# Sale Forecasting for Hot-Drink Productivity Using Naïve Bayesian Classification

Su Su Swe Hlaing, Thaung Myint Htun
*Computer University (Sittway)*
*nayniswe@gmail.com* , *thaungmyinthtun.cu@gmail.com*

## Abstract

*Classification is one of the most popular data mining tasks with a wide ranges of application and lots of algorithms have been proposed to build scalable classifiers. Several data mining techniques and classification methods have been widely applied to extract knowledge from databases. Naïve Bayes is one of the most efficient and effective inductive learning algorithms for machine learning and data mining. Its competitive performance in classification is surprising, because the conditional independence assumption on which it the conditional independence assumption on which it is based, rarely true in real-world applications. This system will present sale forecasting productivity using Naïve Bayesian Classification.*

Keywords: data mining, classification, forecasting, productivity.

## 1. Introduction

Data mining is the process of analyzing large datasets in order to find patterns that can help to isolate key variables to build predictive models for management forecasting .[2].

Data mining is an aid to strategic, tactical and operational decision making in situations where numerous variables, affecting costs or benefits, impinge on the eventual outcome of the course of action that a company might decide to take. The modeling that accompanies data mining assimilates the information on costs and benefits of alternative course of action as visualized in the form of familiar with prediction. Organizations use such information to find new opportunities for growth, choose more effective means to achieve their goals.

Data mining refers to extracting or mining knowledge from large amount of data either in database or data warehouse or other information repositories. Data mining is a step in the knowledge discovery process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of pattern over the data.

## 2. Related Work

Using Naïve Bayes improves the tasks of the mining. Due to its perceived limitations, the simple Bayesian classifier has traditional not been a focus of research not been a focus of research in machine learning. Spiegel Halter and Knill Jones (1984) described a greatly simplified, yet elegant method for combing evidence to make a prediction based on the Naïve Bayes classifier. John and Langley (1995) showed that the Bayesian classifier's performance can be much improved if the traditional treatment of numeric attributes, which assumes Gaussian distributions. The naïve Bayes classifier is used as a competitive classifier due to its robustness and simplicity. See P.Domingo's and M.Pazzani[7], Khin Myo Aye, "Prediction Heart Disease Using Naïve Bayesian Classification".[5], Thu Thu Zan, "Treatment Guidelines for Diabetic Patient By Using Naïve Bayesian Classification".[9], Witt Yee Win, "Classification of Nutritional Status by Using Naïve Bayesian Classification".[8].

## 3. Classification

Classification problems aim to identify the characteristics that indicate the group to which each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. For example, we may want to predict whether individuals can be classified as likely to respond to a direct mail solicitation, vulnerable to switching over to a competing long distance phone service, or a good candidate for a surgical procedure. Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. These existing cases may come from an historical database, such as people who have already undergone a particular medical treatment or moved to a new long distance service. They may come from an experiment in which a sample of the entire database is tested in the real world and the results used to create a classifier. For example, a sample of a mailing list would be sent an offer, and the results of the mailing used to develop a classification model to be applied to the entire database. Sometimes an expert classifies a sample of the database, and this classification is then used to create the model which will be applied to the entire database.

Classification is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown.

### 3.1. Bayesian Classification

Bayesian classification is based on Bayes theorem. Bayesian classifiers have exhibited high accuracy and speed when applied to large databases.

Naïve Bayesian classifier assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence.

Naive Bayes Classification (NBC) is a machine learning method, particularly popular in prediction as well as sale forecasting applications. NBC assumes that the attributes are mutually independent. Although in practice this assumption is not quite true, experience shows that the NBC approach in forecasting application is effective more elaborate learning methods. The relative strength of these approaches comes precisely from the fact that they assume attribute independence, even when the assumption is not completely true. The independence assumption licenses the classifier to collect the evidences about the class from individual attributes separately. So an attribute's contribution of evidence about the class is determined independently from other attributes.

NBC is a highly practical Bayesian learning method. Naïve Bayes classifiers are simple, effective, efficient, robust, and support incremental training. Naïve Bayesian classifiers have proven to be powerful tools for solving classification problems in a variety of domains. Naïve Bayesian classifiers have been successfully applied. Naïve Bayes is based on a Bayesian formulation of the classification problem which uses the simplifying assumption of attribute independence. It is simple to implement and use while giving surprisingly good results.

## 3.2. Bayes Theorem

Let X be a data sample whose class label is unknown. Let H be some hypothesis, such as that the data sample X belongs to a specified class C. For classification problems, the system wants to determine P(H\X), the probability that the hypothesis H holds given the observed data sample X. P(H\X) is the posterior probability, of a posteriori probability, of H conditioned on X. P(H) is the prior probability, or a priori probability of H. The posterior probability, P(X\H) , is based on more information than the priori probability, P(H), which is independent of X. Similarity, P(X/H) is the posterior probability of X conditioned on H. P(X) is the priori probability of X. Bayes theorem is useful that it provides a way of calculating the posterior probability, P(H\X), from P(H), and P(X/H). Bayes theorem is

$$P(X \setminus H) = \frac{P(X \setminus H)P(H)}{P(X)} \; .$$

## 3.3. Naïve Bayes Classifier

The Naïve Bayesian classifier, or Simple Bayesian classifier works as follows:

Each data sample is represented by an n-dimensional feature vector, X= $(x_1, x_2,\ldots, x_n)$, depicting n measurements made on the sample from n attributes, respectively, $A_1, A_2, \ldots., A_n$.

Suppose that there are m class, $C_1, C_2,\ldots, C_m$. Given an unknown data sample, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. The Naïve Bayesian classifier assigns an unknown sample X to the class $C_i$, if an only if

$$P(C_i \setminus X) > P(C_j \setminus X) \text{ for } 1 <= j <= m, j != i.$$

The class $C_i$, for which $P(C_i \setminus X)$ is maximized is called the maximum posteriori hypothesis.

$$P(C_i \setminus X) = \frac{P(X \setminus C_i)P(C_i)}{P(X)}$$

As P(X) is constant for all classes, only $P(X \setminus C_i)$ $P(C_i)$ need to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is , $P(C_1)= P(C_2)=\ldots\ldots= P(C_m)$, and the system would therefore maximize $P(X \setminus C_i)$. Otherwise, the system maximize $P(X \setminus C_i)$ $P(C_i)$. The class prior probabilities may be estimated by $P(C_i)= s_i/ s$, where $s_i$ is the number of training samples of class $C_i$ , and s is the total number of training samples.

Given data sets with many attributes, it would be extremely computationally expensive to compute $P(X \setminus C_i)$, the naïve assumption of class conditional independence is made. There are no dependence relationships among the attributes.

$$P(X \mid C_i) = \prod_{k=1}^{n} P(s_k \mid C_i)$$

The probabilities $P(X_1| C_i)$, $P(X_2| C_i)$,… $P(X_n| C_i)$ can be estimated from the training samples.

$$P(x_k \mid C_i) = g(x_k, \mu c_i, \sigma c_i) = \frac{1}{\sqrt{2\Pi \sigma c_i}} e^{-\frac{(x_k - \mu c_i)}{2\sigma_{c_i}^2}}$$

In order to classify an unknown sample X, $P(X|C_i)\ P(C_i)$ is evaluated for each class $C_i$ . Sample X is assigned to the class $C_i$ if and only if

$$P(X \mid C_i)P(C_i) > P(X \mid Cj)P(C_j)$$

for $1 \leq j \leq m$, $j \neq i$.

In other words, it is assigned to the class $C_i$ for which $P(X|C_i)\ P(C_i)$ is the maximum.

Bayesian classifiers have the minimum error rate in comparison to all other classifiers. Bayesian classifiers are also useful.

## 4. Flow of the System

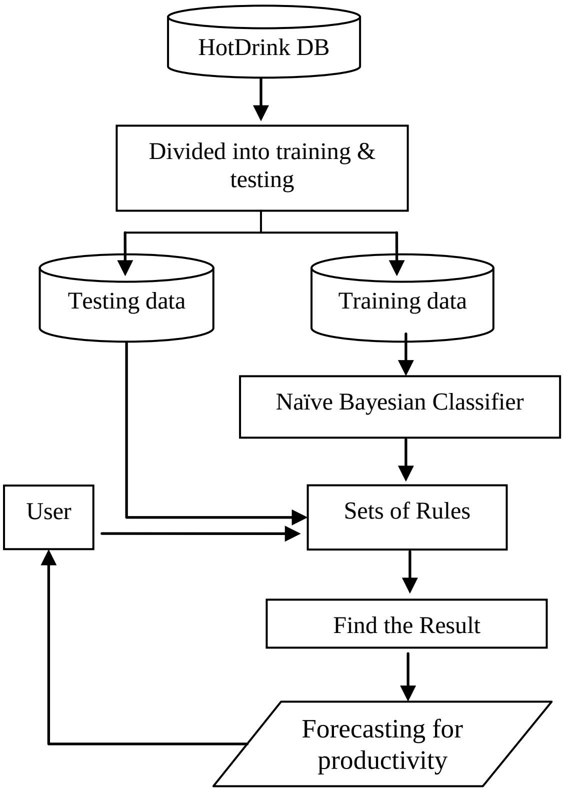


**Figure 1: System flow Diagram**

The system tends to forecast for hot-drink of sale for productivity using Naïve Bayesian Classification. The proposed system of sale forecasting for hot-drink productivity will allow the user to get a quick forecasting results for several beverages of hot-drinks and gives the marketers effective productive advice.

The system use training data to generate rules with Naïve Bayesian Classifier. Training gives the set of rules to testing. Training shows the system's information to user as figure 2. User gives new data to the system. The system sends this data to testing. Then, testing returns result to the system. Finally, system forecasts the sale result and the forecasting result of productivity for beverages of hot-drinks as near future result to the users as figure 3. Figure 1 shows the steps of flow of the proposed system.

## 5. System Implementation

After the user entered the sales data for each product, customer preferences by monthly, yearly and product types can be retrieved.


Training Data Set-PremierCM

| TNo | PurcQty | PurcRate | SaleQty | SaleRate | Balance | Debt | Handover | Advertise | Forecast |
|---|---|---|---|---|---|---|---|---|---|
| t1 | 600 | 2400 | 400 | 2500 | more | no | yes | no | decline |
| t10 | 600 | 2400 | 492 | 2500 | nearlyzero | yes | no | yes | normal |
| t11 | 600 | 2400 | 378 | 2500 | more | yes | yes | no | decline |
| t12 | 600 | 2400 | 384 | 2500 | more | yes | yes | no | decline |
| t13 | 800 | 2500 | 680 | 2600 | more | no | no | yes | normal |
| t14 | 800 | 2500 | 687 | 2600 | nearlyzero | no | no | yes | normal |
| t15 | 800 | 2500 | 740 | 2600 | nearlyzero | no | no | yes | high |
| t16 | 800 | 2500 | 760 | 2600 | nearlyzero | no | no | yes | high |
| t17 | 800 | 2500 | 686 | 2600 | more | yes | no | yes | normal |
| t18 | 800 | 2500 | 490 | 2600 | more | yes | yes | no | decline |
| t19 | 800 | 2500 | 488 | 2600 | more | yes | yes | no | decline |
| t2 | 600 | 2400 | 450 | 2500 | nearlyzero | no | yes | yes | normal |
| t20 | 800 | 2500 | 687 | 2600 | nearlyzero | no | no | yes | normal |
| t21 | 800 | 2500 | 700 | 2600 | nearlyzero | no | no | yes | normal |
| t22 | 800 | 2500 | 774 | 2600 | nearlyzero | no | no | yes | high |
| t23 | 800 | 2500 | 730 | 2600 | nearlyzero | yes | no | yes | normal |
| t24 | 800 | 2500 | 738 | 2600 | nearlyzero | no | no | no | normal |
| t25 | 900 | 2600 | 756 | 2700 | nearlyzero | no | no | yes | normal |
| t26 | 1000 | 2600 | 920 | 2700 | nearlyzero | no | no | yes | high |
| t27 | 1000 | 2600 | 762 | 2700 | more | yes | no | yes | normal |
| t28 | 1000 | 2600 | 953 | 2700 | nearlyzero | no | no | yes | high |
| t3 | 600 | 2400 | 575 | 2500 | nearlyzero | yes | yes | yes | high |
| t30 | 1000 | 2600 | 768 | 2700 | more | yes | no | yes | normal |
| t31 | 1000 | 2600 | 974 | 2700 | nearlyzero | no | no | yes | high |
| t32 | 1000 | 2600 | 980 | 2700 | nearlyzero | no | no | yes | high |
| t4 | 600 | 2400 | 465 | 2500 | nearlyzero | no | no | yes | normal |
| t5 | 600 | 2400 | 392 | 2500 | more | no | no | no | decline |
| t6 | 600 | 2400 | 463 | 2500 | nearlyzero | no | yes | yes | normal |
| t7 | 600 | 2400 | 465 | 2500 | nearlyzero | no | yes | yes | normal |



**Figure 2: Training Data Set for Hot Drink**

The system predicts the class label of an unknown sample using naïve Bayesian classification, given the same training data. The training data are in Figure 2. The data samples are described by the attributes Purc_Qty, Purc_Rate, Sale_Qty, Sale_Rate, Balance, Debt, Handover, Advertisement. The class label attribute, Forecasting, has two distinct values (namely, {high, normal, decline}). Let C1 correspond to the class Forecasting="high", C2 correspond to Forecasting= "normal" and C3 correspond to Forecasting = "decline". The unknown sample X wish to be classified.

Figure 3 is the forecasting result. According to the result, the marketers can know sales forecast for productivity near future for specified hot drink type. Most customers have such varied tastes and preferences that it is not possible to group them into large homogenous populations to develop marketing strategies. In fact, each customer wants to be served according to the individual and unique needs.

FORECASTING RESULT FOR GIVEN YEAR

| No | Year | Month | SaleQty | Balance | Forecasting |
|---|---|---|---|---|---|
| 1 | 2009 | Jan | 680 | more | normal |
| 2 | 2009 | Feb | 687 | nearlyzero | normal |
| 3 | 2009 | Mar | 740 | nearlyzero | high |
| 4 | 2009 | Apr | 760 | nearlyzero | high |
| 5 | 2009 | May | 686 | more | normal |
| 6 | 2009 | Jun | 490 | more | decline |
| 7 | 2009 | Jul | 488 | more | decline |
| 8 | 2009 | Aug | 687 | nearlyzero | normal |
| 9 | 2009 | Sep | 700 | nearlyzero | normal |
| 10 | 2009 | Oct | 774 | nearlyzero | high |
| 11 | 2009 | Nov | 730 | nearlyzero | normal |
| 12 | 2009 | Dec | 738 | nearlyzero | normal |

For New Year, nearly forecasting for productivity should be=Normal

**Figure 3: Sale Forecasting Result for The Hot Drink By Year**

## 6. Conclusion

Data mining can be viewed as a result of the natural evolution of information technology. This paper focus on the classification rules mining that based on Naïve Bayesian Classification. Despite its unrealistic independence assumption, the naïve Bayes classifier is surprisingly effective in practice since its classification decision may often be correct even if its probability estimates are inaccurate.

In this paper, the system has shown that data mining efforts can be useful in forecasting out comes. The information generated by data mining can be usefully applied by marketers to increase hot drinking's sales. By using this forecasting result, the marketers can make decision for future sales of enterprise or firm.

## 7. References


[1] An, Aijun, "Classification Methods", NewYork University, Canada. http://en.wikipedia.org/wiki/Granular computing

[2] Breiman, Leo, "Bagging Predictors', Department of Statistics, University of California at Berkeley, DMS-9212419.

[3] Han & Kamber, "Data Minng Concepts and Techniques" (Second Edition)

[4] Jakulin, A; Bratko, I; Smrke, D; Demsar, J ; Blaz Aupan, B "Attribute Interactions in Medical Data Analysis" Faculty of Computer and Information Science, University of Ljubljana, Trzaska, Ljubljana, Slovenia J.Stefan Institute, Jamova, Ljubljana, Slovenia Dept. of Traumatology, University Clinical Center, Ljubljana, Slovenia Dept. of Human and Mol. Genetics, Baylor College of Medicine, Houston, USA

[5] Khin Myo Aye, "Prediction Heart Disease Using Naïve Bayesian Classification".

[6] Kononenko, "Inductive and Bayesian Learning in Medical Diagnosis" Faculty of Electrical and Computer Engineering Trzaska 25, 61001 Ljubljana, Slovenia

[7] Pazzani, Michael Jones, 'Searching for Dependencies in Bayesian Classifiers". http://www.ics.uci,edu/~pazzani/Publications/Pazza

[8] Seifert, Jeffrey Welter, "Data Mining: An Overview", Analyst in Information Science and Technology Policy Resources, Science and Industry Division, Updated December 16,2004.

[9] Thu Thu Zan, "Treatment Guidelines for Diabetic Patient By Using Naïve Bayesian Classification".

[10] Witt Yee Win, "Classification of Nutritional Status by Using Naïve Bayesian Classification".