

Combination of Naïve Bayes Classifier and K-Nearest Neighbor (cNK) in the Classification Based Predictive Models

Elma Zannatul Ferdousy¹, Md. Mafijul Islam¹ & M. Abdul Matin¹

¹ Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh

Correspondence: Elma Zannatul Ferdousy, Department of Statistics, Jahangirnagar University, Savar, Dhaka - 1342, Bangladesh. Tel: 880-173-001-9575. E-mail: elma@juniv.edu

Received: March 12, 2013 Accepted: April 25, 2013 Online Published: May 22, 2013

doi:10.5539/cis.v6n3p48

URL: <http://dx.doi.org/10.5539/cis.v6n3p48>

Abstract

In this study, we present a new classifier that combines the distance-based algorithm K-Nearest Neighbor and statistical based Naïve Bayes Classifier. That is equipped with the power of both but avoid their weakness. The performance of the proposed algorithm in terms of accuracy is experimented on some standard datasets from the machine-learning repository of University of California and compared with some of the art algorithms. The experiments show that in most of the cases the proposed algorithm outperforms the other to some extent. Finally we apply the algorithm for predicting profitability positions of some financial institutions of Bangladesh using data provided by the central bank.

Keywords: classification, predictive model, K-Nearest Neighbor, Naïve Bayes, data mining

1. Introduction

Classification is one of the most important multivariate techniques used in statistics. It is closely related to prediction and interestingly the classification problem is sometimes called the prediction problem particularly in data mining. In statistics, classification is a procedure in which individual objects are placed into groups based on quantitative information on one or more characteristics inherent in the objects (referred to as traits, variables, characters, etc) and based on a training set of previously labeled objects. The problem can be stated as follows: given training data $\{(x_1, y_1), \dots, (x_n, y_n)\}$ produce a classifier $h: \mathcal{X} \rightarrow Y$, which maps an object $x \in \mathcal{X}$ to its classification label $y \in Y$.

There are several approaches those deal with classification problem. The statistical based algorithm Naïve Bayes Classifier and distance-based algorithm K-Nearest Neighbor are frequently used in prediction problem. In case of Naïve Bayes algorithm one of the main factors is to deal with numerical attributes. It is obvious because in the algorithm one must determine the conditional probability for each possible value of all attributes. To resolve this issue, we have to discretize numerical attributes into several classes by adopting a discretization technique from a wide range of options available. So the technique used for discretization plays an important role over the accuracy of the method. Several attempts have been made to increase the accuracy of the Naïve Bayes algorithm by adopting new discretization scheme (Yang & Webb, 2002).

In case of K Nearest Neighbor algorithm, the situation is quite opposite. It has issue regarding categorical attributes. As the algorithm selects a segment from the training data based on the distance, a distance measurement scheme for the categorical data must be obtained. Generally, it is done by different similarity measurement techniques. Several investigations also have been done to find a proper distance measurement scheme.

In this paper, we propose a new algorithm that combines these two classifiers mentioned above in such a way to resolve both the issues involved. That is, in the new algorithm we do not have to discretize the continuous variables and, at the same time, do not have to measure the distances between categorical attributes. The combination is done in a logical way, which is supposed to increase the performance over the said algorithm, and will be consistent in nature. The implementation of the proposed algorithm, over some standard datasets, reflects this fact.

We also implement new algorithm to determine the profitability position of some financial institutions of Bangladesh. Typically this profitability position is evaluated by the central bank. We predict the profitability

position with the accuracy of about 90% by our proposed model.

In the next section we will have a review of Naïve Bayes and K-Nearest Neighbor. In section 3 we shall mention the proposed algorithm and in section 4 we discuss the experimentation in detail. In section 5 we include the analysis of a dataset which is taken from Bangladesh. Finally, concluding remark and some future research are mentioned in section 6.

2. Naïve Bayes Classifier and K-Nearest Neighbor

In abstract, the probability model for a classifier is a conditional model.

$$P(C | x_1, \dots, x_n)$$

Over a dependent class variable C with a small number of outcomes or classes, conditional on several variables x_1 through x_n . Using Bayes' theorem, we may write:

$$P(C | x_1, \dots, x_n) = \frac{P(C) P(x_1, \dots, x_n | C)}{P(x_1, \dots, x_n)}$$

The numerator is equivalent to the joint probability model $P(C, x_1, \dots, x_n)$ which can be rewritten as follows, using repeated applications of the definition of conditional probability and so forth.

$$\begin{aligned} & P(C, x_1, x_2, \dots, x_n) \\ &= P(C) P(x_1, \dots, x_n | C) \\ &= P(C) P(x_1 | C) P(x_2, \dots, x_n | C, x_1) \\ &= P(C) P(x_1 | C) P(x_2 | C, x_1) P(x_3, \dots, x_n | C, x_1, x_2) \\ &= P(C) P(x_1 | C) P(x_2 | C, x_1) P(x_3 | C, x_1, x_2) P(x_4, \dots, x_n | C, x_1, x_2, x_3) \end{aligned}$$

Now the naive conditional independence assumptions come into play: assume that each attribute x_i is conditionally independent of every other attribute x_j for $j \neq i$. This means that $P(x_i | C, x_j) = P(x_i | C)$ and so the joint model can be expressed as

$$\begin{aligned} P(C, x_1, \dots, x_n) &= P(C) P(x_1 | C) P(x_2 | C) P(x_3 | C) \\ &= P(C) \prod_{i=1}^n P(x_i | C) \end{aligned}$$

This means that under the above independence assumptions, the conditional distribution over the class variable C can be expressed as

$$P(C | x_1, \dots, x_n) = \frac{1}{Z} P(C) \prod_{i=1}^n P(x_i | C)$$

Here, Z is a scaling factor dependent only on x_1, x_2, \dots, x_n , i.e., a constant if the values of the feature variables are known.

The corresponding classifier is the function classify defined as follows:

$$\text{classify}(x_1, \dots, x_n) = \arg \max_C P(C = c) \prod_{i=1}^n P(X_i = x_i | C = c)$$

In case of K-Nearest Neighbor, given a training set D and a test object $z = (x', y')$, the algorithm computes the distance (or similarity) between z and all the training objects $(x, y) \in D$ to determine its nearest-neighbor list, D_z (x is the data of a training object, while y is its class. Likewise, x' is the data of the test object and y' is its class).

Input: D , the set of training objects, and test object $z = (x', y')$.

Process: Compute $d(x', x)$, the distance between z and every object $(x, y) \in D$.

Select $D_z \subseteq D$, the set of k closest training objects to z .

Output: $y' = \arg \max_v \sum_{(x_i, y_i) \in D_z} I(v = y_i)$

3. The Proposed Algorithm

Combining classifiers to improve the accuracy is a common phenomenon now-a-days. Being simpler yet powerful algorithms both Naïve Bayes and KNN are ideal candidate for combination to achieve higher accuracy. Hsiao et al. (2008) proposed a combination for predicting sub cellular location of Eukaryotic proteins. Jiang et al. (2009) proposed another way of combining NB and KNN and showed the performance on some UCI dataset. More recently the performance of some sort of combination of these two methods is investigated in the field of Image classification (Timofte et al., 2012; McCann & Lowe, 2012). Previously, Xie et al. (2002) proposed an improved algorithm called Selective Neighborhood Naïve Bayes (SNNB). Lazkano and Sierra (2003) proposed the combination in a different way that combines the nearest neighbor with Bayesian network. Etzold (2003) introduced the approach of combination of these two classifiers for spam filtering. Jiang et al. (2006) proposed the new algorithm called Dynamic K-Nearest Neighbor Naive Bayes where they have weighted the attributes based on the mutual information between each attribute and class attributes.

The idea of the algorithm which we are going to propose is very simple. To classify a new object we first use the KNN algorithm to find the K Nearest Neighbor from the training dataset. While implementing the KNN, we shall not include the categorical attributes. The distance will be measured only by the numerical attributes. After selecting the K nearest object, we shall then build a model using the Naïve Bayes algorithm, but this time we are using only the categorical attributes. From the model, we shall classify the new object. So, this is a two step process. In the first step, only the numerical attributes are used to select the closest data of the new object. This makes sense, as numerically close objects are supposed to have the same characteristics. Now, after we get K nearest object of the new object, instead of taking simple voting scheme as in KNN, we are looking deeper in the characteristic of the categorical data and the relation to the class. So we use the Naïve Bayes for the purpose. In this way, both the numerical and categorical attributes are used for the classification of a new object without any alteration in the data. Hence, the proposed method keeps the data intact. No discretization or complex similarity measurement is required. Thus, the proposed method is nothing but a ‘combination of Naïve Bayes and K Nearest Neighbor’, in short, we write it as cNK.

Formally, we can describe cNK algorithm (see also Figure 1) as follows:

Step 1. Obtain the K-Nearest Neighbor of a new observation based on the numerical attributes.

Step 2. Use the set of K observations, found in step 1 as training data and use it to build a model exploiting the Naïve Bayes algorithm based only on the categorical attributes.

Step 3. Use the model built in step 2 to classify the new observation.

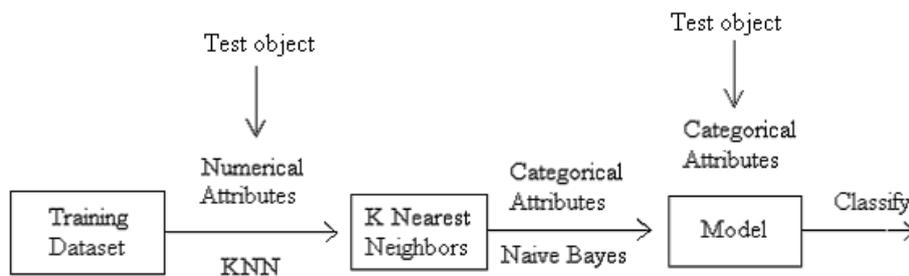


Figure 1. Graphical representation of the proposed algorithm

4. Experimental Evaluation

4.1 Datasets

In order to have a better understanding of how the newly proposed cNK algorithm works. We used numerical data illustration of six datasets obtained from Machine Learning Repository of University of California (<http://archive.ics.uci.edu/ml/datasets.html>) (Table 1). In doing so, we used accuracy as a measure of performance of the algorithm.

Table 1. Description of dataset used

Dataset	Title in Website	Number of Attributes	Number of Classes	Number of Objects
Hepatitis	Hepatitis	19	2	155
CRX	Credit Approval	15	2	690
Australian	Statlog (Australian Credit Approval)	14	2	690
Post-Operative	Post Operative Patient	8	3	90
Heart	Statlog (Heart)	13	2	270
German	Statlog (German Credit Data)	20	2	1000

Hepatitis Dataset: The hepatitis domain was donated by G. Gong in Carnegie-Mellon University on November 1988. It was used by Diaconis and Efron (1983). Gail-Gong reported 80 percent classification accuracy. It was used by Cestnik et al. (1987) and they showed 83 percent classification accuracy.

Crx Dataset: This file concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. There is no missing value in this dataset.

Australian Dataset: This dataset is originated from Quinlan (1997). The title of the dataset is “Australian Credit Approval”. This dataset is used for “simplifying decision trees” in 1987 and “C4.5: Programs for Machine Learning” in 1992. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. This dataset is interesting because there is a good mix of attributes continuous, nominal with small numbers of values, and nominal with larger numbers of values. There are also a few missing values.

Post-Operative Dataset: This dataset is created by Sharon Summers of University of Kansas. The title of the dataset is “Postoperative Patient Data.” It is donated by Jerzy W. Grzymala-Busse in 1993. It is used for program LERS_LB as a tool for knowledge acquisition in nursing by Budihardjo et al. (1991). It is also used for machine learning program LERS_LB 2.5 in knowledge acquisition for expert system development in nursing. The classification task of this database is to determine where patients in a postoperative recovery area should be sent to next. The accuracy of LERS is 48 percent.

Heart Dataset: This database contains no missing values. Here presence or absence of heart disease is predicted on the basis of age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina and oldpeak etc.

German Dataset: The title of this dataset is “German Credit Data”. This dataset classifies people described by a set of attributes as good or bad credit risks. This dataset is provided by Prof. Dr. Hans Hofman. This dataset is used by Eggermont et al. (2004) and Wang et al. (2003).

4.2 Experimental Design

It is necessary to have datasets that contains both numerical and categorical attributes as the newly proposed cNK algorithm treats the numerical and categorical attributes differently. Naïve Bayes and cNK algorithms are implemented by using “class” and “e1071” libraries of R programming language. The 10-fold cross validation method is used to find the accuracy of the algorithm for different values of K. See Figure 2 for stepwise description of the whole experimental design.

For wider comparison, we have chosen four algorithms viz discretize simple Bayes (Dougherty et al., 1995), selective simple Bayes with forward selection (Langley & Sage, 1994), tree augmented Naïve Bayes (Friedman et al., 1997) and lazy Bayesian rule-learning algorithm (Zheng & Webb, 2000). Moreover, sophisticated machine learning technique algorithms (used to estimate the values of the weights of a neural network) the Back Propagation (BP) algorithm (Mitchell, 1997), the SMO algorithm, and the 3-Nearest Neighbor (3NN) algorithm are also used as the representative of the Artificial Neural Network (ANN), the Support Vector Machine (Platt, 1999) and KNN (Aha, 1997), respectively. Subsequently, we compare the performance of BoostFSNB with Bagging decision trees and boosting decision trees that have been proved to be successful in many machine-learning problems (Quinlan, 1997).

- Each dataset is shuffled randomly.
- Producing disjoint training and test sets as follows.
 - First 10% test and 90% training
 - Second 10% test and 90% training
 - Third 10% test and 90% training
 -
 - Last 10% test and 90% training
- For each set of training and test data, run
 1. Naïve Bayes Classifier.
 2. Proposed Algorithm.
- For different values of K run the last algorithm.

Figure 2. Steps of the experimentation procedure

4.3 Experimental Results

We run cNK algorithm over each dataset for different values of K (number of neighbor to build up the Naïve Bayes Model). For experimentation purpose we gradually increase the value of K and run the same test. For distance metric we always use the normalized Euclidian distance. If we plot accuracy against the values of K, get a pick point. Thus from 10 fold cross validation over the training data we find the suitable value of K that gives the highest accuracy. We then use it over the test data.

For Hepatitis, Post-operative and CRX the cNK algorithm (86%, 71%, and 87%) performs better than the Naïve Bayes classifier (83%, 64% and 85%) for K equal to 70, 50 and 400; respectively (See Table 2). For Australian data, the cNK Algorithm (87%) also has some improvement over the accuracy of the Naïve Bayes classifier (86%) for K = 350. It is also clear (Table 2) that for all datasets the cNK algorithm has lower error rate than the Naïve Bayes classifier. These results confirm the initial hypothesis of the study. It is apparent that the cNK Algorithm outperforms the original Naïve Bayes classifier in every domain giving the accuracy up to 3 percent level.

Table 2. Accuracy of the algorithm using 10-fold cross validation

	NB	cNK
Hepatitis	82.67	86.00 (K = 70)
Post-operative	64.44	71.11 (K = 50)
CRX	85.07	86.52 (K = 400)
Australia	86.15	86.96 (K = 350)
Heart	82.96	85.92 (K = 165)
German	75.16	75.60 (K = 575)

Table 3. Comparison of results with some other state-of-the art algorithms

	cNK	NB	DiscNB	Wrapper NB	TAN	LBR	BP	SMO	3-NN
Australian	86.96	86.15	86.22	85.30	86.07	86.10	85.94	84.91	84.96
German	75.40	75.16	75.04	73.79	74.30	74.90	74.86	75.11	72.21
Hepatitis	86.00	82.67	84.28	82.55	84.83	84.91	84.29	85.03	80.85
Heart	85.92	82.96	83.47	78.75	82.97	83.54	82.98	84.03	81.82

Table 4. Comparing the cNK with other attempts to improve the Naïve Bayes

	cNK	BoostFSNB	AdaboostC4.5	BaggingC4.5	Bagging NB	Adaboost NB
Australian	86.96	85.9	84.01	86.38	77.81	81.16
German	75.60	72.83	70.75	73.38	75.13	75.09
Hepatitis	86	84.14	82.38	82.05	84.39	84.23
Heart	85.92	82.42	78.79	79.86	83.24	83.14

The cNK algorithm also outperforms DiscNB, WrapperNB, TAN, LBR, BP, SMO and 3-NN in every domain giving the accuracy up to 2, 7, 3, 2, 3, 2 and 4 percent, respectively (see Table 3). It is observed that the cNK algorithm outperforms the algorithms those tried to improve the classification accuracy of the simple Bayes algorithm. The overall performance of the cNK is quite impressive. The accuracy of the cNK algorithm compared with the simple Bayes algorithm while applying the bagging and boosting procedures of Bauer and Kohavi (1999). The cNK algorithm gives greater accuracy than BoostFSNB, single boosting simple Bayes (AdaBoostNB), AdaBoost C4.5, Bagging C4.5 and Bagging NB for all datasets (Table 4).

5. Predicting Profitability Position of Financial Institutions

The cNK algorithm is implemented to find the profitability position of some financial institutions of a country. The central bank of the country evaluates such profitability position and determines whether it is satisfactory or not. The position is determined by analyzing quite a many attributes. It is performed by some regulatory organizations that involve complex calculations and procedures. However, using the available attributes and the class assigned against them in previous regulation, we may build a model to predict the possible position of any institute provided the value of the attributes (Table 5). Institute type refers to the type of the financial institution namely nationalized commercial banks, government-owned development financial institutions, private commercial banks and foreign commercial banks. Total 'interest income' refers to the income from the interest that it is the biggest source of income for a financial institution. Total 'interest expense' is the amount of money spends in for borrowing money. It usually refers to the amount of money given to the depositors as interest by the financial institutions. Without earning from interest income, now-a-days, most of the financial institutions shift their income source through other diversifying into fee-earning activities, such as corporate cash management, check collection, consumer annual fees on credit cards, and monthly service charges on deposit accounts. These are the noninterest income which may also includes many new activities, such as fees from participation in mutual fund commissions, investment advisor fees in merger and acquisition activities, and securities underwriting fees.

Table 5. Central bank dataset attributes

Attribute Name	Attribute Type	Attribute Name	Attribute Type
Bank type	Categorical	Inter bank deposit	Continuous
Total interest income	Continuous	Non interest	Continuous
Total interest expenses	Continuous	NIBPT	Categorical
Non interest income	Continuous	Total assets	Continuous
Operating income	Continuous	Number of ATM	Continuous
Total deposit	Continuous	Deposit type	Categorical
Total loans	Continuous	Number of loan account	Continuous
Classified loan	Continuous	Number of employees	Continuous
Interest earning assets	Continuous	Interest bearing liabilities	Continuous
Off balance sheet item	Categorical	Class	Categorical

Operating income is the income reported after deducting the expenses without deducting the taxes and interests. Interests earning assets and interest bearing liabilities are the total value of earning assets and liabilities that cost

interest of the financial institutions. Total loan is the amount given as loan by the institution. Classified loans are those that given according to the rules and regulations of the institution but later become suspect by the bank examiners.

Table 6. Accuracy of the proposed algorithm on central bank dataset (2 class)

Value of K	Accuracy
3	81.25
4	79.16
7	79.16
10	81.25
15	83.33
19	87.50
21	89.58
28	87.50
31	87.50

Inter bank deposit is the deposit held by the institutions for other banks, usually a correspondent. Usually a due to account is held. Deposit excluding this is shown in another column. Urban and rural represents the number of urban and rural branch of the institution.

Table 7. Comparing the performance of the algorithms over the central bank dataset

	cNK	NB	DiscNB	TAN	LBR	BP	SMO	3-NN
Central Bank	89.58	77.14	78.35	83.65	82.36	86.12	85.92	81.96

Number of ATM, number of employees, number of deposit accounts, number of loan account fields are self explanatory. Obligations that are contingent liabilities of a bank, and thus do not appear on its balance sheet. In general, off-balance sheet items include the following: direct credit substitutes in which a bank substitutes its own credit for a third party, including standby letters of credit; irrevocable letters of credit that guarantee repayment of commercial paper or tax-exempt securities; risk participations in bankers' acceptances; sale and repurchase agreements; and asset sales with recourse against the seller; interest rate swaps; interest rate options and currency options, and so on.

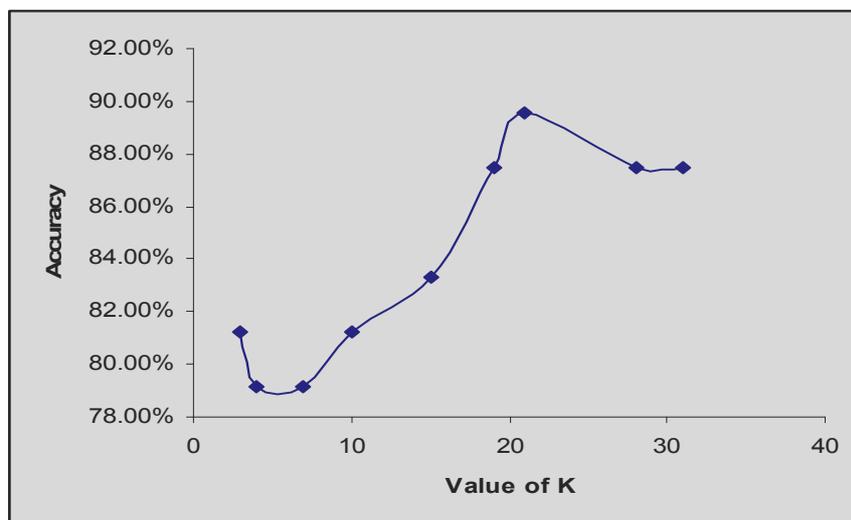


Figure 3. Accuracy of the proposed algorithm without PCA of central bank dataset (2 classes)

This dataset is not used ever before for classification purpose and is used for the first time to measure the accuracy of the Naïve Bayes and cNK algorithm. There are 20 attributes and 48 observations, 15 of them are numerical, 4 of them are categorical and 1 class attribute (Table 5). We apply some of the algorithms over the dataset and the proposed cNK algorithm on the data to predict the profitability position. The best result (89.58%) for cNK comes for the value of $K = 21$ (see Table 6 and Figure 3). The details result is shown in Table 7. It shows that the proposed algorithm gives better result.

6. Conclusion

This paper leads to a new classifier cNK which combines Naïve Bayes and K-Nearest Neighbor. We implement the Naïve Bayes classifier and the cNK algorithm on some standard datasets using R code. The experimental results clearly show that the performance of the proposed algorithm is better than that of the Naïve Bayes in many cases. Some of the datasets are also compared with the other attempts that have tried to improve the accuracy of the simple Bayes algorithm and some other algorithms such as Boost FSNB, DiscNB, WrapperNB, TAN, LBR, BP, SMO, Adaboost C4.5, Bagging C4.5, BaggingNB and AdaboostNB. In most of the cases, the cNK algorithm gives the better result. It also gives better result in case of central bank dataset providing profitability analysis. Though the experiments are done in limited context over a small number of dataset and based on that it cannot be declared the best approach, it shows that the proposed algorithm can outperform other algorithms in several cases. Further investigation can be done to evaluate the performance of the algorithm using extensive experimentation on real and simulated dataset. Based on the evidences it may be concluded that a simple approach has been proposed that may improve the performance of the simple Bayes classifier through combining two classifiers where the data contains continuous and categorical attributes.

References

- Aha, D. (1997). *Lazy learning*. Dordrecht: Kluwer Academic Publishers.
- Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36, 105-139. <http://dx.doi.org/10.1023/A:1007515423169>
- Budihardjo, A., Grzymala-Busse, J., & Woolery, L. (1991). Program LERS_LB 2.5 as a tool for knowledge acquisition in nursing, In *Proceedings of the 4th International Conference on Industrial & Engineering Applications of AI & Expert Systems* (pp. 735-740).
- Cestnik, G., Kononenko, I., & Bratko, I. (1987). Assistant-86: A Knowledge- Elicitation Tool for Sophisticated Users. In I. Bratko, & N. Lavrac (Eds.), *Progress in Machine Learning* (pp. 31-45). Sigma Press.
- Diaconis, P., & Efron, B. (1983). Computer-Intensive Methods in Statistics. *Scientific American*, 248, 116-130. <http://dx.doi.org/10.1038/scientificamerican0583-116>
- Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 194-202). Tahoe City, CA: Morgan Kaufmann.
- Eggermont, J., Kok, J. N., & Kusters, W. A. (2004). *Genetic Programming for data classification: partitioning the search space*. SAC.
- Etzold, D. (2003). *Improving spam filtering by combining Naïve Bayes with simple K-Nearest Neighbor searches*. arXiv:cs/0312004v1 [cs.LG].
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2-3), 131-163. <http://dx.doi.org/10.1023/A:1007465528199>
- Hsiao, H. C. W., Chen, S., Chang, J. P., & Tsai, J. J. P. (2008). Predicting Subcellular Locations of Eukaryotic Proteins Using Bayesian and K-Nearest Neighbor Classifiers. *Journal of Information Science and Engineering*, 24(5), 1361-1375.
- Jiang, L., Wang, D., Cai, Z., Jiang, S., & Yan, X. (2009). Scaling Up the Accuracy of K-Nearest-Neighbour Classifiers: A Naive-Bayes Hybrid. *International Journal of Computers and Applications*, 31(1). <http://dx.doi.org/10.2316/Journal.202.2009.1.202-2453>
- Jiang, L., Zhang, H., & Cai, Z. (2006). Dynamic K-Nearest-Neighbor Naive Bayes with Attribute Weighted. *Proceedings of the 3rd International Conference on Fuzzy Systems and Knowledge Discovery*, LNAI 4223, pp.365-368. Springer Press. http://dx.doi.org/10.1007/11881599_41
- Langley, P., & Sage, S. (1994). Induction of selective Bayesian classifiers. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence* (pp. 399-406). Seattle, WA: Morgan Kaufmann.

- Lazkano, E., & Sierra, B. (2003). BAYES-NEAREST: A New Hybrid Classifier Combining Bayesian Network and Distance Based Algorithms. *EPIA*, 171-183.
- McCann, S., & Lowe, D. (2012). *Local naive bayes nearest neighbor for image classification*. CVPR.
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
- Platt, J. (1999). Using sparseness and analytic QP to speed training of support vector machines. *Advances in neural information processing systems, 11*. MA: MIT Press.
- Quinlan, J. R. (1997). Simplifying decision trees. *Int J Man-Machine Studies*, 27, 221-234. [http://dx.doi.org/10.1016/S0020-7373\(87\)80053-6](http://dx.doi.org/10.1016/S0020-7373(87)80053-6)
- Timofte, R., Tuytelaars, T., & Van Gool, L. (2012), Naive bayes image classification: beyond nearest neighbors. *Proceedings of 11th Asian conference on computer vision - ACCV 2012*, November 5-9, 2012, Daejeon, Korea.
- Wang, K., Zhou, S., Ada, W. F., & Jeffrey, X. Y. (2003). *Mining Changes of Classification by Correspondence Tracing*. SDM.
- Xie, Z., Hsu, W., Liu, Z., & Mong-Li, L. (2002). SNNB: A Selective Neighborhood Based Naïve Bayes for Lazy Learning. *PAKDD*, 104-114.
- Yang, Y., & Webb, G. I. (2002). A Comparative Study of Discretization Methods for Naïve-Bayes Classifiers, In *Proceedings of PKAW 2002, The 2002 Pacific Rim Knowledge Acquisition Workshop, Tokyo, Japan, 2002*, pp. 159-173.
- Zheng, Z., & Webb, G. I. (2000). Lazy learning of Bayesian rules. *Machine Learning*, 41, 53-84. <http://dx.doi.org/10.1023/A:1007613203719>