability that variable  $X$  will take on the value  $x$  and variable  $Y$  will take on the value  $y$ . A conditional probability is the probability that a random variable will take on a particular value given that the outcome from another random variable is known. For example, the conditional probability  $P(Y = y | X = x)$  refers to the probability that a variable  $Y$  will take on the value  $y$ , given that the variable  $X$  is observed to have the value  $x$ . The joint and conditional probabilities for  $X$  and  $Y$  are related in the following manner as shown below.

$$P(X, Y) = P(Y | X) \times P(X) = P(X | Y) \times P(Y) \quad (1)$$

Rearranging the last two expressions leads to the following formula, as shown in equation below:

$$P(Y | X) = P(X | Y)P(Y)/P(X)$$

The above formula is referred to as Bayes' theorem, and can be used to solve the prediction problem.

## 3. PROPOSED APPROACH

Prediction of Online Vehicle Insurance system emphasizes some key areas. These are

- Introduction to Bayes Theorem has been dealt in section 2.
- Proposed Approach has been dealt in section 3 and Algorithm for Online Vehicle Insurance

system has been dealt in section 4 and Bayes' Classifier in section 5.

- Implementation Methodology has been dealt in section 6
- Experimental evaluation and Results has been dealt in section 7.
- Discussion & Conclusion has been dealt in section 8.
- Acknowledgement & References has been dealt in section 9 and 10.

The block diagram for the Prediction of Online Vehicle Insurance system is shown in Figure. 1.

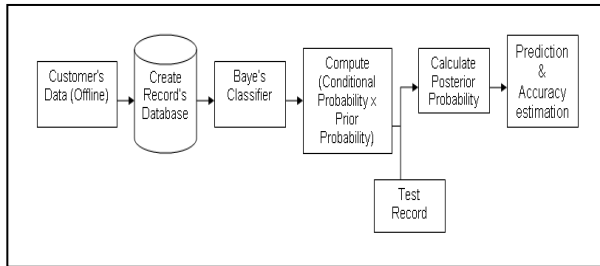


Fig 1: Block diagram – Proposed Prediction System

#### 4. ALGORITHM - ONLINE VEHICLE INSURANCE SYSTEM

A skeleton classifier algorithm called Bayes' Classifier [6] [7] is shown in Algorithm 4.1. The input to this algorithm consists of the records E and the attribute set F.

##### 4.1 Algorithm

- 1: Collect data from customers.
- 2: Create training records database.
- 3: Apply Bayes theorem, to express the posterior probability in terms of prior probability  $P(Y)$ , the **class-conditional** probability  $P(X/Y)$ , and the evidence  $P(X)$
4. The class-conditional probability  $P(X/Y)$ , implementation using naïve Bayes Classifier.
5. Estimate class-conditional probability, assuming that the attributes are conditionally independent.
6. To classify a **test record**, compute the posterior probability for each class  $Y$ , using naïve Bayes Classifier.
7. Estimate the conditional probability  $P(X_i/Y)$  for categorical and continuous attributes.
8. End

#### 5. BAYES CLASSIFIER PRELIMINARIES

Before describing how the Bayes' theorem can be used for classification [8] [9] let us formalize the classification problem from a statistical perspective. Let  $X$  denotes the attribute set and  $Y$  denote the class variable. If the class variable has the non-deterministic relationship with the attributes, then we can treat  $X$  and  $Y$  as random variable and capture their relationship probabilistically using  $P(Y/X)$ . The conditional probability is also known as the posterior probability for  $Y$ , as opposed to its prior probability,  $P(Y)$ .

During the training phase, we need to learn the posterior probabilities  $P(Y/X)$  for every combination of  $X$  and  $Y$  based on information gathered from the training data. By knowing these probabilities, a test record  $X'$  can be classified by finding the class  $Y'$  that maximizes the posterior probability,  $P(Y'/X')$ . To illustrate this approach, consider the task of predicting whether a customer i.e. a vehicle owner will go for online insurance. Table 1 shows a training set with the

following attributes: Vehicle owner, Qualification, and Age. Customers interested for Online insurance are classified as Yes, while others are classified as No.

Table 1. Training Dataset

	Binary	Categorical	Continuos	Class
Tid	Vehicle Owner	Educational Qualification	Age	Online - Insurance
1	Yes	Higher Secondary	70	Yes
2	No	Graduate	60	Yes
3	No	Higher Secondary	30	Yes
4	Yes	Graduate	65	Yes
5	No	Post Graduate	55	No
6	No	Graduate	20	Yes
7	Yes	Post Graduate	70	Yes
8	No	Higher Secondary	40	No
9	No	Graduate	35	Yes
10	No	Higher Secondary	45	No

Suppose we are given a test record with the following attribute set:  $X = (\text{Vehicle owner} = \text{No}, \text{Qualification} = \text{Graduate}, \text{Age} = 60 \text{ years})$ . To classify the record, we need to compute the posterior probabilities  $P(\text{Yes}|X)$  and  $P(\text{No}|X)$  based on information available in the training data. If  $P(\text{Yes}|X) > P(\text{No}|X)$ , then the record is classified as Yes, otherwise it is classified as No, as shown in Table 2.

Table 2. Test Dataset

Tid	Vehicle Owner	Educational Qualification	Age	Online - Insurance
1	No	HS	15	?
2	Yes	Graduate	37	?
3	Yes	Post Graduate	62	?
4	No	Graduate	57	?
5	No	Post Graduate	25	?

Estimating the posterior probabilities accurately for every possible combination of class label and attribute value is a difficult problem because it requires a very large training set, even for a moderate number of attributes. The Bayes theorem is useful because it allows us to express the posterior probability in terms of prior probability  $P(Y)$ , the class-conditional probability  $P(X/Y)$ , and the evidence,  $P(X)$  can be computed as follows:

$$P(Y|X) = P(X|Y)P(Y)/P(X) \quad (2)$$

When comparing the posterior probabilities for different values of  $Y$ , the denominator term,  $P(X)$ , is always constant, and thus can be ignored. The prior probability  $P(Y)$  can be easily estimated from the training set by computing the fraction of training records that belongs to each class. To estimate the class-conditional  $P(X/Y)$ , we are using naïve Bayes Classifier in our proposed work.

Naïve Bayes Classifier: A naïve Bayes classifier [8] estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label  $y$ . The conditional independence can be formally stated as follows:

$$P(X | Y = y) = \prod_{i=1}^d P(X_i | Y = y) \quad (3)$$

Where each attribute set  $X = \{X_1, X_2, \dots, X_d\}$  consists of  $d$  attributes.

Characteristics - Naïve Bayes Classifier: Some of the characteristics have been discussed below.

- They are robust to isolated noise points because such points are averaged out when estimating conditional probabilities from data. Naïve Bayes' classifiers can also handle missing values by ignoring the example during model building and classification.
- They are robust to irrelevant attributes. If  $X_i$  is an irrelevant attribute, then  $P(X_i|Y)$  becomes almost uniformly distributed. The class-conditional probability for  $X_i$  has no impact on the overall computation of the posterior probability.
- Correlated attributes can degrade the performance of Naïve Bayes' classifiers [10] because the conditional independence assumption no longer holds for such attributes.

## 6. IMPLEMENTATION METHODOLOGY - WORKING PRINCIPLE

To illustrate how the algorithm works, consider the problem of predicting whether a customer will go for manual insurance or online insurance. A training set for this problem has been constructed by data collection through customers at different locations in our city. In the example shown in Table 2, each record contains the personal information of a customer, along with a class label indicating whether the customer has shown interest for manual insurance.

With the conditional independence assumption, instead of computing the class-conditional probability for every combination of  $X$ , we only have to estimate the condition probability of each  $X_i$ , given  $Y$ . This approach is more practical because it does not require a very large training set to obtain a good estimate of the probability. To classify a test record, the naïve Bayes classifier [11] [12], computes the posterior probability for each class  $Y$ :

$$P(Y | X) = P(Y) \prod_{i=1}^d P(X_i | Y) / P(X) \quad (4)$$

Since  $P(X)$  is fixed for every  $Y$ , it is sufficient to choose the class that maximizes the numerator term,

$$P(Y) \prod_{i=1}^d P(X_i | Y) / P(X)$$

We describe some approaches for estimating the conditional probabilities  $P(X_i|Y)$  for categorical and continuous attributes.

### 6.1 Conditional Probability Estimation - Categorical Attributes

For a categorical attribute  $X_i$ , the conditional probability  $P(X_i = x_i | Y = y)$  is estimated according to the fraction of training [13] [14] instances in class  $y$  that take on a particular attribute value  $x_i$ . For example, in the training set given in Figure 2, three out of seven people who take online insurance also own a vehicle. As a result, the conditional probability for

$P(\text{Vehicle Owner} = \text{Yes} | \text{Yes})$  is equal to  $3/7$ . Similarly, the condition probability for defaulted borrowers who are single is given by  $P(\text{Qualification} = \text{Graduate} | \text{Yes}) = 2/3$ .

### 6.2 Conditional Probability Estimation - Continuous Attributes

There are two ways to estimate the class-conditional probabilities for continuous attributes in naïve Bayes classifiers:

1. We can discretize each continuous attribute and then replace the continuous attribute value with its corresponding discrete interval. This approach transforms the continuous attributes into ordinal attributes. The conditional probability  $P(X_i | Y = y)$  is estimated by computing corresponding interval for  $X_i$ . The estimation error depends on the discretization strategy, as well as the number of discrete intervals. If the number of intervals is too large, there are too few training records in each interval to provide a reliable estimate for  $P(X_i | Y)$ . On the other hand, if the number of intervals is too small, then some intervals may aggregate records from different classes and we may miss the correct decision boundary.
2. We may assume a certain form of probability distribution for the continuous variable and estimate the parameters of the distribution using the training data [15]. A Gaussian distribution is usually chosen to represent the class – conditional probability for continuous attributes. The distribution is characterized by two parameters, its mean  $\mu$  and variance  $\sigma^2$ . For each class  $y_j$ , the class-conditional probability for attribute  $X_i$  is given below.

$$P(x = v | c) = \frac{1}{\sqrt{2\pi\sigma^2c}} e^{-\frac{(v-\mu_c)^2}{2\sigma^2c}} \quad (5)$$

## 7. EXPERIMENTAL EVALUATION & RESULTS

The parameter  $\mu_{ij}$  can be estimated based on the sample mean of  $X_i(x)$  for all training records that belong to the class  $y_j$ . Similarly,  $\sigma_{ij}^2$  can be estimated from the simple variance ( $s^2$ ) of such training records. For example consider the age attribute as shown in Table 2. The sample mean and variance for this attribute with respect to the class Yes are given below.

Class:  $P(C) = N_c / N$

e.g.,  $P(\text{Yes}) = 7/10$

$P(\text{No}) = 3/10$

For discrete attributes:

$$P(A_i | C_k) = |A_{ik}| / N_c$$

where  $|A_{ik}|$  is number of instances having attribute  $A_i$  and belongs to class  $C_k$

e.g.,  $P(\text{Qualification}=\text{Graduate}|\text{Yes}) = 4/7$

$$P(\text{Owner}=\text{Yes/No})=0$$

For (Age, Class=Yes):

For Class=Yes, Sample mean = 50

$$\text{Sample variance} = ((70-50)^2 + (60-50)^2 + (30-50)^2 + (65-50)^2 + (20-50)^2 + (70-50)^2 + (35-50)^2) / (6)$$

$$= (20)^2 + (10)^2 + (-20)^2 + (15)^2 + (-30)^2 + (20)^2 + (-15)^2 / (6) = (400+100+400+225+900+400+225)/6$$

$$= 2650/6 = 441.66 = 21.00$$

$$\text{Sample variance} = 441.66$$

$$P(\text{Age}=65|\text{Yes}) = 0.024$$

For Class=No, Sample mean = 46.66

$$\text{Sample variance} = ((55-46.66)^2 + (40-46.66)^2 + (45-46.66)^2 + ((8.33)^2 + (-6.66)^2 + (-1.66)^2) / (2) = 58.24$$

$$\text{Sample variance} = 58.24 = 7.63$$

$$P(\text{Age}=65|\text{No}) = 0.940$$

The prior probabilities of each class can be estimated by calculating the fraction of training records that belong to each class. Since there are three records that belongs to the class Yes, and seven records that belongs to the class No,  $P(\text{Yes}) = 0.3$  and  $P(\text{No}) = 0.7$ .

$X = (\text{Owner}=\text{No}, \text{Graduate}, \text{Age}=65\text{years})$

$$P(\text{VOwner}=\text{Yes}|\text{No}) = 0$$

$$P(\text{VOwner}=\text{No}|\text{No}) = 1$$

$$P(\text{VOwner}=\text{Yes}|\text{Yes}) = 3/7$$

$$P(\text{VOwner}=\text{No}|\text{Yes}) = 4/7$$

$$P(\text{Qualification}=\text{HS}|\text{No}) = 2/3$$

$$P(\text{Qualification}=\text{Grad}|\text{No}) = 0$$

$$P(\text{Qualification}=\text{PGrad}|\text{No}) = 1/3$$

$$P(\text{Qualification}=\text{HS}|\text{Yes}) = 2/7$$

$$P(\text{Qualification}=\text{Grad}|\text{Yes}) = 4/7$$

$$P(\text{Qualification}=\text{PGrad}|\text{Yes}) = 1/7$$

For Age:

If Class = Yes: Sample mean = 50

$$\text{Sample variance} = 441.66$$

If Class = No: Sample mean = 46.66

$$\text{Sample variance} = 58.24$$

To predict the class label of a test record  $X = (\text{Vehicle owner}=\text{No}, \text{Qualification}=\text{Graduate}, \text{Age}=65)$ , we need to compute the posterior probabilities  $P(\text{No}|X)$  and  $P(\text{Yes}|X)$ . These posterior probabilities can be estimated by computing

the product between the prior probability  $P(Y)$  and the class-conditional probabilities.

$$P(X|\text{Class}=\text{Yes}) = P(\text{Owner}=\text{No}|\text{Class}=\text{Yes}) \times P(\text{Graduate}|\text{Class}=\text{Yes}) \times P(\text{Age}=65\text{y}|\text{Class}=\text{Yes})$$

$$= 4/7 \times 4/7 \times 0.024 = 0.0078$$

$$P(X|\text{Class}=\text{No}) = P(\text{Owner}=\text{No}|\text{Class}=\text{No}) \times P(\text{Graduate}|\text{Class}=\text{No}) \times P(\text{Age}=65\text{y}|\text{Class}=\text{No})$$

$$= 1 \times 0 \times 0.940 = 0$$

$$P(X|\text{Class}=\text{Yes}) > P(X|\text{Class}=\text{No})$$

Therefore  $P(\text{Yes}|X) > P(\text{No}|X) \Rightarrow \text{Class} = \text{Yes}$

The results obtained after applying Bayes' Classifier to the Test Dataset is shown in Table 3.

**Table 3. Results - Test Dataset**

Tid	Vehicle Owner	Educational Qualification	Age	Online - Insurance
1	No	HS	15	No
2	Yes	Graduate	37	Yes
3	Yes	Post Graduate	62	Yes
4	No	Graduate	57	Yes
5	No	Post Graduate	25	No

Thus the Online Vehicle Insurance system will help the customers, who logs into the system to see the Prediction results.

## 8. DISCUSSION & CONCLUSION

The developed system overcomes the limitations of the manual system, as all the information related to Prediction of Vehicle insurance are available online. To access the developed system the Customer requires a PC and an Internet connection. The system is developed by making use of available tools, techniques and resources that could generate a Secure Web Service. This system is user friendly, cost effective and flexible and the performance of the system is found to be satisfactory.

Compared to client-server environments, a Web service is much more dynamic and secured for such an environment poses unique challenges. In conventional systems user identity is known in advance and can be used for performing access control. In open systems participants may not have any pre-existing relationship and share a common security domain. An important issue is development of access control models, Subject Authentication and Parameter Negotiations are required to restrict access to Web services to authorized users only. In future we will be extending our proposed work, using these features as mentioned above, to develop an online system for Accidental Claim and repair of Damaged Vehicles, using Secure Web Service Negotiation.

## 9. ACKNOWLEDGMENTS

The authors are thankful to Mr. Ashutosh kumar Jha, Mr. Amlan Mondal, Mr. Rakesh kumar Singh, students of Final Year, CSE Deptt, of MCKV Institute of Engineering, Liluah for their involvement in data collection for the said Research work. The authors are also thankful to Prof. Parasar Bandyopadhyay, Principal, MCKVIE, Liluah for giving permission to use the labs. for carrying out the research work.

## 10. REFERENCES

- [1] Cherkassky, V. and Mulier, F. Learning from Data: Concepts, Theory, and Methods. Wiley Interscience, 1998.
- [2] Han, J. and Kamber, M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco, 2001
- [3] Hand, D. J., Mannila, H., and Smyth, P. Principles of Data Mining. MIT Press, 2001
- [4] Hastie, T., Tibshirani, R., and Friedman. J. H. The Elements of Statistical Learning: Data Mining, Inference, Prediction. Springer, New York, 2001
- [5] Hastie, T. and Tibshirani, R. Classification by pairwise coupling. *Annals of Statistics*, 26(2):451-471, 1998.
- [6] Domingos, P. MetaCost: A General Method for Making Classifiers Cost-Sensitive. In Proc. of the 5<sup>th</sup> Intl. Conf. on Knowledge Discovery and Data Mining, pages 155-164, San Diego, CA, August 1999.
- [7] Domingos, P. and Pazzani. M. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*, 29(2-3):103-130, 1997.
- [8] Ramoni, M. and Sebastiani, P. Robust Bayes classifiers. *Artificial Intelligence*, 125:209-226, 2001.
- [9] Langley, P., Iba, W., and Thompson, K. An analysis of Bayesian classifiers. In Proc. of the 10<sup>th</sup> National Conf. on Artificial Intelligence, pages 223-228, 1992.
- [10] Heckerman, D. Bayesian Networks for Data Mining. *Data Mining and knowledge Discovery*, 1(1):79-119, 1997.
- [11] Holte, R. C. Very Simple Classification Rules Perform Well on Most Commonly Used Data sets. *Machine Learning*, 11:63-91, 1993.
- [12] Japkowicz, N. The Class Imbalance Problem: Significance and Strategies. In Proc. of the 2000 Intl. Conf. on Artificial Intelligence: Special Track on Inductive Learning, volume 1, pages 111 -117, Las Vegas, NV, June 2000.
- [13] Joshi, M. V. On Evaluating performance of Classifiers for Rare Classes. In Proc. of the 2002 IEEE Intl. Conf. on Data Mining, Maebashi City, Japan, December 2002.
- [14] Joshi, M. V., Agarwal, R.C., and Kumar, V. Mining Needles in a Haystack: Classifying Rare Classes via Two-Phase Rule Induction. In Proc. of 2001 ACM-SIGMOD Intl. Cong. on Management of Data, pages 91-102, Santa Barbara, CA, June 2001.
- [15] Kubat, M. and Matwin, S. Addressing the Curse of Imbalanced Training Sets: One Sided Selection. In Proc. of the 14<sup>th</sup> Intl. Conf. on Machine Learning, pages 179-186, Nashville, TN, July, 1997.